Executive Function and the Frontal Lobes: A Meta-Analytic Review

Julie A. Alvarez[1] **and Eugene Emory**[2]

Published online: 1 June 2006

Currently, there is debate among scholars regarding how to operationalize and measure executive functions. These functions generally are referred to as "supervisory" cognitive processes because they involve higher level organization and execution of complex thoughts and behavior. Although conceptualizations vary regarding what mental processes actually constitute the "executive function" construct, there has been a historical linkage of these "higher-level" processes with the frontal lobes. In fact, many investigators have used the term "frontal functions" synonymously with "executive functions" despite evidence that contradicts this synonymous usage. The current review provides a critical analysis of lesion and neuroimaging studies using three popular executive function measures (Wisconsin Card Sorting Test, Phonemic Verbal Fluency, and Stroop Color Word Interference Test) in order to examine the validity of the executive function construct in terms of its relation to activation and damage to the frontal lobes. Empirical lesion data are examined via meta-analysis procedures along with formula derivatives. Results reveal mixed evidence that does not support a one-to-one relationship between executive functions and frontal lobe activity. The paper concludes with a discussion of the implications of construing the validity of these neuropsychological tests in anatomical, rather than cognitive and behavioral, terms.

KEY WORDS: Executive function; Frontal lobe; Neuropsychology; Meta-analysis.

Executive functions generally refer to "higher-level" cognitive functions involved in the control and regulation of "lower-level" cognitive processes and goal-directed, future-oriented behavior. Over 2500 scientific articles have been published on this topic in the past 10 years. They have examined the role of executive functions in normal development (e.g., Espy and Kaufmann, 2002), Attention-Deficit/Hyperactivity Disorder (e.g., Sergeant et al., 2002), Antisocial Personality Disorder (e.g., Morgan and Lilienfeld, 2000), Parkinson's disease (e.g., Dagher et al., 2001), and neuropsychiatric disorders including Schizophrenia and Obsessive-Compulsive Disorder (e.g., Nieuwenstein et al., 2001; Perry et al., 2001).

The component processes of executive functions have been investigated by means of factor-analytic studies. The Wisconsin Card Sorting Test (WCST) taps neuropsychological processes involving (a) cognitive flexibility, (b) problem-solving, and (c) response maintenance (Greve et al., 2002). In empirical and theoretical papers, the processes that emerge as component factors underlying executive functions are: (a) inhibition and switching (Baldo et al., 2001; Burgess et al., 1998; Miyake et al., 2000; Rabbitt, 1997; Sergeant et al., 2002; Troyer et al., 1998; Welsh, 2002), (b) working memory (Barcelo and Knight, 2002; Barcelo and Rubia, 1998; Barkley, 1996; Denckla, 1996; Dunbar and Sussman, 1995; Pennington et al., 1996; Sergeant et al., 2002; Stuss et al., 1998; Stuss et al., 2001; Welsh, 2002; Zelazo et al., 1997), and (c) sustained and selective attention (Barcelo, 2001; Barkley, 1996; Manly and Robertson, 1997; Stuss et al., 1998; Stuss et al., 2001).

There remains an on-going debate regarding whether executive functions are regulated by the frontal lobes (e.g., Miyake et al., 2000; Welsh, 2002), leading to ambiguity of definition (Eslinger, 1996; Stuss and Alexander, 2000; Tranel et al., 1994). Tests of abstract reasoning and verbal

[1]Department of Psychology, Emory University, Atlanta, Georgia, 30322, USA.

[2]Center for Prenatal Assessment and Human Development, Emory University, Atlanta, Georgia, USA.

fluency commonly are referred to as "frontal lobe" measures because persons with severe lesions in this area do poorly on them (e.g., Benton, 1968; Milner, 1963). Eventually it became standard practice to conclude that individuals who perform poorly on executive function tests had a "frontal lobe deficit" (Stuss et al., 2000). In fact, it "is virtually impossible to find a discussion of prefrontal lobe lesions that does not make reference to disturbances of executive functions and, in parallel fashion, there is rarely a discussion of disturbances of executive functions that does not make reference to dysfunction in prefrontal brain regions" (Tranel et al., 1994, p. 126). Despite the circularity of linking anatomy (frontal lobes) with a neuropsychological construct (executive functions), the frontal lobes continue to be linked to measures of executive function (Duke and Kaszniak, 2000; Duffy and Campbell, 2001). In addition to this circularity, the sensitivity and specificity of executive function measures to lesions in the frontal lobes is inconsistent (e.g., Bigler, 1988; Costa, 1988; Wang, 1987). Several researchers have found that persons with frontal lesions perform within normal limits on executive function tests (e.g., Ahola et al., 1996; Damasio, 1994; Eslinger and Damasio, 1985; Heck and Bryer, 1986; Shallice and Burgess, 1991b) and others have found that persons with non-frontal or diffuse lesions perform as poorly as persons with frontal lesions on these tests (e.g., Anderson et al., 1991; Axelrod et al., 1996; Crockett et al., 1986; Grafman et al., 1990; Heaton, 1981; Robinson et al., 1980).

Because of the fuzziness surrounding the relationship between executive functions and the frontal lobes (e.g., Denckla, 1996; Stuss and Alexander, 2000), there is a limited literature establishing the construct validity of classic executive function tasks (e.g., Barcelo, 2001; Kafer and Hunter, 1997; Phillips, 1997; Rabbitt, 1997; Reitan and Wolfson, 1994). The current paper aims to examine the executive function construct and its method of assessment by reviewing lesion and neuroimaging studies using three executive function measures: Wisconsin Card Sorting Test (WCST), Phonemic Verbal Fluency, and Stroop Color Word Interference Test (Stroop). These three tasks were chosen because they are among the most frequently used executive function measures (e.g., Baddeley, 1996; Butler et al., 1991; Carlin et al., 2000; Goodglass and Kaplan, 1979; MacLeod, 1991; Stuss and Levine, 2002) and there are several theoretical papers examining the underlying cognitive processes involved in each (e.g., Greve et al., 2002; Miyake et al., 2000). Lesion studies will address whether these tests distinguish between persons with frontal lobe injuries and controls, and functional neuroimaging studies will address whether the frontal lobes are activated when healthy individuals are performing

these tasks. Lesion studies also will be examined quantitatively across all three measures (where effect size data are available). There are only three known meta-analytic studies in this area, but they are limited to the WCST (Demakis, 2003; Rhodes, 2004) and verbal fluency tests (Henry and Crawford, 2004).

We shall limit the current review to studies of adult populations because of changes in both executive functions and frontal lobe development among children (e.g., Levin et al., 1997; Malloy and Richardson, 2001; Welsh, 2002). The paper will conclude with a brief review of alternative executive function measures to augment existing protocols, such as the gambling task and the Multiple Errands Test (Bechara et al., 1994; Burgess, 2000), along with possible directions for future research.

FRONTAL-SUBCORTICAL CIRCUITRY

Several researchers (e.g., Cummings, 1995; Duke and Kaszniak, 2000; Sbordone, 2000; Stuss and Benson, 1984) suggest that there are three principal frontal-subcortical circuits involved in cognitive, emotional, and motivational processes: (a) dorsolateral; (b) ventromedial; and (c) orbitofrontal. The dorsolateral frontal cortex projects primarily to the dorsolateral head of the caudate nucleus and has been linked to *executive functions*, including verbal and design fluency, ability to maintain and shift set (as measured by the WCST), planning, response inhibition, working memory, organizational skills, reasoning, problem-solving, and abstract thinking (Cummings, 1993; Duke and Kaszniak, 2000; Ettlinger et al., 1975; Grafman and Litvan, 1999; Jonides et al., 1993; Malloy and Richardson, 2001; Milner, 1971; Stuss et al., 2000). The ventromedial circuit, which is involved in motivation, begins in the anterior cingulate and projects to the nucleus accumbens. Lesions to this region often produce apathy, decreased social interaction, and psychomotor retardation (Sbordone, 2000). The orbitofrontal cortex projects to the ventromedial caudate nucleus and is linked to socially appropriate behavior. Lesions to this area cause disinhibition, impulsivity, and antisocial behavior (Blumer and Benson, 1975; Cummings, 1995).

In addition to the primary projections noted above, the frontal lobes have multiple connections to cortical, subcortical and brain stem sites and should "be conceived as one aspect of an executive system involving many structures of the central nervous system" (Duffy and Campbell, 2001, p. 116). The basis of "higher-level" cognitive functions such as inhibition, flexibility of thinking, problem solving, planning, impulse control, concept formation, abstract thinking, and creativity often arise from much

Table 1. Procedures and Scoring of Executive Function Measures

Test	Items	Instruction(s)	Response	Scoring
WCST	Geometric figures on 128 response cards	Match response cards to four key cards	Match cards to color, form, and number	Categories, Errors
Verbal Fluency	Specific letters	Say as many words as possible beginning with a specific letter	Generate words beginning with "F, A, S" in 60 seconds for each letter	Total number of words generated (minus errors and repetitions)
Stroop	3 sheets of 100 items (black words, color patches, color words)	Read words, colors, and color words printed in incongruous colored ink	Word reading and color naming	Total time in seconds; interference difference score

Note. WCST = Wisconsin Card Sorting Test. This table presents an extremely condensed version of the procedures and scoring protocols of the executive function measures. More details may be found in Heaton et al. (1993), Benton and Hamsher (1976), and Stuss et al. (2001).

simpler, "lower-level" forms of cognition and behavior. Thus, the concept of executive function must be broad enough to include anatomical structures that represent a diverse and diffuse portion of the central nervous system.

EXECUTIVE FUNCTION MEASURES

Wisconsin Card Sorting Test

Origin and History

The WCST (Heaton et al., 1993) often has been cited as *the* most frequently used measure of executive functioning (Baddeley, 1996; Barcelo and Knight, 2002; Reitan and Wolfson, 1994; Spreen and Strauss, 1998; Stuss and Benson, 1986; Stuss and Levine, 2002), and is regularly used by over 70% of neuropsychologists (Butler et al., 1991). The original construction of the task (Berg, 1948) was based partly on sorting test methods to assess abstract reasoning and set shifting in humans and animals (Goldstein, and Scheerer, 1941; Vigotsky, 1934; Weigl, 1941; Zable and Harlow, 1946). Milner (1963) adapted the procedure developed by Grant and Berg (1948), and her version became the model for the current "standard" administration of the WCST (known as the "Heaton version"). The WCST eventually became a popular neuropsychological test (Butler et al., 1991) due to its reported sensitivity to frontal lobe lesions (e.g., Drewe, 1974; Milner, 1963) and the publication of a standardized procedure (Heaton, 1981). Table 1 presents a condensed version of the WCST procedures and scoring protocols (as well as those for the additional two executive function measures to be reviewed in this paper). Readers interested in a more detailed description of procedures and scoring should consult the respective manuals (Benton and Hamsher, 1976; Heaton et al., 1993; Stuss et al., 2001).

Lesion Studies

Twenty-five studies examining the effect of various brain lesions on WCST performance in adult populations were reviewed (see Table 2). There have been two previous qualitative reviews of this literature (Mountain and Snow, 1993; Reitan and Wolfson, 1994) and two meta-analytic studies in this area (Demakis, 2003; Rhodes, 2004). The state of the evidence regarding WCST sensitivity and specificity to frontal lobe lesions follows.

Twelve studies indicate that adults with frontal lobe lesions perform worse than healthy controls, and 10/16 studies suggest that persons with frontal lobe lesions perform worse than persons with non-frontal lobe lesions (see Table 2). In contrast, two single-case design studies found that persons with focal frontal lobe damage exhibited *no* deficits compared to *norms* (Eslinger and Damasio, 1985; Heck and Bryer, 1986), and four studies found *no* differences between frontal and diffuse or basal ganglia comparison groups (Eslinger and Grattan, 1993; Heaton, 1981; Heaton et al., 1993; Robinson et al., 1980). In general, these studies support the sensitivity of the WCST to frontal lobe lesions, as compared to non-frontal lobe lesions, but they do not support the specificity of the test to frontal lobe lesions.

In addition, five studies *failed* to demonstrate significant differences between frontal and non-frontal groups (Anderson et al., 1991; Axelrod et al., 1996; Crockett et al., 1986; Grafman et al., 1990; van den Broek et al., 1993). As seen in Table 2, these five studies do not appear to differ from those in which persons with frontal lobe lesions had impaired performance relative to controls. Further evidence that the WCST is a sensitive but not specific marker of frontal lobe damage is provided by studies that find poorer performance on the instrument among patients with a posterior lesion (Teuber et al., 1951) and patients with thalamic lesions (Wallesch et al., 1983) relative to controls.

Table 2. WCST Lesion Studies for Qualitative Review

First author	Year	Test	Lesion Etiology	Comparison group	Sample Size[a]	Mean Age[b]	Results
Anderson	1991	Standard	Diverse	Non-frontals	73	50.1	No significant differences between frontal and non-frontal groups
Axelrod	1996	Standard	Diverse	(1) Non-frontals	109	–	No significant differences between frontal and non-frontal groups
				(2) Normals	110	–	Frontal group performed significantly worse than the normal group
Barcelo	2002	Modified	Stroke	Normals	14	64.5	Frontal group performed significantly worse than the normal group
Bornstein	1986	Standard[c]	Diverse	Non-frontals	53	36.6	Frontal group performed significantly worse than the non-frontal group
Crockett	1986	Standard	–	Non-frontals	40	40.6	No significant differences between frontal and non-frontal groups
de Oliveira-Souza	2001	Modified	Gun shot wound– right frontal lobe	None	1	28	Patient with injury continued to perform very poorly even after 52 administrations
Drewe	1974	Standard	Diverse	Non-frontals	85	34.6	Frontal group performed significantly worse than the non-frontal group
Eslinger	1985	Standard[c]	Tumor	None	1	44	Patient with bilateral ablation of orbital and mesial frontal cortices demonstrated superior WCST performance
Eslinger	1993	Standard	28/30 cases were unilateral CVAs	(1) Non-frontals	20	59.2	Frontal group performed significantly worse than the non-frontal group
				(2) Normals	20	63.4	Frontal group performed significantly worse than the normal group
Ettlinger	1975	Standard	Unilateral frontal-cortical excisions for relief of focal epilepsy	–	–	–	Patients with DLPFC excisions performed worse than those with orbital excisions; patients with left frontal lesions performed worse than those with right frontal lesions
Grafman	1990	Standard[c]	Penetrating brain wounds	(1) Non-frontals	309	–	No significant differences between frontal and non-frontal groups
				(2) Normals	210	–	Frontal group performed significantly worse than the normal group
Grafman	1986	Standard[c]	Penetrating brain wounds	Non-frontals	112	36.4	Patients with DLPFC lesions performed worse than patients with orbitofrontal and non-frontal lesions
Heaton	1981	Standard	Diverse	(1) Non-frontals	78	42.5	Frontal group performed significantly worse than the non-frontal group
				(2) Normals	193	39.5	Frontal group performed significantly worse than the normal group
Heaton	1993	Standard	Diverse	(1) Non-frontals	113	41.9	Frontal group performed significantly worse than the non-frontal group
				(2) Normals	415	44.8	Frontal group performed significantly worse than the normal group
Heck	1986	Standard[c]	Encephalitis	None	1	41	Patient with bilateral atrophy of OPFC and DLPFC exhibited no deficits
Janowsky	1989	Standard	Diverse	Normals	18	62.4	Frontal group performed significantly worse than the normal group
Milner[d]	1963	Standard	Unilateral frontal excisions for relief of epilepsy	(1) Non-frontals	71	26.5	Both frontal groups performed significantly worse than both non-frontal groups
				(2) Non-frontals	23	30.6	
Nelson	1976	Modified	Diverse	Non-frontals	53	43.0	Frontal group performed significantly worse than the non-frontal group
Podell	1995	Standard	Diverse	(1) Non-frontals	21	40.1	No results reported for frontal versus non-frontal groups
				(2) Normals	24	40.9	Frontal group performed significantly worse than the normal group

Table 2. (Continued)

First author	Year	Test	Lesion Etiology	Comparison group	Sample Size[a]	Mean Age[b]	Results
Robinson	1980	Standard	Diverse	(1) Non-frontals	46	42.8	Frontal group performed significantly worse than the non-frontal group
				(2) Normals	146	39.2	Frontal group performed significantly worse than the normal group
Stuss	1983	Modified	Bilateral OPFC lesions due to SZ	Normals	21	–	Frontal group performed worse than the normal group in 2nd set of 64 cards only
Stuss	2000	Modified	Diverse	(1) Non-frontals	46	50	DLPFC and superior medial (but not inferior medial) frontal groups performed significantly worse than the non-frontal and normal groups
				(2) Normals	51	53	
Teuber	1951	Modified	Gun shot wounds	Non-frontals	40	–	"Posterior" group made more errors than "anterior" group
van den Broek	1993	Modified	Diverse	(1) Non-frontals	76	49.8	No significant differences between frontal and non-frontal groups
				(2) Normals	139	45.9	Frontal group performed significantly worse than the normal group
Wallesch	1983	Modified	Diverse thalamic lesions	Normals	23	60.3	Patients with thalamic lesions performed worse than normals (no frontal group)

Note. Dashes indicate that no information was provided. CVA = cerebral vascular accident; DLPFC = dorsolateral prefrontal cortex; OPFC = orbito prefrontal cortex; SZ = schizophrenia.

[a]Combined sample size for frontal and comparison groups.

[b]Mean age of frontal and comparison groups.

[c]Standard methodology is assumed, although it is not explicitly stated in the article which method was used.

[d]Milner (1963) tested two groups of frontal and non-frontal surgical patients: group 1 was tested before surgery and approximately 18 days after surgery, and group 2 was tested only once between 2 weeks and 15 years after surgery.

Functional Neuroimaging Studies

There is one known qualitative review of brain activation during WCST performance (Barcelo, 2001). Five additional studies not reported by Barcelo (2001) are reviewed here (Catefau et al., 1998; Haines et al., 1994; Rezai et al., 1993; Riehemann et al., 2001; Weinberger et al., 1986). The central question of this section is whether the frontal lobes are activated during WCST performance in healthy adult populations (see Table 3). For studies that included disordered populations as well in their analyses (e.g., persons with schizophrenia), only the data for the healthy control group were reported in Table 3. Event-related potentials, electroencephalographic, and magnetoencephalographic (MEG) studies (e.g., Barcelo et al., 2000; Çiçek, and Nalçaci, 2001) tend to have poor spatial resolution (Cabeza and Nyberg, 2000) and were excluded from this review. In addition, the event-related potential literature has been reviewed extensively elsewhere (Barcelo, 2001).

Several studies found increased activation in the dorsolateral prefrontal cortex during performance of the WCST (see Table 3), which is commensurate with the results of five lesion studies (Ettlinger et al., 1975; Grafman et al., 1986; Heck and Bryer, 1986; Milner, 1963; Milner, 1971; Stuss et al., 2000). Activation occurs in other frontal areas as well, including the ventromedial and orbitofrontal cortices. Collectively, these results seem to suggest that a bilaterally intact prefrontal cortex, especially the dorsolateral prefrontal cortex, is necessary for "normal" WCST performance, but it is unclear whether the WCST primarily activates the left or right prefrontal cortex (see Table 3).

A number of *non*-frontal brain regions are activated by the WCST, including the inferior parietal cortex (Berman et al., 1995; Nagahama et al., 1997; Nagahama et al., 1996; Tien et al., 1998), basal ganglia (Mentzel et al., 1998), temporo-parietal association cortex (Konishi et al., 1998), and occipito-temporal, temporal pole, and occipital cortices (Marenco et al., 1993; Nagahama et al., 1996; Ragland et al., 1998). These results are consistent with lesion studies reporting no significant differences between frontal groups and non-frontal, diffuse, or basal ganglia comparison groups (Anderson et al., 1991; Axelrod et al., 1996; Crockett et al., 1986; Eslinger and Grattan, 1993; Grafman et al., 1990; Heaton, 1981; Heaton et al., 1993;

Table 3.   WCST Neuroimaging Studies for Qualitative Review

First author	Year	WCST Test	Control Test(s)	Imaging Technique(s)	Sample Size	Mean Age	Areas of Increased Activation
Berman	1995	Modified computerized	Matching-to-sample task	PET O[15] MRI T-2	40	25.0	DLPFC, OPFC, VMPFC, and several non-frontal areas, especially inferior parietal lobules
Catefau	1998	–	Relaxation	SPECT	13	19.8	Left inferior cingulate and left posterior frontal region
Haines	1994	–	Number matching task	rCBF	11	25-62[a]	Left DLPFC
Kawasaki	1993	Standard	Relaxation	SPECT	10	28.9	Left DLPFC
Konishi	1999	Modified computerized	–5 s from previous series	fMRI 1.5-T	7	24-40[a]	Posterior part of the bilateral inferior frontal sulci
Konishi	1998	Modified computerized	–5 s from previous series	fMRI 1.5-T	7	–	DLPFC, VMPFC, and temporo-parietal cortices
Marenco	1993	Standard computerized	Matching-to-sample task	SPECT	17	31.35	Right DLPFC and left occipital cortex
Mentzel	1998	Standard computerized	Relaxation	fMRI 1.5-T	31	28.8	Strongest activation in right DLPFC and VMPFC; also increased activation in basal ganglia and mesial thalamus
Nagahama	1997	Modified computerized	Number matching task	PET O[15]	(1) 6	22.2	Both groups had increased activation in VMPFC, left DLPFC, left inferior parietal lobule, and left striate and prestriate cortices. The activation was significantly lower, however, in the elderly subjects
					(2) 6	67.7	
Nagahama	1996	Modified computerized	Matching-to-sample task	PET O[15]	18	24.7	VMPFC, DLPFC, IPL, left superior occipital gyrus, and left cerebellum
Nagahama	1998	Modified computerized	Matching-to-sample task	PET O[15]	6	22-26[a]	DLPFC, VMPFC, parieto-occipital cortex, and inferior occipital gyrus
Parellada	1998	Standard	Relaxation	SPECT	15	22.53	Inferior and superior DLPFC
Ragland	1998	Modified	Relaxation and number matching task	PET O[15] MRI	15	29.6	Inferior frontal, occipitotemporal, and temporal pole regions
Rezai	1993	Modified computerized	Relaxation	SPECT	13	41.54	Left lateralized frontal region
Riehemann	2001	Modified computerized	Color-only card sorting	fMRI 1.5-T	9	36.3	Right middle frontal gyrus
Tien	1998	Modified computerized	Matching-to-sample task	SPECT	5	19-24[a]	DLPFC, right VMPFC, and inferior parietal cortex
Volz	1997	Modified computerized	Relaxation and tapping	fMRI 1.5-T	31	28.8	Right DLPFC and right VMPFC; and, to a lesser extent, in medial thalamic nuclei
Weinberger	1986	Modified computerized	Number matching	rCBF	25	30.7	Bilateral DLPFC

Note. Dashes indicate that no information was provided. DLPFC = dorsolateral prefrontal cortex; fMRI = functional magnetic resonance imaging; IPL = inferior parietal lobules; MRI = magnetic resonance imaging; OPFC = orbito prefrontal cortex; PET = positron emission tomography; rCBF = regional cerebral blood flow; SPECT = single photon emission computed tomography; VMPFC = ventromedial prefrontal cortex.
[a]No mean age provided.

Robinson et al., 1980; van den Broek et al., 1993). That the WCST activates a widespread network of brain regions is consistent with distributed neuronal network modeling accounts of the test (e.g., Dehaene and Changeux, 1991). A parsimonious explanation of the WCST results supports the idea that a distributed network of neural circuits is activated when task demands involve integrated functioning. For example, activities of daily living, such as planning a trip to the store, involve overt and covert behavior components. At the overt level, the individual may search for the appropriate writing instruments, write

down directions, and make a list of items. At the covert level, the individual may engage long-term and short-term memory functions, visualize a path to the store and where items are located, and plan a budget that is within the parameters of the resources available. One could refer to these activities as internal and external (or implicit and explicit) representations of cognitive ability that fall within the purview of executive functions. A fundamental tenet of this review is that executive functions recruit from a wide range of functional abilities that are orchestrated in part by the frontal lobes. Thus, any high-level cognitive

task is likely to require participation of both subcortical and cortical regions, many of which have neural paths leading to the frontal lobes.

Over half of the WCST neuroimaging studies had sample sizes less than 15 subjects, slightly less than half did not include adequate control tasks in their experimental designs, and all used the subtraction method to compare activation between a target and reference condition, a method that has been called into question by several authors (see Cabeza and Nyberg, 2000 for elaboration). There are additional delimiting factors affecting the interpretation of the WCST imaging results, including no examination of non-frontal brain areas in the analyses, simply averaging brain activation across the entire duration of the WCST, and poor temporal resolution of the hemodynamic neuroimaging procedures (e.g., PET and fMRI). Despite these concerns, the results seem to suggest that WCST performance activates a distributed neural network involving both frontal and non-frontal brain regions. Thus, like the lesion studies, the neuroimaging studies indicate that the WCST is a sensitive, but not specific, marker of frontal lobe functioning.

Phonemic Verbal Fluency

Origin and History

Verbal fluency is one of the most frequently used measures of executive functioning (Baddeley, 1996; Baldo et al., 2001; Goodglass and Kaplan, 1979; Stuss and Levine, 2002; Warbuton et al., 1996), and is used regularly by approximately 50% of neuropsychologists (Butler et al., 1991). The two types of verbal fluency tasks are phonemic and semantic. Phonemic fluency tasks require participants to say (or write) as many words as possible beginning with a specific letter. Semantic fluency tasks require participants to say (or write) as many words as possible within a certain category (e.g., animals). In general, persons with frontal lobe damage demonstrate impaired phonemic fluency, while their semantic fluency remains relatively intact (Troyer et al., 1998).

Although Feuchtwanger (1923, cited in Zangwill, 1966) reported that persons with frontal lobe damage had a decrease in the production of spontaneous speech, Thurstone (1938) developed a *written* test of verbal fluency called the Word Fluency Test (TWFT), which was the first standardized procedure for the psychometric assessment of word fluency. An oral word fluency task developed by clinical neuropsychologists about 30 years after the development of the TWFT is the most popular phonemic fluency task for brain-damaged populations (Benton

and Hamsher, 1976). The Controlled Oral Word Association test (COWA) of Benton and Hamsher (1976) requires participants to generate as many words orally as possible within 60 seconds beginning with the letters "F," "A", and "S".

Lesion Studies

Sixteen studies examining the effect of various brain lesions on phonemic verbal fluency performance in adult populations are reviewed below (see Table 4). There is one known qualitative review (Reitan and Wolfson, 1994) and one meta-analytic study (Henry and Crawford, 2004) in this area. Our updated review will summarize the data that both support and challenge the sensitivity and specificity of phonemic verbal fluency to frontal lobe lesions.

Ten (out of 10) studies found that persons with frontal lobe lesions produce significantly fewer words than healthy controls, and eight (out of nine) studies indicate that persons with frontal lobe lesions perform worse than persons with non-frontal lobe lesions (see Table 4). In support of these findings, Henry and Crawford (2004) conducted a meta-analysis of 31 studies with 1,791 subjects and found that individuals with focal frontal lesions had larger deficits in phonemic ($r = .52$) verbal fluency as compared to healthy controls. Collectively, these results suggest that phonemic verbal fluency is sensitive to frontal lobe lesions, but does not address whether it is specific to frontal lobe lesions. To address the specificity issue, persons with any other type of brain damage would have to perform as well as healthy controls, and persons with frontal lobe lesions would have to perform significantly worse than all other brain-damaged groups on this task.

Diffuse and frontal lobe lesioned patients show no differences in total number of words produced, yet both do worse than healthy controls (Pendleton et al., 1982). Alzheimer's patients perform as poorly as frontal lobe lesioned patients (Miller, 1984) and persons with both frontal *and* non-frontal left hemisphere lesions perform worse than persons with right hemisphere frontal and non-frontal lesions (Perret, 1974) and healthy controls (Stuss et al., 1998). Thus, these studies do not support the specificity of phonemic verbal fluency tasks to frontal lesions.

Persons with *left* frontal lesions often perform significantly worse than any other brain-damaged group (Baldo et al., 2001; Pendleton et al., 1982; Perret, 1974; Ramier and Hecaen, 1970; Stuss et al., 1998; Troyer et al., 1998), but right frontal (Bornstein, 1986; Miceli et al., 1981; Miller, 1984; Pendleton et al., 1982; Perret, 1974; Ramier and Hecaen, 1970; Troyer et al., 1998) and bilateral frontal

Table 4. Verbal Fluency Lesion Studies for Qualitative Review

First author	Year	Test or Letters Used	Lesion Etiology	Comparison group	Sample Size[a]	Mean Age[b]	Results
Baldo	2001	COWA	Diverse	Normals	22	66.9	Frontal group produced fewer words than the normal group. Patients with left lesions were more impaired than those with right lesions
Benton	1968	COWA	Diverse	None	25	41.5	Patients with left and bilateral frontal lesions produced fewer words than patients with right frontal lesions
Bornstein	1986	COWA[c]	Diverse	Non-frontals	53	36.6	Left and right frontal groups produced fewer words than left and right non-frontal groups. Left frontal and non-frontal groups produced fewer words than right frontal and non-frontal groups
Butler	1993	COWA	Diverse	Normals	34	37.8	Frontal group produced significantly fewer words than the normal group
Crockett	1986	COWA	–	Non-frontals	40	40.6	Frontal group produced significantly fewer words than the non-frontal group
Ferstl	2002	–	Diverse	(1) Normals	20	41.6	Patients with left frontal lesions produced fewer words than normals and patients without left frontal lesions
				(2) Non-frontals	20	42.9	
Janowsky	1989	COWA	Diverse	Normals	18	62.4	Patients with left/ bilateral frontal lesions produced fewer words than normals
Manes	2002	COWA	Diverse	Normals	32	53.4	Frontal group produced fewer words than normals
Miceli	1981	COWA	Diverse	Non-frontals	149	–	Patients with right frontal lesions produced fewer words than patients with right non-frontal lesions. No differences between left hemisphere groups
Miller	1984	COWA	–	Normals[d]	60	53	Left and right frontal groups produced fewer words than normals
Milner	1964	TWFT	–	Non-frontals	18	–	Patients with left frontal lesions produced fewer words than patients with left temporal lobe lesions
Pendleton	1982	TWFT	Diverse	(1) Normals	181	39.8	Frontal group produced fewer words than normal and non-frontal groups. Left frontal group performed worse than right frontal group
				(2) Non-frontals	86	43.6	
Perret	1974	Modified TWFT	Diverse	(1) Normals	76	41.8	Frontal group produced fewer words than normal and non-frontal groups. All left hemisphere patients performed worse than all right hemisphere patients, with the left frontal group performing the worst
				(2) Non-frontals	118	39.5	
Ramier	1970	P, F, L	Diverse	Non-frontals	76	–	Right and left frontal groups produced fewer words than right and left non-frontal groups, with the left frontal group performing the worst
Stuss	1998	COWA	Diverse	(1) Normals	91	53.8	Left frontal and non-frontal groups performed worse than normals. Patients with left DLPFC and/or striatal lesions performed the worst
				(2) Non-frontals	74	52.1	
Troyer	1998	COWA	Diverse	Normals	90	54.4	Frontal group produced fewer words than normals, with the left DLPFC and SMF groups performing the worst

Note. Dashes indicate that no information was provided. COWA = Controlled Oral Word Association test; DLPFC = dorsolateral prefrontal cortex; SMF = superior medial frontal cortex; TWFT = Thurstone's Word Fluency Test.
[a]Combined sample size for frontal and comparison group.
[b]Mean age of frontal and comparison groups.
[c]F, A, S letters are assumed, although it is not explicitly stated which letters were used in the article.
[d]Control participants either had no identifiable neurological disease or had extracranial nervous system pathology (e.g., peripheral nerve damage).

(Benton, 1968; Janowsky et al., 1989) lesions also impair phonemic verbal fluency performance. Although non-frontal and right-sided lesions have been found to impair phonemic verbal fluency, impaired verbal fluency typi-cally is a product of left-sided lateralization of frontal lobe damage (Ramier and Hecaen, 1970).

There are several limitations to the lesion studies in Table 4 that may be related to the lack of specificity

Table 5.   Verbal Fluency Neuroimaging Studies for Qualitative Review

First author	Year	VF Test	Control Test(s)	Imaging Technique(s)	Sample Size	Mean Age	Areas of Increased Activation
Frith	1995	Modified[a]	Repetition task	PET	6	57.2	Left DLPFC, ACC, and thalamus
Frith	1991	Modified[b]	Repetition task	PET	6	25-45[c]	DLPFC and ACC
Parks	1988	Modified[d]	Relaxation	PET	51[e]	53.9	Frontal, temporal, and parietal lobes with greatest activation in bilateral temporal lobes
Paulesu	1997	Modified[f]	Relaxation	fMRI 1.5T	6	28.3	Anterior triangular and posterior opercular portions of the left IFG, left insula, and left dorsomedial and pulvinar thalamic nuclei
Phelps	1997	Modified[g]	Repetition task and antonym task	fMRI 2.1T	6	–	Left IFG and ACC
Warkentin	1997	COWA	Relaxation	rCBF	22	36.1	Broca's area, SMA, and left DLPFC

Note. Dashes indicate that no information was provided. ACC = anterior cingulate cortex; COWA = Controlled Oral Word Association test; DLPFC = dorsolateral prefrontal cortex; IFG = inferior frontal gyrus; rCBF = regional cerebral blood flow; SMA = supplementary motor area; VF = verbal fluency.
[a]Every five seconds an examiner named a letter, and the participant needed to provide a word beginning with that letter. The same letter was repeated 10 times consecutively.
[b]Every two seconds, the participant heard the word "next" and responded with a word beginning with "S" on the first occasion and "F" on the second.
[c]No mean age provided.
[d]Examiner would say one of 20 letters from the alphabet and the participant had 60 seconds to respond with as many words beginning with that letter. The test continued for 30 minutes and, thus, 10 letters had to be repeated.
[e]16 participants were scanned during VF performance; 35 "controls" were scanned in a resting state.
[f]Participants were asked to generate words covertly during the test and then were asked to recall the words after the experiment.
[g]Participants were asked to generate eight words beginning with a letter. This process was repeated with 16 different letters.

of phonemic verbal fluency tasks to frontal lobe lesions. First, several studies did not include appropriate control groups (e.g., Benton, 1968; Stuss et al., 1998). Second, several studies did not indicate the exact localization of the lesions, which may have confounded the "purity" of the groups' compositions (e.g., Butler et al., 1993). Third, some studies used the COWA while others used the TWFT. Differences between written and oral word fluency tasks, as well as the time interval differences between these tasks, may have impacted the findings. Lastly, some of these studies (e.g., Miceli et al., 1981) excluded persons with dysphasia and others provided no information regarding the incidence of dysphasia in the study participants (e.g., Milner, 1964). Reitan and Wolfson (1994) noted "that the incidence and possible effects of dysphasia should be identified when tests that require production or processing of verbal material are used in comparative assessment of brain-damaged groups" (p. 172) due to the fact that the presence of dysphasia could be causing the limited verbal productions, rather than the lesion location per se.

Functional Neuroimaging Studies

A few studies have examined whether the frontal lobes are activated during phonemic verbal fluency performance in healthy adult populations (see Table 5). Stud-ies that used semantic fluency tasks were excluded from this review because semantic fluency remains relatively intact in persons with frontal lobe damage (Troyer et al., 1998).

As seen in Table 5, the results of the phonemic verbal fluency tasks vary widely across studies. Some consistent results, however, have emerged from an examination of the data. These studies found increased activation in the left dorsolateral prefrontal cortex (Frith et al., 1995; Frith et al., 1991; Warkentin and Passant, 1997), anterior cingulate (Frith et al., 1995; Frith et al., 1991; Phelps et al., 1997), and left inferior frontal gyrus (Paulesu et al., 1997; Phelps et al., 1997). The findings of increased activation in frontal areas along with the finding of Parks et al. (1988) of increased overall frontal lobe activation suggest that an intact frontal cortex, especially the left side, is required for phonemic verbal fluency performance. Phonemic verbal fluency also activates a number of non-frontal brain areas, including the thalamus (Frith et al., 1995; Paulesu et al., 1997), parietal lobes (Parks et al., 1988), and temporal lobes (Parks et al., 1988).

These brain-imaging studies suffer from similar methodological shortcomings as those using the WCST. Despite the limitations to the neuroimaging studies, the results suggest that phonemic verbal fluency performance activates a number of frontal and non-frontal brain areas, indicating the sensitivity, but not specificity, of phonemic verbal fluency tasks to frontal lobe functioning.

Stroop Color Word Interference Test

Origin and History

The Stroop test is one of the most extensively studied measures of selective attention (Blenner, 1993; Carter et al., 1995; Goodglass and Kaplan, 1979; Lezak, 1995; MacLeod, 1991; Stuss et al., 2001) used by approximately 50% of neuropsychologists (Butler et al., 1991). The test often consists of three sets of stimuli: (a) color words printed in black ink; (b) color patches or colored X's; and (c) color words printed in incongruous colored ink (e.g., the word "RED" printed in blue ink). The participant must read the color words on the first sheet, the colors on the second sheet, and the color of the ink (i.e., not the words) on the third sheet. In the latter task, the normal tendency to read the words, rather than the color of the ink in which the words are printed, elicits a significant slowing in reaction time (RT) called the "Stroop effect" or the "interference effect." Stroop (1935) found that healthy college students' mean RT increased by 74 percent from naming color patches to naming the incongruous colored ink in which color words were printed.

Nearly 50 years before Stroop (1935) published his seminal paper on attentional interference using the paradigm described above (now known as the "Stroop test"), Cattell (1886) had previously shown that it took participants longer to name colors (and pictures of objects) than it did to name the corresponding words of the colors and objects. He also found that it took participants a shorter amount of time to *recognize* a color than it did a word or letter, but it took longer to *name* the color than the word or letter "because in the case of words and letters the association between the idea and name has taken place so often that the process has become automatic, whereas in the case of colours [sic] and pictures we must by a voluntary effort choose the name" (p. 65). In other words, Cattell argued that there is a distinction between automatic and voluntary attentional processes involved in naming colors and words, that is, attending to the lexical features of words is an automatic process while attending to ink color is not. As MacLeod (1991) noted, Cattell's hypothesis has influenced his contemporaries (e.g., James, 1890; Quantz, 1897) as well as future psychologists (e.g., Stroop, 1935; Posner and Synder, 1975).

There have been over one thousand articles published on the Stroop effect over the past 67 years. A review of the entire Stroop literature is beyond the scope of this paper (interested readers should consult MacLeod, 1991). This section will focus only on how the test came to be used by neuropsychologists for the detection of frontal lobe impairment. It appears that the Stroop came to be

used by neuropsychologists as a "frontal lobe test" based on the results of *one* study which found that persons with left frontal lobe lesions displayed significantly longer interference trial RTs than persons with non-frontal lobe lesions (Perret, 1974). Only five additional studies have been located that examine the role of frontal lobe lesions on Stroop performance since that time, even though the test is widely used as a measure of frontal lobe functioning (Butler et al., 1991; Stuss et al., 2001).

Lesion Studies

The last qualitative review of studies employing the Stroop test (MacLeod, 1991) did not address the question of sensitivity and specificity of the task to frontal lobe lesions. Only two studies have found that persons with frontal lobe lesions perform worse than healthy controls (Stuss et al., 2001; Vendrell et al., 1995). Another study found that bilateral medial frontal lesions increase "susceptibility" to the Stroop effect (Holst and Vilkki, 1988, p. 80). In sum, only certain areas of the frontal lobes appear to underlie Stroop performance, namely lateral and superior medial, not orbitofrontal (see Table 6).

Two other studies have found differences between frontal and non-frontal groups. Persons with left frontal lobe lesions perform significantly worse on the incongruent color naming condition than persons with right frontal, right non-frontal, and left non-frontal damage, and persons with left-sided frontal lobe lesions produce significantly slower performance on all three conditions than persons with non-frontal lobe lesions (Perret, 1974; Stuss et al., 2001). Conversely, Blenner (1993) found no differences between groups with frontal and temporal lobe lesions, although the combined lesioned group performed worse than a normal group on all three conditions.

Collectively, the results of the Stroop lesion studies are less consistent than the results of the WCST and phonemic verbal fluency lesion studies. The Stroop test is sensitive to lateral and superior medial lesions of the frontal lobes, but it is not specific to overall frontal lobe functioning.

Functional Neuroimaging Studies

In general, the hemodynamic brain imaging data both support and challenge the sensitivity and specificity of the Stroop test to frontal lobe functioning (see Table 7). Despite a number of differences between the studies, there are some consistent findings. A general conclusion to be drawn is that increased activation in the anterior cingulate cortex is a critical brain region for selective attention (e.g.,

Table 6. Stroop Lesion Studies for Qualitative Review

First author	Year	Test	Lesion Etiology	Comparison group	Sample Size[a]	Mean Age[b]	Results
Blenner	1993	Standard	Diverse	(1) Normals	36	40.8	The combined frontal and temporal lobe lesioned groups performed worse than the normal group on all three conditions. There were no differences between the frontal and temporal lobe lesioned groups
				(2) Non-frontals	35	42.3	
Holst	1988	–	–	–	–	–	Persons with bilateral medial frontal lesions demonstrated "increased susceptibility" to the Stroop effect
Perret	1974	Modified	Diverse	Non-frontals	118	39.5	Persons with left frontal lesions performed worse than persons with right frontal, right non-frontal and left non-frontal lesions on the interference condition
Stuss	1981	Standard	Bilateral OFC leucotomies due to schizophrenia	Normals	26	–	There were no differences between the frontal and normal groups
Stuss	2001	Standard	Diverse	(1) Normals	63	52.9	Persons with frontal lesions, particularly left-sided, are slower than the normal and non-frontal groups on all three conditions
				(2) Non-frontals	51	49.9	
Vendrell	1995	Modified computerized	Diverse	Normals	64	40.5	The frontal group made significantly more errors than the normal group. There were no RT differences between groups

Note. Dashes indicate that no information was provided. OFC = orbitofrontal cortex; RT = reaction time.
[a]Combined sample size for frontal and comparison group.
[b]Mean age of frontal and comparison groups.

Bench et al., 1993; Carter et al., 1995; Pardo et al., 1990; Posner and Dehaene, 1994; Posner and Petersen, 1990). The one study that did not find increased activation in the anterior cingulate cortex differed from the others in experimental design (Banich et al., 2000). Despite the one negative finding, it appears that the anterior cingulate cortex does play an important role in Stroop performance. In fact, Peterson et al. (1999) stated that it acts as a "*central executor* [emphasis added] that coordinates and integrates the task-oriented sensory, receptive and expressive language, alerting, working memory, response selection, motor planning, and motor response processes within the CNS [central nervous system]" (p. 1253).

In addition to the prominent role of the anterior cingulate during Stroop performance, several studies also indicate that the Stroop test activates the middle frontal gyrus (Banich et al., 2000; Bush et al., 1998; Leung et al., 2000; Taylor et al., 1997), parietal lobe regions (Brown et al., 1999; Bush et al., 1998; Carter et al., 1995; Leung et al., 2000; Peterson et al., 1999; Taylor et al., 1997), motor areas (Bush et al., 1998; Pardo et al., 1990; Peterson et al., 1999), and temporal lobe regions (Bush et al., 1998; Carter et al., 1995; Leung et al., 2000; Pardo et al., 1990). Furthermore, several studies demonstrate that the Stroop activates a distributed neural network of brain regions

(Brown et al., 1999; Bush et al., 1998; Carter et al., 1995; Leung et al., 2000; Pardo et al., 1990; Peterson et al., 1999; Taylor et al., 1997). Moreover, the finding that a task as complex as the Stroop activates a large number of brain areas is consistent with parallel distributed processing models of the Stroop effect (e.g., Cohen et al., 1990). Despite a number of methodological limitations to the Stroop neuroimaging studies, the results suggest that Stroop performance activates a distributed neuronal network of frontal and non-frontal brain regions.

SUMMARY OF QUALITATIVE REVIEW

A qualitative review of three popular executive function measures (WCST, phonemic verbal fluency, and Stroop) suggests that these measures are sensitive, but not specific, indicators of frontal lobe damage. Typically, persons with frontal lobe lesions perform more poorly than healthy controls on these tests, although several studies indicate that patients perform within normal limits (e.g., Ahola et al., 1996; Damasio, 1994; Eslinger and Damasio, 1985; Heck and Bryer, 1986; Shallice and Burgess, 1991b). Moreover, persons with frontal lobe lesions usually perform worse than persons with non-frontal lobe

Table 7.   Stroop Neuroimaging Studies for Qualitative Review

First author	Year	Stroop Test	Control Test(s)	Imaging Technique(s)	Sample Size	Mean Age	Areas of Increased Activation
Banich	2000	Modified computerized	Neutral words	fMRI 1.5-T	10	–	Bilateral inferior and middle frontal gyrus for incongruent stimuli
Bench	1993	Modified computerized	Colored crosses	PET O[15]	6	21-34[a]	Right ACC and right frontal polar region
Brown	1999	Modified/covert	Colored squares and color-neutral words	fMRI 1.5-T	8	< 55[a]	Bilateral ACC, right parietal cortex, and left pars opercularis compared to both control conditions
Bush	1998	Computerized "Counting Stroop"[b]	Counting neutral words[b]	fMRI 1.5-T	9	24.2	ACC, middle frontal gyri, premotor and primary motor cortex, inferior temporal gyrus, and superior parietal lobule
Carter	1995	Modified computerized	(1) Neutral animal words (2) Congruent color words	PET O[15]	9	34.3[c]	(1) Bilateral ACC, right IPL, right superior temporal gyrus, and left precentral gyrus (2) Right ACC, right frontal polar cortex, bilateral IPL, bilateral insula, and left lingual gyrus
Leung	2000	Modified	Congruent color words	fMRI 1.5-T	13	28	ACC, insula, inferior frontal, middle frontal, parietal, and mid-temporal regions
Pardo	1990	Modified	Congruent color words	PET O[15]	8	26	Right ACC, bilateral peristriate, left premotor and left postcentral foci, left putamen, SMA, inferior ACC, and right temporal lobe
Peterson	1999	Modified	Congruent color words	fMRI 1.5-T	34	29.3	Right ACC, DLPFC, left midsensorimotor cortex, left inferior motor strip, bilateral IPL, left peristriate cortex, caudate, and thalamic nuclei
Taylor	1997	Modified	Neutral words	PET O[15]	12	–	ACC, right middle frontal gyrus, left inferior frontal gyrus, left parietal region, and left insula

Note. Dashes indicate that no information was provided. ACC = anterior cingulate cortex; DLPFC = dorsolateral prefrontal cortex; IPL = inferior parietal lobule; SMA = supplementary motor area.
[a]No mean age provided.
[b]Participants had to determine the number (ranging from 1–4) of number words ("one," "two," "three," or "four") on a screen for the interference trial and the number (1–4) of animal words ("dog," "cat," "bird," and "mouse") for the neutral trial.
[c]Mean age was provided for 15 participants, although the study only reported data for nine of these participants (and their mean age was not reported).

lesions, but some studies found that persons with non-frontal or diffuse brain lesions do as poorly as frontal lobe lesion patients (e.g., Anderson et al., 1991; Axelrod et al., 1996; Crockett et al., 1986; Grafman et al., 1990; Heaton, 1981; Robinson et al., 1980). Thus, overall there have been inconsistent findings regarding the sensitivity and specificity of these three executive function measures to lesions in the frontal lobes, indicating that these tasks should not be used as "frontal lobe tests" per se, but rather as tests of specific executive functions (e.g., problem-solving, cognitive fluency).

In addition, functional neuroimaging studies using these three executive function measures demonstrated that the tests activate a distributed neural network of frontal and non-frontal brain regions. In other words, the brain imaging data do not implicate the frontal lobes as the *only* brain region responsible for executive functions. It is not surprising, however, that multiple brain areas are involved in cognitive processes as complex as the executive functions tapped by these measures, including shift-

ing and maintaining cognitive set, inhibition of prepotent responses, selective attention, and planning. Moreover, the frontal lobe regions have multiple connections with various other cortical, subcortical, and brain stem sites and, thus, the frontal lobes should "be conceived as one aspect of an executive system involving many structures of the central nervous system" (Duffy and Campbell, 2001, p. 116). Commensurate with the results of the lesion studies, the neuroimaging data support the sensitivity, but not specificity, of these three executive function measures to frontal lobe functioning.

RATIONALE FOR QUANTITATIVE REVIEW

In order to examine further the relationship between these three executive function measures and frontal lobe damage, lesion studies were examined quantitatively. The meta-analytic approach increases statistical power, permits the estimation of a population effect size, and allows

an examination of variables that may be moderating the relationship between lesion location and performance on the executive function measures.

As a rationale for the quantitative analyses, the qualitative review by itself relies only on the statistical significance of the original findings, which may be unreliable due to low statistical power (Schmidt, 1992). Thus, studies with small sample sizes may produce results that are "not significant" despite the presence of large effect sizes. The qualitative review was performed in order to determine potential moderator variables, inclusion criteria, and coding strategies. The procedures and results of the meta-analysis are presented next.

METHOD

Search Strategy

Several strategies were employed to identify studies for inclusion in the meta-analysis. First, searches of computerized databases, including PsycINFO and MEDLINE, were conducted using keywords such as "executive function," "frontal lobe," "Wisconsin card sort," "Stroop," "verbal fluency," and "word fluency," as well as variants on these terms. After collecting all available published articles and abstracted studies, their reference sections were scanned in order to locate additional articles that may have been missed in the previous searches. Lastly, authors were contacted to inquire whether additional research had been conducted that would have been overlooked by the previous search methods.

Inclusion Criteria

Studies needed to satisfy the following criteria to be included in the meta-analysis:

1. The sample consisted of adult participants only (i.e., the mean or median age of the sample was equal to or above 18 years).
2. The study did not consist solely of persons with "suspected" frontal lobe damage, such as psychiatric populations (e.g., persons with schizophrenia) or demented populations (e.g., persons with Alzheimer's disease or frontal lobe dementia).
3. The study either had a healthy control group or a non-frontal lobe lesioned control group.
4. The study included verification of frontal lobe damage either through a brain imaging technique (e.g., CT scan or MRI) or through surgical reports.

5. The study employed the standard version of the WCST (Milner, 1963; Heaton, 1981; Heaton et al., 1993), the COWA test of phonemic verbal fluency (Benton and Hamsher, 1976), or the standard version of the Stroop (Stroop, 1935; Stuss et al., 2001).
6. The study reported the following scores, depending on which test was administered: (a) number of perseverative errors on the WCST because "perseverative errors are regarded as the main signs of frontal dysfunction" (Barcelo and Knight, 2002, p. 349); (b) total number of "FAS" words generated on the COWA test; and (c) interference trial RT for the Stroop.
7. Adequate data (i.e., means and standard deviations, *t*-values, *F*-values, or *p*-values) were provided for calculation of effect sizes.

Study Sample

Initially, there were 52 lesion studies selected for the meta-analysis. A total of 27 studies were included after evaluation of whether each study satisfied the inclusion criteria. These 27 studies tested 1,992 participants with sample sizes ranging from 18 to 415. The mean age of the samples ranged from 26.5 to 66.9 (with a mean age of the entire sample of 45.33, SD = 10.14).

Examination of Moderator Variables

The qualitative review identified potential moderators of the relationship between executive function measures and the frontal lobes, including type of test, comparison group, and age. Each variable is discussed briefly below.

Type of Test

Several studies have found relatively low intercorrelations ($r < .40$) among executive function tests (e.g., Cockburn, 1995; Crockett et al., 1986; Duncan et al., 1997; Humes et al., 1997; Miyake et al., 2000; Welsh et al., 1999), indicating the possibility that each executive function test measures something unique (Duncan et al., 1997; Rabbitt, 1997; Vandierendonck, 2000). Despite the low correlations among executive function measures, it is important to determine whether there are higher correlations among executive function measures than between executive function measures and other measures *not* hypothesized to tap executive functions (e.g., recognition

memory). One study found that the median correlation between four executive function tests and five non-executive function tests was .29, while the median correlation among the four executive function measures was only .26 (Duncan et al., 1997), indicating that the executive function measures are no more related to one another than they are to other tasks. Furthermore, Miyake et al. (2000) found that correlations among executive function measures were higher when they were thought to tap the same underlying cognitive process while correlations were lower when they were thought to tap different cognitive processes, suggesting that executive function measures show signs of both convergent and discriminant validity.

Generally, the results seem to suggest that type of executive function measure may moderate the relationship between frontal lobe functioning and test performance. Type of test was coded as a categorical variable.

Comparison Group

The qualitative review of lesion studies indicated that most studies using a healthy comparison group found that the executive function measures were sensitive to frontal lobe lesions, while many of the studies using a non-frontal lobe lesioned comparison group did not find differences between groups. Thus, type of comparison group was explored as a possible moderator variable of the relationship between executive function measures and integrity of the frontal lobes. It was hypothesized that studies comparing persons with lesions in the frontal lobes to lesions in posterior brain regions would yield smaller effect sizes than studies comparing persons with lesions in the frontal lobes to healthy control participants. Comparison group was coded as a categorical variable.

Age

A moderating effect of age was predicted based on several studies that found significant correlations between executive function measures and age of participants (Anderson et al., 1991; Axelrod and Henry, 1992; Berg, 1948; Bryan and Luszcz, 2000; Crockett et al., 1986; Grafman and Litvan, 1999; Heaton et al., 1993; Little and Hartley, 2000; Nagahama et al., 1997; Nelson, 1976; Pendleton et al., 1982; Rhodes, 2004; van den Broek et al., 1993; Wang, 1987; Zelazo et al., 1997). For instance, Little and Hartley (2000) found that the interference effect on the Stroop test was greater for older adults than for younger adults. Additionally, Heaton et al. (1993) found a quadratic relationship between WCST performance and age. Scores improved during childhood (ages 6-1/2 to 19)

then stabilized during adulthood (ages 20 to 50) and finally declined at an accelerated rate during late adulthood (ages 60 to 90). Moreover, Malloy and Richardson (2001) found that the frontal lobes do not fully mature until adolescence and that there is a greater loss of neurons during normal aging in the frontal lobes than in posterior regions. In a meta-analysis conducted by Rhodes (2004) robust age effects were found on the number of categories achieved and the number of perseverative errors committed. These effects were moderated by education and test version. Thus, age appears to impact executive functions on a psychological level (i.e., executive function test performance) and a neuroanatomical level (i.e., frontal lobe development and degeneration). Mean age was coded as a continuous variable for each study in which this information was provided.

RESULTS

Primary Analysis

Details regarding the statistical procedures employed in this meta-analysis are reported in Appendix A. The effect sizes were averaged across executive function measures to produce an unweighted grand mean effect size of large magnitude (Cohen, 1992), $d = -.83$, with a 95 percent confidence interval of -1.08 to $-.58$. The effect sizes were then weighted by their respective sample sizes, yielding a weighted grand mean effect size of moderate magnitude (Cohen, 1992), $d_+ = -.78$, with a 95 percent confidence interval of $-.88$ to $-.68$. The fail-safe N statistic (Orwin, 1983) indicated that 78 additional studies with null results (i.e., effect sizes equal to zero) would be necessary to reduce the weighted mean effect size to a non-significant level.

A test of homogeneity using the weighted effect sizes was significant, $Q(26) = 99.11$, $p < .0001$, indicating that the effect sizes come from two or more populations. Thus, there are potential moderator variables that may be impacting the relationship between executive function measures and frontal lobe functioning.

Analysis of Moderator Variables

Details regarding the statistical procedures for the moderator analyses are reported in Appendix B. Due to the finding of significant heterogeneity of the effect sizes, several factors were examined as potential moderator variables.
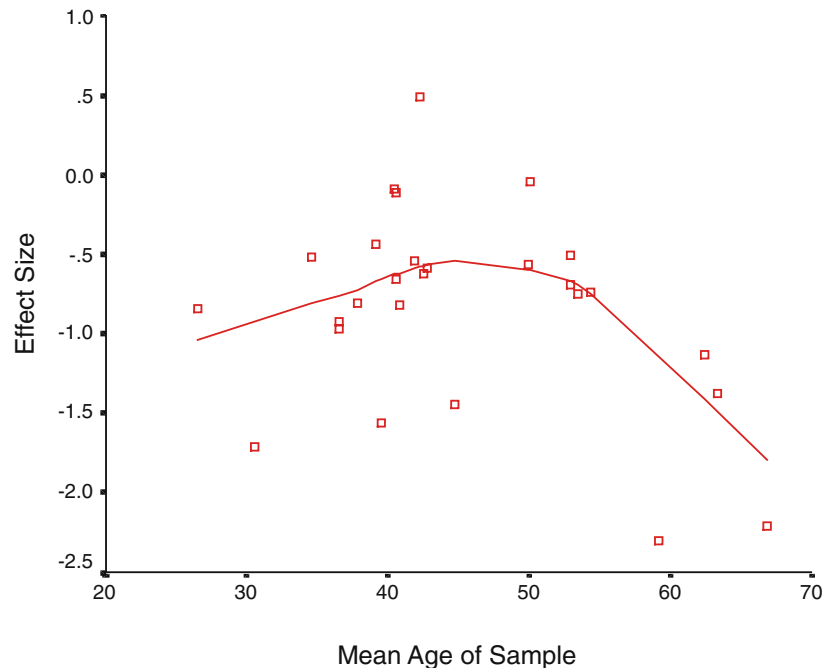
Fig. 1. Scatterplot of relationship between unweighted effect sizes and mean age of samples in the meta-analysis.

Type of Test

Type of test was a significant moderator of the relation between executive function test performance and integrity of the frontal lobes, $Q_b = 97.85$, $p < .0001$. However, there was significant within-group heterogeneity at each level of the variable. Follow-up contrasts revealed significant differences between each comparison (i.e., WCST v. verbal fluency, WCST v. Stroop, and verbal fluency vs. Stroop). Overall, the results indicated greater sensitivity to frontal lobe damage for the WCST ($d = -0.97$) and phonemic verbal fluency ($d = -0.80$) than for the Stroop test ($d = -0.30$).

Comparison Group

Type of control group also was a significant moderator, $Q_b = 109.47$, $p < .0001$. An analysis of the results supported the hypothesis that studies comparing a frontal group to a non-frontal group yield a smaller grand mean effect size than studies comparing a frontal group to a healthy control group, $d = -0.57$ and $d = -1.05$, respectively.

Age

A scatterplot of effect sizes (see Fig. 1) suggests that effect sizes are larger for younger and older adult

groups, but are smaller (i.e., closer to zero) for middle-aged adults. Regression analysis confirms that there is a quadratic relationship between age and effect size, $\beta = -4.422$, $t = -3.306$, $p = .003$, and the LOWESS line fit to the scatterplot of age and effect size is arched, suggesting that effect size is curvilinearly related to age. Although somewhat counterintuitive, frontal lobe damage in younger and older adults may be more detrimental due to developmental and degenerative processes and, thus, it may cause greater impairment in these individuals on executive function measures (leading to the larger observed effect sizes).

DISCUSSION

Summary of Qualitative and Quantitative Reviews

The results of the qualitative and quantitative reviews of the WCST, phonemic verbal fluency tasks, and Stroop test suggest that these measures are sensitive (not specific) indicators of frontal lobe damage, but there are inconsistencies in the results. The WCST has the strongest and most consistent relationship to the frontal lobes, phonemic verbal fluency has the second strongest relationship, and the Stroop test has a less consistent and weaker relationship than the WCST and phonemic verbal fluency.

It may be that the three tests are tapping different underlying cognitive process and, therefore, the construct of executive function may not be unitary. Several authors have suggested that the executive function construct is "fractionable" (Baddeley, 1996; Bryan and Luszcz, 2000; Burgess, 1997; Burgess et al., 1998; Denckla, 1994; Duke and Kaszniak, 2000; Duncan et al., 1995; Miyake et al., 2000; Owen et al., 1995; Robbins, 1998; Shallice, 1988; Shallice and Burgess, 1991a; Stuss and Levine, 2002; Stuss et al., 1995; Vandierendonck, 2000; Zelazo et al., 1997) and that "there is no frontal homunculus, no unitary executive function" (Stuss and Alexander, 2000, p. 291). That executive functioning may involve participation of diffuse areas of the brain and that different tests of this function appear to be tapping varied cognitive processes does not mean that the construct lacks unity. On the contrary, if one assumes that "executive" is distinct from "non-executive" function, the implication would be of greater overall coordination of brain activity as a necessary condition for higher-level cognitive processing. Thus, tasks designed to tap executive function naturally would be sensitive to frontal lobe damage, but not specific to focal frontal lesions because executive functioning requires participation and coordination of activity among diffuse anatomical and functional brain areas. It simply may be that the frontal lobes participate to a greater extent than other areas of the brain in functions considered to be "executive." Without input, however, from other cortical and subcortical areas executive functioning would be compromised. Therefore, it may be more worthwhile to conceptualize executive functions as a "macroconstruct" in which multiple executive function subprocesses work in conjunction to solve complex problems and execute complicated decisions (Zelazo et al., 1997).

Alternatively, there are several additional potential moderators that were not explored due to insufficient data, including general intelligence, exact localization of lesion, and time since injury. Differences in task administration and scoring procedures as well as poor psychometric properties also may have influenced the findings. Surprisingly, there are only a handful of studies examining the reliability and validity of executive function measures, and these studies usually find low reliability and inadequate validity (Bowden et al., 1998; Humes et al., 1997; Kafer and Hunter, 1997; Miyake et al., 2000; Schnirman et al., 1998; Stuss and Alexander, 2000; Vandierendonck, 2000). Parks et al. (1992) suggested that parallel distributed processing (PDP) models of executive function tasks may help to circumvent some of these problems with reliability and validity. It is important, however, to discuss briefly what PDP modeling is before explaining how it addresses the reliability and validity of executive function measures.

The theory behind PDP modeling developed out of early psychological associationist ideas such as Hebbian learning principles (Hebb, 1949). Subsequently, investigators produced empirical physiological data that were compatible with the psychological theories (Parks et al., 1992). Following the advent of "supercomputers" that could integrate the physiological and psychological data, PDP modeling was developed. In general, PDP "refers to a complex mathematical methodology used to model neuropsychological functions and other neurobehavioral tasks" (Parks et al., 1992, p. 215). PDP methodology addresses reliability insofar as each network model is internally consistent due to neuroanatomical and biological constraints. In terms of validity, computer simulations of experimental data (e.g., WCST scores) have replicated actual neuropsychological test performance in persons with frontal lobe damage and healthy controls (Levine and Prueitt, 1989).

One might logically ask if it is necessary to require frontal lobe involvement in order to qualify a test as a measure of executive functioning. Some circularity of reasoning emerges in the argument that the construct validity of executive function tests should be established on the basis of their sensitivity and specificity to frontal lobe damage. Rather, it should be established on their ability "to assess the theoretical concept of executive function and the group of cognitive processes it entails" (Bryan and Luszcz, 2000, p. 41). None of the tasks reviewed measure the entire executive function domain because it is not a unitary construct (e.g., Miyake et al., 2000; Vandierendonck, 2000). Moreover, executive functions depend on the integrity of other "lower-level" aspects of cognition that were not specifically assessed in the majority of these studies, including visual-spatial perception, visual and auditory attention, and short- and/or long-term memory (Phillips, 1997). In other words, people may display impairments on these tasks due to a deficit in one of the "lower-level" cognitive processes that underlie the target executive function, rather than due to frontal lobe dysfunction.

In summary, the use of executive function tests as "frontal lobe indicators" is not supported by the data reviewed (i.e., the articles failed to demonstrate the *specificity* of these measures to frontal lobe functioning). Discussing the validity of these tests solely in terms of the frontal lobes, however, not only confounds psychology with anatomy but it also ignores the importance of linking the neuropsychological construct of executive functions to behaviors that are both measurable and important in the real world. Despite the seemingly paradigmatic shift within psychology where the study of behavior has become the study of the brain (i.e., many psychologists now are studying which brain regions underlie certain

behaviors rather than studying the behaviors themselves), it is important to ground the executive function construct in the measurement of observable behaviors that have real-world significance. There is a need for more ecologically valid executive function measures (Burgess et al., 1998; Cripe, 1996; Ready et al., 2001; Sbordone, 1996; Wilson, 1993).

Several authors have begun to examine the executive function construct in this manner and they have developed new tasks and measurement tools for this endeavor (e.g., Bechara et al., 1994; Shallice and Burgess, 1991a). Ironically, many of these procedures originated from the lack of traditional executive function measures to detect impairment in persons with frontal lobe lesions (e.g., Eslinger and Damasio, 1985; Shallice and Burgess, 1991b). Nevertheless, the development of these procedures represents a movement away from the "strict localizationist approach" of clinical neuropsychology (Duffy and Campbell, 2001, p. 113), where "psychology and anatomy are inseparable" (Tranel et al., 1994, p. 126) to a more integrative approach that incorporates the behavioral, theoretical, cognitive, and neuroanatomical approaches. This movement has resulted in the development of "alternative" executive function measures.

Alternative Executive Function Measures

Several researchers (e.g., Eslinger and Damasio, 1985; Shallice and Burgess, 1991b) have demonstrated executive function impairments in persons who performed within normal limits to exceptionally well on standard neuropsychological executive function measures (e.g., WCST) and standard IQ tests (e.g., Wechsler Adult Intelligence Scale). The executive function deficits manifested themselves only in complex "real-life" situations constructed by the examiners, such as shopping tasks. One of these tasks is called the Multiple Errands Test (MET) and it requires participants to buy various grocery items on a shopping list with money given to them by the examiner (Burgess, 2000). They also are given a written copy of instructions asking them to find out specific information, be at a particular location at a certain time, and follow a number of rules such as "you must not enter a shop other than to buy something" (Burgess, 2000, p. 281). Another alternative executive function measure is the Cognitive Estimates Test (CET; Shallice and Evans, 1978) which requires participants to provide a reasonable estimate to a series of ten questions to which they are unlikely to know the answer (e.g., "What is the length of an average man's spine?"). An increasingly popular, alternative executive function measure is the gam-

bling task (Bechara et al., 1994), which measures real-life decision-making skills and sensitivity to future consequences. The gambling task requires participants to choose a card from one of four decks that have different monetary rewards and punishments of which the participants are unaware, and it measures the ability to estimate which decks are risky and which are more profitable over time.

Impaired performance on these alternative executive function measures are said to reflect a "dysexecutive syndrome" (Burgess et al., 1998) rather than a "frontal lobe syndrome" (Stuss and Benson, 1984). The change in terminology not only represents a movement away from the linkage of psychology and anatomy but also a movement towards more ecologically valid indicators of executive functions. In other words, persons with a dysexecutive syndrome have difficulties with decision-making, risk-taking, and problem-solving that are not measured adequately by the classic neuropsychological executive function measures (Damasio, 1994). These difficulties significantly impair their ability to work and/or attend school and function well interpersonally (Grafman and Litvan, 1999).

SUMMARY AND FUTURE DIRECTIONS

There has been a long-standing tradition within clinical neuropsychology to link the "highest cognitive functions" such as planning, organization, decision-making, problem-solving, and logical analysis with the largest and most enigmatic brain region, the frontal lobes (Luria, 1966; Reitan and Wolfson, 1994). Before systematic studies were carried out to illuminate the functions of the frontal lobes, "higher-level" processes were attributed to the anterior brain regions because neurological studies already had mapped the majority of "lower-level" functions onto posterior brain areas (Reitan and Wolfson, 1994).

In our review, we found inconsistent support for the historical association between executive functions and the fontal lobes. Rather, the results indicated the sensitivity, but not specificity, of these measures to frontal lobe functioning. In other words, both frontal and non-frontal brain regions are necessary for intact executive functions. One can ask why is sensitivity fairly robust and reliable among commonly used tests of executive function, yet specificity is modest at best? The answer may reside with the notion that executive function is a "macroconstruct" (Zelazo et al., 1997), that is, multiple executive function subprocesses (e.g., working memory, inhibition, and selective attention) work in conjunction to solve complex

problems and execute complicated decisions. Thus, participation of the frontal lobes in virtually any "executive process" is probably a necessary, but largely insufficient, requirement.

In the past decade, there has been a growing interest in studying executive functions in both normal and disordered populations. It has been found that persons with executive function deficits are significantly impaired in their ability to work, attend school, and function well interpersonally (Damasio, 1994; Grafman and Litvan, 1999). For instance, a study by Bayless et al. (1989) found that low scores on the Tinker Toy Test (an alternative executive function measure of planning, goal formulation and execution) were strongly predictive of unemployment. In addition, scores on the Behavioral Assessment of Vocational Skills test (a newer and more ecologically valid executive function measure), as compared to more classic neuropsychological tests (e.g., Trails A & B and the Wechsler Adult Intelligence Scale-Revised), were found to be the only significant predictor of vocational performance (Butler et al., 1993). Furthermore, the Behavioral Assessment of the Dysexecutive Syndrome (a series of six "real-life" tests hypothesized to cause difficulties in persons with executive function deficits) was a better predictor of executive functions in real-world situations than the WCST (Wilson, 1993). Future research should be devoted to the development of ecologically valid executive function measures and more emphasis should be placed on the *remediation* of executive function deficits considering their often profound negative impact on social and occupational functioning. Investigators also should conduct additional studies examining the underlying cognitive subprocesses of the executive function construct. Finally, clinical neuropsychologists may consider abandoning the conceptualization of executive functions in terms of the frontal lobes in favor of a more integrative approach that incorporates behavioral, theoretical, cognitive, and neuroanatomical approaches.

APPENDIX A

Formulas and Procedures for Primary Meta-Analysis

Preliminary Issues

All computations described below were carried out using the following three computer programs – DSTAT (Johnson, 1989), Meta-Analysis Programs (Schwarzer, 1989), and Statistical Package for the Social Sciences (SPSS) – and by hand using formulas cited in Cooper and

Hedges (1994), Hedges and Olkin (1985), Orwin (1983), and Rosenthal (1991).

Calculation of Effect Sizes

The unbiased estimator *d* was chosen as the effect size estimator for the meta-analysis in this paper. Although there are a variety of effect size indicators from which to choose (Glass, 1976; Rosenthal, 1991), the decision to use *d* was based on theoretical and practical matters. First, *d* was selected because the data to be examined were represented as differences between groups (usually means and standard deviations). Estimators in the *d* family are better suited than estimators in the *r* family for studying the strength and direction of mean differences between groups (Hedges and Olkin, 1985). Second, the unbiased estimator *d* adjusts for bias, that is, it adjusts each effect size to control for a standard discrepancy between the sample effect size and the population effect size (Hedges and Olkin, 1985). Third, with respect to more practical concerns, the majority of studies reported means and standard deviations, which reduced the need to transform original findings to another effect size index such as *r*. Lastly, for the present review, negative effect sizes indicate that persons with frontal damage performed worse than the control group (i.e., either healthy controls or persons with non-frontal damage). For instance, a negative effect size on the COWA indicates that the frontal group produced fewer words than the control group. For those samples in which a lower score reflected better performance (e.g., fewer perseverative errors on the WCST), the sign of the effect size was reversed.

When means and standard deviations were available, the following formula was used to calculate *d* (Rosenthal, 1991):

$$d = M_1 - M_2/\sigma_{\text{pooled}} \quad \text{(A.1)}$$

where M_1 and M_2 are means for groups one and two, respectively, and, σ_{pooled} (the pooled within-group standard deviation) was computed as follows:

$$\sigma_{\text{pooled}} = \sqrt{\frac{N_1 SD_1^2 + N_2 SD_2^2}{N_1 + N_2 - 2} + \left(\frac{1}{N_1} + \frac{1}{N_2}\right)} \quad \text{(A.2)}$$

When means and standard deviations were not available, *t*-values, *F*-values, or *p*-values were used to calculate *d* using DSTAT (Johnson, 1989) or Meta-Analysis Programs (Schwarzer, 1989).

Nonindependence

Several studies in this analysis employed more than one executive function measure (e.g., WCST *and* verbal fluency). In addition, the same researcher or team of researchers often conducted more than one study examining the relationship between executive function measures and the frontal lobes. While including these studies would violate the assumption of independent samples (Hedges and Olkin, 1985), excluding them would result in too few studies to conduct a meta-analysis with adequate statistical power. Thus, four possible approaches to dealing with this problem (Mullen, 1989) were considered: (a) use each effect size as if it came from an independent sample; (b) use the results from the best executive function measure; (c) conduct individual meta-analyses for each executive function measure; or (d) average the effect sizes of the different executive function measures within each study to form one estimate. The last three approaches were discarded for the following three reasons, respectively: (a) there are no criteria available for determining which executive function measure is the "best"; (b) there are too few studies to conduct individual meta-analyses for each executive function measure; and (c) averaging the effect sizes of the different executive function measures within each study would lose information about the uniqueness of each measure and its possible moderating effect on the relationship between executive functions and the frontal lobes. Thus, the first approach was chosen (i.e., each effect size was treated as if it came from an independent sample).

COMBINING EFFECT SIZES

The individual effect sizes from the 27 studies listed in Table 8 were averaged to form an unweighted grand mean estimate of the relationship between frontal lobe lesions and performance on the WCST, phonemic verbal fluency, and the Stroop test using the following formula:

$$\overline{d} = \frac{\sum_{i=1}^{k} d_i}{k} \tag{A.3}$$

where k is the number of effect sizes combined and d_i is the aggregated effect size from the ith study.

Next, effect sizes were combined after weighting each effect size by its sample size:

$$d_+ = \frac{\sum_{i=1}^{k} \frac{d_i}{\hat{\sigma}^2 d_i}}{\sum_{i=1}^{k} \frac{1}{\hat{\sigma}^2 d_i}} \tag{A.4}$$

where the variance of d is defined as:

$$\hat{\sigma}^2_{d_i} = \frac{n_F + n_C}{n_F n_C} + \frac{d_i^2}{2(n_F + n_C)} \tag{A.5}$$

where n_F is the sample size for the focal group and n_C is the sample size for the control group. Weighting the studies by sample size allowed more emphasis to be placed on studies with larger samples, thereby producing more precise effect size estimates (Hedges and Olkin, 1985).

For purposes of interpretation, the strength of obtained effect sizes was evaluated according to criteria outlined in Cohen (1992), that is, a d of .20 indicates a small effect, .50 indicates a moderate effect, and .80 indicates a large effect.

TESTS OF HOMOGENEITY OF EFFECT SIZES

The Q statistic outlined in formula 6 indicates whether the amount of variance in the 27 studies used to obtain an estimate of the population effect is greater than what would be expected based upon sampling error alone (Hedges and Olkin, 1985).

$$Q = \sum_{i=1}^{k} \frac{d_i^2}{\hat{\sigma}^2 d_i} - \frac{\left(\sum_{i=1}^{k} \frac{d_i^2}{\hat{\sigma}^2 d_i}\right)^2}{\sum_{i=1}^{k} \frac{1}{\hat{\sigma}^2 d_i}} \tag{A.6}$$

ESTIMATION OF THE "FAIL-SAFE N"

There is a risk that the studies sampled for this meta-analysis only comprise a subset of the existing research on executive function measures and frontal lobe functioning due to the problem that journals often accept for publication only those studies with "significant" results. Although there were studies included in this meta-analysis that did not find significant differences between groups, the 27 studies in this quantitative review may not adequately represent the entire population of studies on this topic. Thus, a "fail-safe N" statistic was computed as an estimate of the number of additional studies with null results (i.e., effect sizes equal to zero) that would be needed to reduce the weighted combined mean effect size to nonsignificance (Orwin, 1983). A fail-safe N was calculated using the following formula:

$$N_{\text{fail-safe}} = \frac{N_{\text{total}}(d_+ - d_{\text{crit}})}{d_{\text{crit}}} \tag{A.7}$$

Table 8. Lesion Studies Included in Meta-Analysis

First author	Year	Test	Comparison group	Sample Size[a]	Mean Age[b]	d	95% Confidence Interval
Anderson	1991	WCST	Non-frontals	73	50.1	−0.04	−0.53/+0.45
Baldo	2001	COWA	Normals	22	66.9	−2.21	−3.27/−1.15
Blenner	1993	Stroop	(1) Normals	36	40.8	−0.81	−1.49/−0.14
			(2) Non-frontals	35	42.3	+0.49	−0.19/+1.16
Bornstein	1986	WCST	Non-frontals	53	36.6	−0.98	−1.55/−0.40
Bornstein	1986	COWA	Non-frontals	53	36.6	−0.92	−1.49/−0.35
Butler	1993	COWA	Normals	34	37.8	−0.81	−1.51/−0.11
Crockett	1986	WCST	Non-frontals	40	40.6	−0.11	−0.73/+0.53
Crockett	1986	COWA	Non-frontals	40	40.6	−0.66	−1.30/−0.02
Drewe	1974	WCST	Non-frontals	85	34.6	−0.52	−0.95/−0.08
Eslinger	1993	WCST	(1) Normals	20	63.4	−1.38	−2.35/−0.40
			(2) Non-frontals	20	59.2	−2.31	−3.44/−1.18
Heaton	1981	WCST	(1) Normals	193	39.5	−1.56	−1.93/−1.19
			(2) Non-frontals	78	42.5	−0.62	−1.08/−0.17
Heaton	1993	WCST	(1) Normals	415	44.8	−1.45	−1.74/−1.16
			(2) Non-frontals	113	41.9	−0.55	−0.92/−0.17
Janowsky	1989	COWA	Normals	18	62.4	−1.13	−2.15/−0.11
Manes	2002	COWA	Normals	32	53.4	−0.75	−1.47/−0.02
Miller	1984	COWA	Normals[c]	60	53	−0.69	−1.22/−0.17
Milner[d]	1963	WCST	(1) Non-frontals	71	26.5	−0.84	−1.39/−0.29
			(2) Non-frontals	23	30.6	−1.71	−2.73/−0.69
Robinson	1980	WCST	(1) Normals	146	39.2	−0.44	−0.88/+0.01
			(2) Non-frontals	46	42.8	−0.59	−1.18/−0.00
Stuss	2001	Stroop	(1) Normals	63	52.9	−0.50	−1.01/+0.01
			(2) Non-frontals	51	49.9	−0.56	−1.19/+0.06
Troyer	1998	COWA	Normals	90	54.4	−0.73	−1.17/−0.30
Vendrell	1995	Stroop	Normals	64	40.5	−0.09	−0.58/+0.40

Note. COWA = Controlled Oral Word Association test; WCST = Wisconsin Card Sorting Test.
[a]Combined sample size for frontal and comparison groups.
[b]Mean age of frontal and comparison groups.
[c]Control participants either had no identifiable neurological disease or had extracranial nervous system pathology (e.g., peripheral nerve damage).
[d]Milner (1963) tested two groups of frontal and non-frontal surgical patients: group 1 was tested before surgery and approximately 18 days after surgery, and group 2 was tested only once between 2 weeks and 15 years after surgery.

where N_{total} is the number of effect sizes included in the meta-analysis, d_{crit} is the critical value of d, and d_+ is the weighted mean effect size.

APPENDIX B

Formulas and Procedures for Moderator Analyses

Preliminary Issues

All computations described below were carried out using the following three computer programs – DSTAT (Johnson, 1989), Meta-Analysis Programs (Schwarzer, 1989), and Statistical Package for the Social Sciences (SPSS) – and by hand using formulas cited in Cooper and Hedges (1994).

Examination of Moderator Variables

Analyses of potential moderator variables were conducted by means of fixed effects strategies based on whether the variables of interest were categorical or continuous (Hedges, 1994). For categorical moderators (i.e., type of test and comparison group), estimates of both between (Q_b) and within (Q_w) group variances were derived and tested along the χ^2 distribution to determine whether they were statistically significant moderators. A significant Q_b indicates that there is significant variability between the groups that comprise the categorical

moderator variable that is greater than what would be expected simply by chance. On the other hand, a significant Q_w indicates that there is still significant variability within each effect size that is not being explained by the categorical moderator. Thus, a categorical moderator variable explains all of the heterogeneity present in the grand mean effect size only when the variance is significant between groups (and is not significant within groups). Similar to the analysis of variance (ANOVA), the Q_b statistic provides an omnibus test for between-group differences. Follow-up contrasts should be conducted when Q_b is significant and there are three or more levels of the moderator variable. [Note that only "type of test" had three levels (i.e., WCST, verbal fluency, and Stroop). A significant Q_w was obtained along with a significant Q_b for this moderator variable and, thus, a contrast analysis was conducted accordingly.]

The following computational formulas, which weight effect sizes by sample size, were used for the two categorical moderator analyses (Hedges, 1994):

$$TW = \sum_{i=1}^{j} (n_i - 3) \qquad \text{(B.8)}$$

$$TWD = \sum_{i=1}^{j} (n_i - 3)d_i \qquad \text{(B.9)}$$

$$TWDS = \sum_{i=1}^{j} (n_i - 3)d_i^2 \qquad \text{(B.10)}$$

$$Q_{\text{Total}} = TWDS - \frac{(TWD)^2}{TW} \qquad \text{(B.11)}$$

$$Q_{wi} = TWDS_i - \frac{TWD_i^2}{TW_i} \qquad \text{(B.12)}$$

$$Q_b = Q_T - Q_w \qquad \text{(B.13)}$$

[For formulas 8–10, $j =$ total number of cases in each sub-group.]

Weighted least squares regression procedures (WLS) were used to test the continuous moderator, age (Hedges, 1994). Effect sizes were weighted by their sample size and then regressed onto the relevant predictor variable (i.e., age). A Z-test of the unstandardized regression coefficient (b) was used to determine the statistical significance (i.e., whether b differed significantly from zero) of the moderator using formulas 14–16; 95% confidence intervals were constructed according to formula 16:

$$Z_j = \frac{b_j}{S_j} \qquad \text{(B.14)}$$

$$S_j = \frac{SE_j}{\sqrt{MS_{\text{error}}}} \qquad \text{(B.15)}$$

$$b_j \pm 1.96(S_j) \qquad \text{(B.16)}$$

$j = 1, \ldots, k;\ k =$ total # of predictors in equation.

ACKNOWLEDGMENT

REFERENCES

Ahola, K., Vilkki, J., and Servo, A. (1996). Frontal tests do not detect frontal infarctions after ruptured intracranial aneurysm. *Brain and Cognition* **31**: 1–16.

Anderson, S. W., Damasio, H., Jones, R. D., and Tranel, D. (1991). Wisconsin Card Sorting Test performance as a measure of frontal lobe damage. *J. Clinical and Experimental Neuropsychology* **13**: 909–922.

Axelrod, B. N., Goldman, R. S., Heaton, R. K., Curtiss, G., Thompson, L. T., Chelune, G. J., and Kay, G. G. (1996). Discriminability of the Wisconsin Card Sorting Test using the standardization sample. *J. Clin. Exp. Neuropsyc.* **18**: 338–342.

Axelrod, B. N., and Henry, R. R. (1992). Age-related performance on the Wisconsin Card Sorting, Similarities, and Controlled Oral Word Association Tests. *Clinical Neuropsychologist* **6**: 16–26.

Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology* **49A**: 5–28.

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

Banich, M. T., Milham, M. P., Atchley, R., Cohen, N. J., Webb, A., Wszalek, T., Kramer, A. F., Liang, Z., Barad, V., Gullett, D., Shah, C., and Brown, C. (2000). Prefrontal regions play a predominant role in imposing an attentional 'set': Evidence from fMRI. *Cognitive Brain Research* **10**: 1–9.

Barcelo, F. (2001). Does the Wisconsin Card Sorting Test measure prefrontal function? *Spanish Journal of Psychology* **4**: 79–100.

Barcelo, F., and Knight, R. T. (2002). Both random and perseverative errors underlie WCST deficits in prefrontal patients. *Neuropsychologia* **40**: 349–356.

Barcelo, F., Munoz-Cespedes, J. M., Pozo, M. A., and Rubia, F. J. (2000). Attentional set shifting modulates the target P3b response in the Wisconsin card sorting test. *Neuropsychologia* **38**: 1342–1355.

Barcelo, F., and Rubia, F. J. (1998). Non-frontal P3b-like activity evoked by the Wisconsin Card Sorting Test. *Neuroreport* **9**: 747–751.

Barkley, R. A. (1996). Linkages between attention and executive functions. In: Lyon, G. R. and Krasnegor, N. A. (Eds.), *Attention, memory, and executive function* (pp. 307–325). Baltimore, MD, US: Paul H. Brookes Publishing Co.

Bayless, J. D., Varney, N. R., and Roberts, R. J. (1989). Tinker Toy Test performance and vocational outcome in patients with closed-head injuries. *J. Clin. Exp. Neuropsyc.* **11**: 913–917.

Bechara, A., Damasio, A. R., Damasio, H., and Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* **50**: 7–15.

Bench, C. J., Frith, C. D., Grasby, P. M., Friston, K. J., Paulesu, E., Frackowiak, R. S. J., and Dolan, R. J. (1993). Investigations of the functional anatomy of attention using the Stroop test. *Neuropsychologia* **31**: 907–922.

Benton, A. L. (1968). Differential behavioral effects in frontal lobe disease. *Neuropsychologia* **6**: 53–60.

Benton, A. L., and Hamsher, K. (1976). *Multilingual aphasia examination*. Iowa City: University of Iowa Press.

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology* **39**: 15–22.

Berman, K. F., Ostrem, J. L., Randolph, C., Gold, J., Goldberg, T. E., Coppola, R., Carson, R. E., Herscovitch, P., and Weinberger, D. R. (1995). Physiological activation of a cortical network during performance of the Wisconsin Card Sorting Test: A positron emission tomography study. *Neuropsychologia* **33**: 1027–1046.

Bigler, E. D. (1988). Frontal lobe damage and neuropsychological assessment. *Archives of Clinical Neuropsychology* **3**: 279–297.

Blenner, J. L. (1993). The discriminant capacity of the Stroop test in tumor neurosurgical patients and its relationship to the visual evoked potential measure. *Personality and Individual Differences* **15**: 99–102.

Blumer, D., and Benson, D. F. (1975). Personality changes with frontal and temporal lobe lesions. In: Benson, D. F. and Blumer, D. (Eds.), *Psychiatric aspects of neurologic disease* (pp. 151–170). New York: Grune and Stratton.

Bornstein, R. A. (1986). Contribution of various neuropsychological measures to detection of frontal lobe impairment. *International Journal of Clinical Neuropsychology* **8**: 18–22.

Bowden, S. C., Fowler, K. S., Bell, R. C., Whelan, G., Clifford, C. C., Ritter, A. J., and Long, C. M. (1998). The reliability and internal validity of the Wisconsin Card Sorting Test. *Neuropsychological Rehabilitation* **8**: 243–254.

Brown, G. G., Kindermann, S. S., Siegle, G. J., Granholm, E., Wong, E. C., and Buxton, R. B. (1999). Brain activation and pupil response during covert performance of the Stroop Color Word task. *Journal of the International Neuropsychological Society* **5**: 308–319.

Bryan, J., and Luszcz, M. A. (2000). Measurement of executive function: Consideration for detecting adult age differences. *J. Clin. Exp. Neuropsyc.* **22**: 40–55.

Burgess, P. W. (1997). Theory and methodology in executive function research. In: Rabbitt, P. (Ed.), *Methodology of frontal and executive function* (pp. 81–116). Hove, UK: Psychology Press.

Burgess, P. W. (2000). Strategy application disorder: The role of the frontal lobes in human multitasking. *Psychological Research* **63**: 279–288.

Burgess, P. W., Alderman, N., Ernslie, H., Evans, J., and Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society* **4**: 547–558.

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

Butler, M., Retzlaff, P. D., and Vanderploeg, R. (1991). Neuropsychological test usage. *Professional Psychology: Research and Practice* **22**: 510–512.

Butler, R. W., Rorsman, I., Hill, J. M., and Tuma, R. (1993). The effects of frontal brain impairment on fluency: Simple and complex paradigms. *Neuropsychology* **7**: 519–529.

Cabeza, R., and Nyberg, L. (2000). Imaging cognition: An empirical review of PET studies with normal subjects. *Journal of Cognitive Neuroscience* **12**: 1–47.

Carlin, D., Bonerba, J., Phipps, M., Alexander, G., Shapiro, M., and Grafman, J. (2000). Planning impairments in frontal lobe demen-

tia and frontal lobe lesion patients. *Neuropsychologia* **38**: 655–665.

Carter, C. S., Mintun, M., and Cohen, J. D. (1995). Interference and facilitation effects during selective attention: an H215O PET study of Stroop task performance. *Neuroimage* **2**: 264–72.

Catefau, A. M., Parellada, E., Lomena, F., Bernardo, M., Setoain, J., Catarineu, S., Pavia, J., and Herranz, R. (1998). Role of the cingulate gyrus during the Wisconsin Card Sorting Test: A single photon emission computed tomography study in normal volunteers. *Psychiatry Research and Neuroimaging* **83**: 67–74.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind* **11**: 63–65.

Çiçek, M., and Nalçaci, E. (2001). Interhemispheric asymmetry of EEG alpha activity at rest and during the Wisconsin Card Sorting Test: Relations with performance. *Biological Psychology* **58**: 75–88.

Cockburn, J. (1995). Performance on the Tower of London test after severe head injury. *Journal of the International Neuropsychological Society* **1**: 537–544.

Cohen, J. (1992). A power primer. *Psychological Bulletin* **112**: 155–159.

Cohen, J. D., Dunbar, K., and McClelland, J. L. (1990). A parallel distributed processing model of the Stroop effect. *Psychological Review* **97**: 332–361.

Cooper, H., and Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.

Costa, L. (1988). Clinical neuropsychology: Prospects and problems. *Clinical Neuropsychologist* **2**: 3–11.

Cripe, L. I. (1996). The ecological validity of executive function testing. In: Sbordone, R. J. and Long, C. J. (Eds.), *Ecological validity of neuropsychological testing* (pp. 171–202). Delray Beach, FL: GR/St Lucie Press, Inc.

Crockett, D., Bilsker, D., Hurwitz, T., Kozak, J. (1986). Clinical utility of three measures of frontal lobe dysfunction in neuropsychiatric samples. *International Journal of Neuroscience* **30**: 241–248.

Cummings, J. L. (1993). Frontal-subcortical circuits and human behavior. *Archives of Neurology* **50**: 873–880.

Cummings, J. L. (1995). Anatomic and behavioral aspects of frontalsubcortical circuits. In: Grafman, J., Holyoak, K. J. and Boller, F. (Eds.), Structure and functions of the human prefrontal cortex. *Annals of the New York Academy of Sciences, 769*, 1–13.

Dagher, A., Owen, A. M., Boecker, H., and Brooks, D. J. (2001). The role of the striatum and hippocampus in planning: A PET activation study in Parkinson's disease. *Brain* **124**: 1020–1032.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and human brain*. New York: Grosset/Putnam.

de Oliveira-Souza, R., Moll, J., Tovar Moll, F., and de Oliveira, D. L. G. (2001). Executive amnesia in a patient with pre-frontal damage due to a gunshot wound. *Neurocase* **7**: 383–389.

Dehaene, S. and Changeux, J. P. (1991). The Wisconsin Card Sorting Test: theoretical analysis and modeling in a neuronal network. *Cerebral Cortex* **1**: 62–79.

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**: 255–264.

Denckla, M. B. (1994). Measurement of executive function. In: Lyon, G. R. (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 117–142). Baltimore, MD, US: Paul H. Brookes Publishing Co.

Denckla, M. B. (1996). A theory and model of executive function: A neuropsychological perspective. In: Lyon, G. R. and Krasnegor, N. A. (Eds.), *Attention, memory, and executive function* (pp. 263–278). Baltimore, MD: Paul H. Brookes Publishing Company.

Drewe, E. A. (1974). The effect of type and area of brain lesion on Wisconsin Card Sorting Performance. *Cortex* **10**: 159–170.

Duke, L. M., and Kaszniak, A. W. (2000). Executive functions in degenerative dementias: A comparative review. *Neuropsychology Review* **10**: 75–99.

Duffy, J. D., and Campbell, J. J. III. (2001). Regional prefrontal syndromes: A theoretical and clinical overview. In: Salloway, S. P.,

Malloy, P. F. and Duffy, J. D. (Eds.), *The frontal lobes and neuropsychiatric illness* (pp. 113–123). Washington, DC, US: American Psychiatric Publishing, Inc.

Dunbar, K., and Sussman, D. (1995). Toward a cognitive account of frontal lobe function: Simulating frontal lobe deficits in normal subjects. In: Grafman, J., Holyoak, K. J. and Boller, F. (Eds.), *Structure and functions of the human prefrontal cortex. Annals of the New York Academy of Sciences* (Vol. 769, pp. 289–304). New York, NY: New York Academy of Sciences.

Duncan, J., Burgess, P., and Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia* **33**: 261–268.

Duncan, J., Johnson, R., Swales, M., and Freer, C. (1997). Frontal lobe deficits after head injury: Unity and diversity of function. *Cognitive Neuropsychology* **14**: 713–741.

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

Eslinger, P. J., and Damasio, A. R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient EVR. *Neurology* **35**: 1731–1741.

Eslinger, P. J., and Grattan, L. M. (1993). Frontal lobe and frontal-striatal substrates for different forms of human cognitive flexibility. *Neuropsychologia* **31**: 17–28.

Espy, K. A., and Kaufmann, P. M. (2002). Individual differences in the development of executive function in children: Lessons from the delayed response and A-not-B tasks. In: Molfese, D. L. and Molfese, V. J. (Eds.), *Developmental variations in learning: Applications to social, executive function, language, and reading skills* (pp. 113–137). Mahwah, NJ, US: Lawrence Erlbaum Associates.

Ettlinger, G., Teuber, H., and Milner, B. (1975). The seventeenth International Symposium of Neuropsychology. *Neuropsychologia* **13**: 125–133.

Ferstl, E. C., Guthke, T., and von Cramon, D. Y. (2002). Text comprehension after brain injury: Left prefrontal lesions affect inference processes. *Neuropsychology* **16**: 292–308.

Frith, C. D., Friston, K. J., Herold, S., Silbersweig, D., Fletcher, P., Cahill, C., Dolan, R. J., Frackowiak, R. S., and Liddle, P. F. (1995). Regional brain activity in chronic schizophrenia patients during the performance of a verbal fluency task. *British Journal of Psychiatry* **167**: 343–349.

Frith, C. D., Friston, K., Liddle, P. F., and Fracknowiak, R. S. J. (1991). Willed action and the prefrontal cortex in man: A study with PET. *Proceedings of the Royal Society of London* **244**: 241–248.

Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher* **10**: 3–8.

Goldstein, K., and Scheerer, M. (1941). Abstract reasoning and concrete behavior: An experimental study with special tests. *Psychological Monographs* **53**: 1–151.

Goodglass, H., and Kaplan, E. (1979). Assessment of cognitive deficit in the brain-injured patient. In: Gazzaniga, M. S. (Ed.), *Handbook of behavioral neurology* (Vol. 2, pp. 3–22). NY: Plenum Press.

Grafman, J., Jonas, B., and Salazar, A. (1990). Wisconsin Card Sorting Test performance based on location and size of neuroanatomical lesion in Vietnam veterans with penetrating head injury. *Perceptual & Motor Skills* **71**: 1120–1122.

Grafman, J., and Litvan, I. (1999). Recognizing the importance of deficits in executive functions. *Lancet* **354**: 1921–1923.

Grafman, J., Vance, S. C., Weingartner, H., Salazar, A. M., and Amin, D. (1986). The effects of lateralized frontal lesions on mood regulation. *Brain* **109**: 1127–1148.

Grant, D. A., and Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology* **38**: 404–411.

Greve, K. W., Love, J. M., Sherwin, E., Mathias, C. W., Ramzinski, P., and Levy, J. (2002). Wisconsin Card Sorting Test in chronic severe traumatic brain injury: Factor structure and performance subgroups. *Brain Injury* **16**: 29–40.

Haines, J., Hoffman, W., Hanada, J., Brown, P., Labs, S., Loberg, D., Walter, C. T. (1994). Further evaluation of prefrontal lobe function data in schizophrenic patients during Wisconsin Card Sorting Test. *American Journal of Psychiatry* **151**: 1842–1843.

Heaton, R. K. (1981). *Wisconsin Card Sorting Test Manual*. Odessa, FL: Psychological Assessment Resources.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., and Curtiss, G. (1993). *Wisconsin Card Sorting Test Manual: Revised and Expanded*. Odessa, FL: Psychological Assessment Resources.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Heck, E. T., and Bryer, J. B. (1986). Superior sorting and categorizing ability in a case of bilateral frontal atrophy: An exception to the rule. *Journal of Clinical & Experimental Neuropsychology* **8**: 313–316.

Hedges, L. V. (1994). Fixed effects models. In: Cooper, H. and Hedges, L. V. (Eds.), *The handbook of research synthesis* (pp. 285–299). New York, NY: Russell Sage Foundation.

Hedges, L. V., and Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego: Academic Press.

Henry, J. D., and Crawford, J. R. (2004). *A meta-analytic review of verbal fluency performance following focal cortical lesions. Neuropsychology* **18**: 284–295.

Holst, P., and Vilkki, J. (1988). Effect of frontomedial lesions on performance on the Stroop Test and word fluency tasks. *J. Clin. Exp. Neuropsyc.* **10**: 79–80.

Humes, G. E., Welsh, M. C., Retzlaff, P., and Cookson, N. (1997). Towers of Hanoi and London: Reliability of two executive function tasks. *Assessment* **4**: 249–257.

James, W. (1890). *The principles of psychology*. New York: Holt.

Janowsky, J. S., Shimamura, A. P., Kritchevsky, M., and Squire, L. R. (1989). Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behavioral Neuroscience* **103**: 548–560.

Johnson, B. T. (1989). *DSTAT: Software for the meta-analytic review of research literatures* (Version 1.00) [Computer software]. Lawrence Erlbaum Associates, Inc.

Jonides, J., Smith, E. E., Koeppe, R. A., Awh, E., Minoshima, S., and Mintun, M. A. (1993). Spatial working memory in humans as revealed by PET. *Nature* **363**: 623–625.

Kafer, K. L., and Hunter, M. (1997). On testing the face validity of planning/problem-solving tasks in a normal population. *Journal of the International Neuropsychological Society* **3**: 108–119.

Kawasaki, Y., Maeda, Y., Suzuki, M., Urata, K., Higashima, M., Kiba, K., Yamaguchi, N., Matsuda, H., and Hisada, K. (1993). SPECT analysis of regional cerebral blood flow changes in patients with schizophrenia during the Wisconsin Card Sorting Test. *Schizophrenia Research* **10**: 109–116.

Konishi, S., Kawazu, M., Uchida, I., Kikyo, H., Asakura, I., and Miyashita, Y. (1999). Contribution of working memory to transient activation in human inferior prefrontal cortex during performance of the Wisconsin Card Sorting Test. *Cerebral Cortex* **9**: 745–753.

Konishi, S., Nakajima, K., Uchida, I., Kameyama, M., Nakahara, K., Sekihara, K., and Miyashita, Y. (1998). Transient activation of inferior prefrontal cortex during cognitive set shifting. *Nature Neuroscience* **1**: 80–84.

Leung, H., Skudlarski, P., Gatenby, J. C., Peterson, B. S., and Gore, J. C. (2000). An event-related functional MRI study of the Stroop color word interference task. *Cerebral Cortex* **10**: 552–560.

Levin, H. S., Song, J., Scheibel, R. S., Fletcher, J. M., Harward, H., Lilly, M., and Goldstein, F. (1997). Concept formation and problem-solving following closed head injury in children. *Journal of the International Neuropsychological Society* **3**: 598–607.

Levine, D. S., and Prueitt, P. S. (1989). Modeling some effects of frontal lobe damage: Novelty and perseveration. *Neural Networks* **2**: 103–116.

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Little, D. M., and Hartley, A. A. (2000). Further evidence that negative priming in the Stroop color-word task is equivalent in older and younger adults. *Psychology* and *Aging* **15**: 9–17.

Luria, A. R. (1966). *Higher cortical functions in man*. NY: Basic Books.

Lyon, G. R., and Krasnegor, N. A. (Eds.). (1996). *Attention, memory, and executive function*. Baltimore, MD, US: Paul H. Brookes Publishing Co.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin* **109**: 163–203.

Malloy, P. F., and Richardson, E. D. (2001). Assessment of frontal lobe functions. In: Salloway, S. P., Malloy, P. F., and Duffy, J. D. (Eds.), *The frontal lobes and neuropsychiatric illness* (pp. 125–137). Washington, DC, US: American Psychiatric Publishing, Inc.

Manes, F., Sahakian, B., Clark, L., Rogers, R., Antoun, N., Aitken, M., and Robbins, T. (2002). Decision-making processes following damage to the prefrontal cortex. *Brain* **125**: 624–639.

Manly, T., and Robertson, I. H. (1997). Sustained attention and the frontal lobes. In: Rabbitt, P. (Ed.), *Methodology of frontal and executive function* (pp. 135–153). Hove, UK: Psychology Press.

Marenco, S., Coppola, R., Daniel, D. G., Zigun, J. R., and Weinberger, D. R. (1993). Regional cerebral blood flow during the Wisconsin Card Sorting Test in normal subjects studied by xenon-133 dynamic SPECT: Comparison of absolute values, percent distribution values, and covariance analysis. *Psychiatry Research: Neuroimaging* **50**: 177–192.

Mentzel, H. J., Gaser, C., Volz, H. P., Rzanny, R., Hager, F., Sauer, H., and Kaiser, W. A. (1998). Cognitive stimulation with the Wisconsin Card Sorting Tests: Functional MR imaging at 1.5 T. *Radiology* **207**: 399–404.

Miceli, G., Caltagirone, C., Gainotti, G., Masullo, C., and Silveri, M. C. (1981). Neuropsychological correlates of localized cerebral lesions in non-aphasic brain-damaged patients. *Journal of Clinical Neuropsychology* **3**: 53–63.

Miller, E. (1984). Verbal fluency as a function of a measure of verbal intelligence and in relation to different types of cerebral pathology. *British Journal of Clinical Psychology* **23**: 53–57.

Milner, B. (1963). Effects of different brain lesions on card sorting: The role of the frontal lobes. *Archives of Neurology* **9**: 100–110.

Milner, B. (1964). Some effects of frontal lobectomy in man. In: Warren, J. M. and Akert, K. (Eds.), *The frontal granular cortex and behavior* (pp. 313–335). New York: McGraw Hill.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin* **27**: 272–277.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and 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**: 49–100.

Morgan, A. B., and Lilienfeld, S. O. (2000). A meta-analytic review of the relation between antisocial behavior and neuropsychological measures of executive function. *Clinical Psychology Review* **20**: 113–136.

Mountain, M. A., and Snow, W. G. (1993). Wisconsin Card Sorting Test as a measure of frontal pathology: A review. *Clinical Neuropsychologist* **7**: 108–118.

Mullen, B. (1989). *Advanced BASIC meta-analysis*. New Jersey: Lawrence Erlbaum Associates.

Nagahama, Y., Fukuyama, H., Yamaguchi, H., Katsumi, Y., Magata, Y., Shibasaki, H., and Kimura, J. (1997). Age-related changes in cerebral blood flow activation during performance of a card sorting test. *Experimental Brain Research* **114**: 571–577.

Nagahama, Y., Fukuyama, H., Yamauchi, H., Matsuzaki, S., Konishi, J., Shibasaki, H., and Kimura, J. (1996). Cerebral activation during performance of a card sorting test. *Brain* **119**: 1667–1675.

Nagahama, Y., Sadato, N., Yamauchi, H., Katsumi, Y., Hayashi, T., Fukuyama, H., Kimura, J., Shibasaki, H., and Yonekura, Y. (1998). Neural activity during attention shifts between object features. *Neuroreport: an International Journal for the Rapid Communication of Research in Neuroscience* **9**: 2633–2638.

Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex* **12**: 313–324.

Nieuwenstein, M. R., Aleman, A., and de Haan, E. H. F. (2001). Relationship between symptom dimensions and neurocognitive functioning in schizophrenia: A meta-analysis of WCST and CPT studies. *Journal of Psychiatric Research* **35**: 119–125.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics* **8**: 157–159.

Owen, A. M., Sahakian, B. J., Hodges, J. R., Summers, B. A., Polkey, C. E., Robbins, T. W. (1995). Dopamine-dependent frontostriatal planning deficits in early Parkinson's disease. *Neuropsychology* **9**: 126–140.

Pardo, J. V., Pardo, P. J., Janer, K. W., and Raichle, M. E. (1990). The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proceedings of the National Academy of Sciences* **87**: 256–259.

Parellada, E., Catafau, A. M., Bernardo, M., Lomena, F., Catarineu, S., and Gonzalez-Monclus, E. (1998). The resting and activation issue of hypofrontality: A single photon emission computed tomography study in neuroleptic-naive and neuroleptic-free schizophrenic female patients. *Biological Psychiatry* **44**: 787–790.

Parks, R. W., Levine, D. S., Long, D. L., Crockett, D. J., Dalton, I. E., Weingartner, H., Fedio, P., Coburn, K. L., Siler, G., Matthews, J. R., and Becker, R. E. (1992). Parallel distributed processing and neuropsychology: A neural network model of Wisconsin Card Sorting and verbal fluency. *Neuropsychology Review* **3**: 213–233.

Parks, R. W., Loewenstein, D. A., Dodrill, K. L., Barker, W. W., Yoshii, F., Chang, J. Y., Emran, A., Apicella, A., Sheramata, W. A., and Duara, R. (1988). Cerebral metabolic effects of a verbal fluency test: A PET scan study. *Journal of Clinical & Experimental Neuropsychology* **10**: 565–575.

Paulesu, E., Goldacre, B., Scifo, P., Cappa, S. F., Gilardi, M. C., Castiglioni, I., Perani, D., and Fazio, F. (1997). Functional heterogeneity of the left inferior frontal cortex as revealed by fMRI. *Neuroreport* **8**: 2011–2017.

Pendleton, M. G., Heaton, R. K., Lehman, R. A., and Hulihan, D. M. (1982). Diagnostic utility of the Thurstone Word Fluency Test in neuropsychological evaluations. *Journal of Clinical Neuropsychology* **4**: 307–317.

Pennington, B. F., Bennetto, L., McAleer, O., and Roberts, R. J. (1996). Executive functions and working memory: Theoretical and measurement issues. In: Lyon, G. R. and Krasnegor, N. A. (Eds.), *Attention, memory, and executive function* (pp. 327–348). Baltimore, MD: Paul H. Brookes Publishing Company.

Perret, E. (1974). The left frontal lobe of man and the suppression of habitual responses in verbal categorical behavior. *Neuropsychologia* **12**: 323–330.

Perry, W., Potterat, E. G., and Braff, D. L. (2001). Self-monitoring enhances Wisconsin Card Sorting Test performance in patients with schizophrenia: Performance is improved by simply asking patients to verbalize their sorting strategy. *Journal of the International Neuropsychological Society* **7**: 344–352.

Peterson, B. S., Skudlarski, P., Gatenby, J. C., Zhang, H., Anderson, A. W., and Gore, J. C. (1999). An fMRI study of Stroop Word-Color Interference: Evidence for cingulate subregions subserving multiple distributed attentional systems. *Biological Psychiatry* **45**: 1237–1258.

Phelps, E. A., Hyder, F., Blamire, A. M., and Shulman, R. G. (1997). FMRI of the prefrontal cortex during overt verbal fluency. *Neuroreport* **8**: 561–565.

Phillips, L. H. (1997). Do "frontal tests" measure executive function? Issues of assessment and evidence from fluency tests. In: Rabbitt, P. (Ed.), *Methodology of frontal and executive function* (pp. 191–213). Hove, UK: Psychology Press.

Podell, K., Lovell, M., Zimmerman, M., and Goldberg, E. (1995). The cognitive bias task and lateralized frontal lobe functions in males. *Journal of Neuropsychiatry and Clinical Neurosciences* **7**: 491–501.

Posner, M. I., and Dehaene, S. (1994). Attentional networks. *Trends in Neuroscience* **17**: 75–79.

Posner, M. I., and Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience* **13**: 25–42.

Posner, M. I., and Snyder, C. R. R. (1975). Attention and cognitive control. In: Solso, R. L. (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

Quantz, J. O. (1897). Problems in the psychology of reading. *Psychological Review Monographs* **2**: (No. 5).

Rabbitt, P. (1997). Introduction: Methodologies and models in the study of executive function. In: Rabbitt, P. (Ed.), *Methodology of frontal and executive function* (pp. 1–38). Hove, UK: Psychology Press.

Ragland, J. D., Gur, R. C., Glahn, D. C., Censits, D. M., Smith, R. J., Lazarev, M. G., Alavi, A., and Gur, R. E. (1998). Frontotemporal cerebral blood flow change during executive and declarative memory tasks in schizophrenia: A positron emission tomography study. *Neuropsychology* **12**: 399–413.

Ramier, A. M., and Hécaen, H. (1970). Rôle respectif des atteintes frontales et de la latéralisation lésionnelle dans les déficits de la "fluence verbale." *Revue Neurologique* **123**: 17–22.

Ready, R. E., Stierman, L., and Paulsen, J. S. (2001). Ecological validity of neuropsychological and personality measures of executive functions. *Clinical Neuropsychologist* **15**: 314–323.

Reitan, R. M., and Wolfson, D. (1994). A selective and critical review of neuropsychological deficits and the frontal lobes. *Neuropsychology Review* **4**: 161–197.

Rezai, K., Andreasen, N. C., Alliger, R., Cohen, G., Swayze, V., and O'Leary, D. S. (1993). The neuropsychology of the prefrontal cortex. *Archives of Neurology* **50**: 636–642.

Rhodes, M. G. (2004). *Age-related differences in performance on the Wisconsin Card Sorting Test: A meta-analytic review. Psychology and Aging* **19**: 482–494.

Riehemann, S., Volz, H., Stuetzer, P., Smesny, S., Gaser, C., and Sauer, H. (2001). Hypofrontality in neuroleptic-naive schizophrenic patients during the Wisconsin Card Sorting Test-A fMRI study. *European Archives of Psychiatry & Clinical Neuroscience* **251**: 66–71.

Robbins, T. W. (1998). Dissociating executive functions of the prefrontal cortex. In: Roberts, A. C., Robbins, T. W. and Weiskrantz, L. (Eds.), *The prefrontal cortex: Executive and cognitive functions* (pp. 117–130). New York, NY: Oxford University Press.

Robinson, A. L., Heaton, R. K., Lehman, R. A. W., and 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**: 605–614.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (rev. ed.). Thousand Oaks, CA: Sage Publications, Inc.

Sbordone, R. J. (1996). Ecological validity: Some critical issues for the neuropsychologist. In: Sbordone, R. J. and Long, C. J. (Eds.), *Ecological validity of neuropsychological testing* (pp. 15–41). Delray Beach, FL: GR/St Lucie Press, Inc.

Sbordone, R. J. (2000). The executive functions of the brain. In: Groth-Marnat, G. (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 437–456). New York, NY, US: John Wiley & Sons, Inc.

Schmidt, F. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist,* **47**: 1173–1181.

Schnirman, G. M., Welsh, M. C., and Retzlaff, P. D. (1998). Development of the Tower of London–Revised. *Assessment* **5**: 355–360.

Schwarzer, R. (1989). *Meta-analysis programs* (Version 5.3) [Computer software]. Berlin: Author.

Sergeant, J. A., Geurts, H., and Oosterlaan, J. (2002). How specific is a deficit of executive functioning for Attention-Deficit/Hyperactivity Disorder? *Behavioural Brain Research* **130**: 3–28.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge: Cambridge University Press.

Shallice, T., and Burgess, P. (1991a). Higher-order cognitive impairments and frontal lobe lesions in man. In: Levin, H. S., Eisenberg, H.

M. and Benton, A. L. (Eds.), *Frontal lobe function and dysfunction* (pp. 125–138). New York: Oxford University Press.

Shallice, T., and Burgess, P. (1991b). Deficits in strategy application following frontal lobe damage in man. *Brain* **114**: 727–741.

Shallice, T., and Evans, M. E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex* **14**: 294–303.

Spreen, O., and Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms, and commentary* (2nd ed.). New York: Oxford University Press.

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

Stuss, D. T., and Alexander, M. P. (2000). Executive functions and the frontal lobes: A conceptual view. *Psychological Research* **63**: 289–298.

Stuss, D. T., Alexander, M. P., Hamer, L., Palumbo, C., Dempster, R., Binns, M., Levine, B., and Izukawa, D. (1998). The effects of focal anterior and posterior brain lesions on verbal fluency. *Journal of the International Neuropsychological Society* **4**: 265–278.

Stuss, D. T., and Benson, D. F. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin* **95**: 3–28.

Stuss, D. T., and Benson, D. F. (1986). *The frontal lobes.* NY: Raven Press.

Stuss, D. T., Benson, D. F., Kaplan, E. F., Weir, W. S., and Della Malva, C. (1981). Leucotomized and nonleucotomized schizophrenics: Comparison on tests of attention. *Biological Psychiatry* **16**: 1085–1100.

Stuss, D. T., Benson, D. F., Kaplan, E. F., Weir, W. S., Naeser, M. A., Lieberman, I., and Ferrill, D. (1983). The involvement of orbitofrontal cerebrum in cognitive tasks. *Neuropsychologia* **21**: 235–248.

Stuss, D. T., Floden, D., Alexander, M. P., Levine, B., and Katz, D. (2001). Stroop performance in focal lesion patients: Dissociation of processes and frontal lobe lesion location. *Neuropsychologia* **39**: 771–786.

Stuss, D. T., and Levine, B. (2002). Adult clinical neuropsychology: Lessons from studies of the frontal lobes. *Annual Review of Psychology* **53**: 401–433.

Stuss, D. T., Levine, B., Alexander, M. P., Hong, J., Palumbo, C., Hamer, L., Murphy, K. J., and Isukawa, 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**: 388–402.

Stuss, D. T., Shallice, T., Alexander, M. P., and Picton, T. W. (1995). A multidisciplinary approach to anterior attentional functions. *Annals of the New York Academy of Sciences* **769**: 191–211.

Taylor, S. F., Kornblum, S., Lauber, E. J., Minoshima, S., and Koeppe, R. A. (1997). Isolation of specific interference processing in the Stroop task: PET activation studies. *Neuroimage* **6**: 81–92.

Teuber, H. L., Battersby, W. S., and Bender, M. B. (1951). Performance of complex visual tasks after cerebral lesions. *Journal of Nervous and Mental Disease* **114**: 413–429.

Thurstone, L. L. (1938). *Primary mental abilities.* Chicago: University of Chicago Press.

Tien, A. Y., Schlaepfer, T. E., Orr, W., and Pearlson, G. D. (1998). SPECT brain blood flow changes with continuous ligand infusion during previously learned WCST performance. *Psychiatry Research and Neuroimaging* **82**: 47–52.

Tranel, D., Anderson, S. W., and Benton, A. (1994). Development of the concept of "executive function" and its relationship to the frontal lobes. In: Boller, F. and Grafman, J. (Eds.), *Handbook of Neuropsychology* (Vol. 9, pp. 125–148). Amsterdam: Elsevier.

Troyer, A. K., Moscovitch, M., Winocur, G., Alexander, M. P., and Stuss, D. (1998). Clustering and switching on verbal fluency: The effects of focal frontal- and temporal-lobe lesions. *Neuropsychologia* **36**: 499–504.

van den Broek, M. D., Bradshaw, C. M., and Szabadi, E. (1993). Utility of the Modified Wisconsin Card Sorting Test in neuropsychological assessment. *British Journal of Clinical Psychology* **32**: 333–343.

Vandierendonck, A. (2000). Executive functions and task switching. *Psychologica Belgica* **40**: 211–226.

Vendrell, P., Junque, C., Pujol, J., Jurado, M. A., Molet, J., and Grafman, J. (1995). The role of prefrontal regions in the Stroop task. *Neuropsychologia* **33**: 341–352.

Vigotsky, L. (1934). Thought in schizophrenia. *Archives of Neurology and Psychiatry* **31**: 1063–1077.

Volz, H., Gaser, C., Häger, F., Rzanny, R., Mentzel, H., Kreitschmann-Andermahr, I., Kaiser, W. A., and Sauer, H. (1997). Brain activation during cognitive stimulation with the Wisconsin Card Sorting Test–A functional MRI study on healthy volunteers and schizophrenics. *Psychiatry Research: Neuroimaging* **75**: 145–157.

Wallesch, C. W., Kornhuber, H. H., Kunz, T., and Brunner, R. J. (1983). Neuropsychological deficits associated with small unilateral thalamic lesions. *Brain* **106**: 141–152.

Wang, L. (1987). Concept formation and frontal lobe function: The search for a clinical frontal lobe test. In: Perecman, E. (Ed.), *The frontal lobes revisited* (pp. 189–205). New York: IRBN Press.

Warbuton, E., Wise, R. J. S., Price, C. J., Weiller, C., Hadar, U., Ramsay, S., and Frackowiak, R. S. J. (1996). Noun and verb retrieval by normal subjects: Studies with PET. *Brain* **119**: 159–179.

Warkentin, S., and Passant, U. (1997). Functional imaging of the frontal lobes in organic dementia. *Dementia & Geriatric Cognitive Disorders* **8**: 105–109.

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

Weinberger, D. R., Berman, K. F., and Zec, R. F. (1986). Physiological dysfunction of dorsolateral prefrontal cortex in schizophrenia: I. Regional cerebral blood flow (rCBF) evidence. *Archives of General Psychiatry* **43**: 114–124.

Welsh, M. C. (2002). Developmental and clinical variations in executive functions. In: Molfese, D. L. and Molfese, V. J. (Eds.), *Developmental variations in learning: Applications to social, executive function, language, and reading skills* (pp. 139–185). Mahwah, NJ, US: Lawrence Erlbaum Associates.

Welsh, M. C., Satterlee-Cartmell, T., and Stine, M. (1999). Towers of Hanoi and London: Contribution of working memory and inhibition to performance. *Brain & Cognition* **41**: 231–242.

Wilson, B. A. (1993). Ecological validity of neuropsychological assessment: Do neuropsychological indexes predict performance in everyday activities? *Applied & Preventive Psychology* **2**: 209–215.

Zable, M., and Harlow, H. F. (1946). The performance of rhesus monkeys on a series of object quality and positional discriminations and discrimination reversals. *Journal of Comparative Psychology* **39**: 13–23.

Zangwill, O. L. (1966). Psychological deficits associated with frontal-lobe lesions. *International Journal of Neurology* **5**: 395–402.

Zelazo, P. D., Carter, A., Reznick, J. S., and Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology* **1**: 198–226.