Cognitive Psychology **41,** 49–100 (2000)
doi:10.1006/cogp.1999.0734, available online at http://www.idealibrary.com on IDEAL®

The Unity and Diversity of Executive Functions and Their Contributions to Complex ''Frontal Lobe'' Tasks: A Latent Variable Analysis

Akira Miyake, Naomi P. Friedman, Michael J. Emerson,
Alexander H. Witzki, and Amy Howerter

University of Colorado at Boulder

and

Tor D. Wager

University of Michigan

This individual differences study examined the separability of three often postulated executive functions—mental set shifting (''Shifting''), information updating and monitoring (''Updating''), and inhibition of prepotent responses (''Inhibition'')—and their roles in complex ''frontal lobe'' or ''executive'' tasks. One hundred thirty-seven college students performed a set of relatively simple experimental tasks that are considered to predominantly tap each target executive function as well as a set of frequently used executive tasks: the Wisconsin Card Sorting Test (WCST), Tower of Hanoi (TOH), random number generation (RNG), operation span, and dual tasking. Confirmatory factor analysis indicated that the three target executive functions are moderately correlated with one another, but are clearly separable. Moreover, structural equation modeling suggested that the three functions

We thank Anna Ficken, Timi Iddings, Silvie Kilworth, Vandana Passi, Juan Quezada, Bob Slevc, and Neal Wolcott for their help in running the experiments and scoring data. We also thank Greg Carey for his statistical advice; John R. Crawford, John Duncan, Priti Shah, and an anonymous reviewer for their comments on a draft of this article; Dan Kimberg and Jim Parker for making the versions of the Wisconsin Card Sorting Test and the Stroop task programs, respectively, available to us; and Ernie Mross for programming the Tower of Hanoi task. This research was supported in part by a National Science Foundation (NSF) KDI/LIS Grant (IBN–9873492), a University of Colorado Council of Research and Creative Work Grants-in-Aid award, and an NSF Graduate Fellowship. A preliminary version of this research was presented at the 11th Annual Convention of the American Psychological Society, Denver, Colorado in June 1999.

Correspondence and reprint requests concerning this article should be addressed to Akira Miyake, Department of Psychology, University of Colorado at Boulder, Campus Box 345, Boulder, CO 80309–0345. Electronic mail may also be sent to MIYAKE@PSYCH.COLORADO.EDU.

0010-0285/00 $35.00
Copyright © 2000 by Academic Press
All rights of reproduction in any form reserved.

contribute differentially to performance on complex executive tasks. Specifically, WCST performance was related most strongly to Shifting, TOH to Inhibition, RNG to Inhibition and Updating, and operation span to Updating. Dual task performance was not related to any of the three target functions. These results suggest that it is important to recognize both the unity and diversity of executive functions and that latent variable analysis is a useful approach to studying the organization and roles of executive functions. © 2000 Academic Press

Cognitive psychology has made considerable progress over the last few decades and has developed sophisticated theories and models about specific cognitive domains or processes (such as object perception, word recognition, syntactic parsing, etc.). Despite this headway, there still remain a number of theoretical issues or phenomena about which little can be said. According to Monsell (1996), one such ‘‘embarrassing zone of almost total ignorance’’ (p. 93) concerns how specific cognitive processes are controlled and coordinated during the performance of complex cognitive tasks. In other words, the field still lacks a compelling theory of executive functions—general-purpose control mechanisms that modulate the operation of various cognitive subprocesses and thereby regulate the dynamics of human cognition.

The main goal of the present article is to provide a necessary empirical basis for developing a theory that specifies how executive functions are organized and what roles they play in complex cognition. Toward this goal, we report an individual differences study of executive functions. Specifically, we focus on three of the most frequently postulated executive functions in the literature—shifting of mental sets, monitoring and updating of working memory representations, and inhibition of prepotent responses—and specify how separable these functions are and how they contribute to so-called frontal lobe or executive tasks.

Research on executive functions has historical roots in neuropsychological studies of patients with frontal lobe damage. It has been known for a long time that patients with damage to the frontal lobes, including the well-known patient Phineas Gage, demonstrate severe problems in the control and regulation of their behavior and cannot function well in their everyday lives. Although some of these patients demonstrate remarkably intact performance on various well-defined cognitive tasks from neuropsychological test batteries and IQ tests (e.g., Damasio, 1994; Shallice & Burgess, 1991), they tend to show, as a group, some impairments on a host of complex frontal lobe or executive tasks. These tasks include, among others, the Wisconsin Card Sorting Test (WCST) and the Tower of Hanoi (TOH) task and its variant, the Tower of London task. Although these tasks are complex and poor performance on them could arise for many different reasons, they have nonetheless become the primary research tools for studying the organization and roles of executive functions in neuropsychological studies of brain-damaged patients and, more recently, in individual differences studies of normal populations from different age groups. In particular, these frontal lobe or executive tasks have provided a basis for many proposals regarding the nature of the cognitive deficits that frontal lobe patients exhibit as well as the nature of the control functions that the normal, intact frontal lobes seem to perform.[1]

One of the most prominent cognitive frameworks that has been associated with the study of executive functions is Baddeley's (1986) influential multicomponent model of working memory. This model includes three components, two of which are specialized for the maintenance of speech-based, phonological information (the phonological loop) and visual and spatial information (the visuospatial sketchpad), respectively. In addition to these two ''slave'' systems, the model also includes a central control structure called the central executive, which is considered responsible for the control and regulation of cognitive processes (i.e., executive functions) and is often linked to the functioning of the frontal lobes. Baddeley (1986) also proposed that Norman and Shallice's (1986; Shallice, 1988) Supervisory Attentional System (SAS), originally constructed as a model of attentional control of behavior in normals as well as neuropsychological patients, may be a candidate model of the central executive.

One important research question that has been a source of controversy in both neuropsychological and cognitive studies of executive functions is an issue raised by Teuber (1972) in his review entitled *Unity and diversity of frontal lobe functions* and recently revisited by Duncan and his colleagues (Duncan, Johnson, Swales, & Freer, 1997). Specifically, to what extent can different functions often attributed to the frontal lobes or to the central executive (or SAS) be considered unitary in the sense that they are reflections of the same underlying mechanism or ability?

At least in the early stages of theoretical development, both the central executive and the SAS had a unitary flavor, without including any distinct subfunctions or subcomponents. In addition, some recent conceptions of executive functions suggest that there is some common basis or a unifying mechanism that can characterize the nature of deficits in frontal lobe patients or the functions of the frontal lobes (e.g., Duncan, Emslie, Williams, Johnson, & Freer, 1996; Duncan et al., 1997; Engle, Kane, & Tuholski, 1999a; Kimberg & Farah, 1993).

In contrast, there is also some evidence for the nonunitary nature of frontal

[1] Despite the fact that the phrases ''frontal lobe'' and ''executive'' are often used interchangeably, they are not conceptually identical (Baddeley, Della Sala, Gray, Papagno, & Spinnler, 1997). Although there is strong evidence that the frontal lobes may play an important role in executive control of behavior, some frontal lobe patients do not show any problems with frontal lobe tasks (Shallice & Burgess, 1991), whereas some patients who have lesions outside the frontal lobes can demonstrate severe impairments on them (Anderson, Damasio, Jones, & Tranel, 1991; Reitan & Wolfson, 1994). Such findings suggest that the anatomical term ''frontal lobe'' and the functional term ''executive'' are not necessarily synonymous. For this reason, we will use the term ''executive tasks,'' rather than ''frontal lobe tasks,'' for the rest of the article.

lobe or executive functions (Baddeley, 1996). One line of evidence comes from clinical observations, which indicate some dissociations in performance among the executive tasks. For example, some patients may fail on the WCST, but not on the TOH, whereas others may show the opposite pattern, suggesting that executive functions may not be completely unitary (e.g., Godefroy, Cabaret, Petit-Chenal, Pruvo, & Rousseaux, 1999; Shallice, 1988).

Another line of evidence for the nonunitary nature of executive functions comes from a number of individual differences studies, the main focus of the present article. These studies examined a wide range of target populations, including normal young adults (Lehto, 1996), normal elderly adults (Lowe & Rabbitt, 1997; Robbins et al., 1998), brain-damaged adults (Burgess, 1997; Burgess, Alderman, Evans, Emslie, & Wilson, 1998; Duncan et al., 1997), and children with neurocognitive pathologies (Levin et al., 1996; Schachar, Tannock, & Logan, 1993; Welsh, Pennington, & Groisser, 1991). Despite differences in the target populations, these studies are similar in the sense that they all employed a battery of widely used executive tasks like the WCST and TOH and examined how well these tasks correlated with one another by performing correlation–regression analyses and, in many cases, exploratory factor analyses (EFA). Although details of the results differ from study to study, a highly consistent pattern that holds across these individual differences studies is that the intercorrelations among different executive tasks are low (usually $r = .40$ or less) and are often not statistically significant. EFA also tends to yield multiple separable factors (rather than a single unitary factor) for a battery of executive tasks. The results from these individual differences studies are often used to argue that the functions of the frontal lobe or the central executive (or SAS) are not unitary and hence need to be fractionated.

Although it has provided useful insights, the typical correlational or factor-analytic approach has several important weaknesses or limitations (Baddeley, Della Sala, Gray, Papagno, & Spinnler, 1997; Rabbitt, 1997a). One major weakness is that, although the finding of low correlations among executive tasks seems to be robust across studies, it is not completely clear whether such reported lack of correlations is indeed a reflection of the independence of underlying executive functions (Miyake & Shah, 1999). It is quite possible that striking differences in nonexecutive processing requirements (e.g., language and visuospatial processing) have simply masked the existence of some underlying commonalities among the chosen executive tasks. More generally, this issue highlights the so-called task impurity problem, a particularly vexing issue in studies of executive functions (Burgess, 1997; Phillips, 1997). Because executive functions necessarily manifest themselves by operating on other cognitive processes, any executive task strongly implicates other cognitive processes that are not directly relevant to the target executive function. For these reasons, a low score on a single executive test does not necessarily mean inefficient or impaired executive functioning. Similarly, low zero-order correlations or multiple separable factors may also not be due to dissociable executive functions (Miyake & Shah, 1999).

This task impurity problem is further compounded by the observation that complex executive tasks tend to suffer from relatively low internal and/or test–retest reliability (Denckla, 1996; Rabbitt, 1997b). Although the reasons for the low reliabilities are not completely clear, one possibility is that people adopt different strategies on different occasions (or even within a session) when performing these tasks. Also, the involvement of executive control functions is generally considered strongest when the task is novel (Rabbitt, 1997b). Thus, repeated encounters with the task may reduce its effectiveness in actually capturing the target executive process, thereby yielding low reliability. Regardless of what factors contribute to the reliability problem, an important point for our current discussion is that measures with low reliabilities necessarily lead to low correlations with other measures. Thus, low zero-order correlations among executive tasks could be a reflection of low reliabilities of the measures themselves, rather than a reflection of independence of underlying executive functions tapped by individual tasks.

Another important problem associated with the reliance on prevalent complex executive tasks like the WCST and TOH is that, despite their wide acceptance as measures of executive functioning, their construct validities are not well established (Phillips, 1997; Rabbitt, 1997b; Reitan & Wolfson, 1994). Many popular executive tasks seem to have been validated only to the rather loose criterion of being somewhat sensitive to frontal lobe damage (i.e., at least some frontal lobe patients show difficulty performing the tasks), and the precise nature of executive processes implicated in the performance of these tasks is underspecified, to say the least. In other words, there is a paucity of rigorous theoretical analysis and independent empirical evidence regarding what these executive tasks really measure.

This unclarity of the underlying abilities tapped by these complex executive tasks is reflected in a proliferation of terms and concepts used to characterize the task requirements of different executive tests. The WCST, for example, has been suggested by different researchers as a measure of ‘‘mental set shifting,’’ ‘‘inhibition,’’ ‘‘flexibility,’’ ‘‘problem solving,’’ and ‘‘categorization,’’ just to name a few. Although these suggestions may sound reasonable at an intuitive level, no independent testing of them has been reported. Another related consequence of the unclarity as to what these executive tests really measure is the difficulty of interpreting what construct(s) different factors obtained in many EFA studies of executive functions really represent. Interpretations given to obtained factors often seem quite arbitrary and post-hoc. For example, a factor that loaded highly on WCST, verbal fluency, and design fluency tests was interpreted as a ‘‘Conceptual/Productivity’’ factor by Levin et al. (1996). Although such interpretation difficulties reflect, in large part, the characteristics of the EFA technique, not knowing what executive functions these complex tasks really tap is likely another important reason.[2]

Taken together, all of these problems seriously undermine the usefulness of typical correlational, factor-analytic studies for theorizing about the organization of executive functions and their roles in complex cognition. Although we do not deny the utility of these methods as exploratory tools, new approaches that overcome these problems are clearly needed for further theoretical development. We argue that one such promising approach is latent variable analysis.

THE PRESENT STUDY

In this article, we report an individual differences study of executive functions that we believe alleviates at least some of the problems that have plagued the typical individual differences approach. Specifically, we focus on three executive functions that are frequently postulated in the literature, carefully select multiple tasks that tap each target executive function, and examine the extent of unity or diversity of these three executive functions at the level of *latent variables* (i.e., what is shared among the multiple exemplar tasks for each executive function), rather than at the level of *manifest variables* (i.e., individual tasks). In other words, we statistically ‘‘extract’’ what is common among the tasks selected to tap a putative executive function and use that ‘‘purer’’ latent variable factor to examine how different executive functions relate to one another.

As will become clear, this latent variable approach has a number of important advantages over a more typical individual differences approach. For example, the emphasis on latent variables (as opposed to manifest variables) should minimize the task impurity problem. In addition, this study examines, also at the level of latent variables, how each target executive function contributes to performance on a number of complex executive tasks such as the WCST and TOH. Such an attempt should provide useful insights as to what each complex executive task really measures and hence will likely contribute to the alleviation of the construct validity problem as well.

The Three Executive Functions Examined in This Study

We focus on the following three executive functions: (a) shifting between tasks or mental sets, (b) updating and monitoring of working memory representations, and (c) inhibition of dominant or prepotent responses. All three are frequently postulated in the literature as important executive functions

[2] Such interpretation difficulties may be exacerbated if an orthogonal rotation technique (a default option of most statistical programs), which does not allow extracted factors to correlate with each other, is used when there is good reason to suspect some interfactor correlations (Fabrigar, Wegener, MacCullum, & Strahan, 1999).

(e.g., Baddeley, 1996; Logan, 1985; Lyon & Krasnegor, 1996; Rabbitt, 1997a; Smith & Jonides, 1999). We chose to focus on these three functions for several reasons. First, they seem to be relatively circumscribed, lower level functions (in comparison to some other often postulated executive functions like ‘‘planning’’) and hence can be operationally defined in a fairly precise manner. Second, for these three executive functions, a number of well studied, relatively simple cognitive tasks that we believed would primarily tap each target function were available. Third, and perhaps most importantly, the three target functions are likely to be implicated in the performance of more complex, conventional executive tests. For example, the WCST has often been suggested as a test that measures set shifting (for shifting between sorting principles) as well as inhibition (for suppressing inappropriate responses). Thus, a good understanding of these three executive functions may provide a basis for specifying what traditional executive tests really measure.

Below, we define and review these three executive functions and briefly discuss the tasks we chose as measures of each executive function. Many of these tasks seem to implicate the frontal lobes, although performance on them would certainly rely on other brain regions as well. Details of each task are provided under Method.

Shifting between tasks or mental sets (‘‘Shifting’’). The first executive function concerns shifting back and forth between multiple tasks, operations, or mental sets (Monsell, 1996). Also referred to as ‘‘attention switching’’ or ‘‘task switching,’’ this ability (henceforth, called ‘‘Shifting’’ for short) has been proposed as a candidate executive function and appears to be important in understanding both failures of cognitive control in brain-damaged patients and laboratory tasks that require participants to shift between tasks (Monsell, 1996). In addition, models of attentional control like SAS (Norman & Shallice, 1986) often assume that the ability to shift between tasks or mental sets is an important aspect of executive control.

The tasks we chose to tap the Shifting function are the plus–minus task (Jersild, 1927), the number–letter task (Rogers & Monsell, 1995), and the local–global task. All of these tasks require shifting between mental sets, although the specific operations that need to be switched back and forth are rather different across tasks; thus, what the three chosen tasks have in common is likely to be the Shifting requirement, rather than other idiosyncratic task requirements not related to the target executive function. Previous studies have shown conclusively that shifting mental sets incurs a measurable temporal cost (e.g., Jersild, 1927; Rogers & Monsell, 1995), particularly when the shifting must be driven internally, rather than by external cues (Spector & Biederman, 1976).

Perhaps the most common explanation of this function is that the Shifting process involves the disengagement of an irrelevant task set and the subsequent active engagement of a relevant task set. Although still prevalent, this conceptualization of Shifting may be too simplistic. Recent work suggests that, when a new operation (say, subtracting 3) must be performed on a set of stimuli (e.g., a list of two-digit numbers), it may be necessary to overcome proactive interference or negative priming due to having previously performed a different operation (e.g., adding 3) on the same type of stimuli (Allport & Wylie, in press). Thus, individual differences in the Shifting ability may not be a simple reflection of the ability to engage and disengage appropriate task sets per se, but may also (or even instead) involve the ability to perform a new operation in the face of proactive interference or negative priming.

Despite the apparent similarity, the notion of Shifting that we focus on in this article is not synonymous with the abilities involved in spatially shifting or switching visual attention by making appropriate voluntary eye movements or covertly moving visual attention. Posner and Raichle (1994) argued that different neural circuits may mediate the shifting of visual attention and more executive-oriented shifts that involve, for example, the conscious fulfilling of instructions, although these networks seem to interact with each other. More specifically, visual attention shifting may be regulated primarily by the parietal lobes and the mid-brain (or the ‘‘posterior attention network’’), whereas more executive-oriented shifts may be regulated primarily by the frontal lobes, including the anterior cingulate (or the ‘‘anterior attention network’’).

In fact, there is a growing body of neuropsychological and neurophysiological evidence indicating that shifting between tasks or mental sets involves the frontal lobes, although not necessarily to the exclusion of other brain regions. For example, an event-related potential (ERP) study has indicated that shifting between two tasks activated frontal as well as bioccipital and parietal regions (Moulden et al., 1998). In addition, one key symptom of frontal lobe impairments, perseveration or repeating the same response over and over even when it is clearly no longer appropriate, is often interpreted in terms of difficulty in shifting mental set (Luria, 1966; Stuss & Benson, 1986). With regard to the tasks used in the current study, we know of no neuroimaging studies demonstrating that the frontal lobes are implicated in the performance on these specific tasks. However, there is some neuropsychological evidence indicating that patients with damage to the left frontal lobes demonstrate a significant shifting impairment compared to age- and IQ-matched controls on a simplified version of the number–letter task, at least in a task condition that is most similar to the task used in this study (Rogers et al., 1998).

Updating and monitoring of working memory representations (‘‘Updating’’). The second target executive function, updating and monitoring of working memory representations (‘‘Updating’’ for short), is closely linked to the notion of working memory (Jonides & Smith, 1997; Lehto, 1996), which in turn is often associated with the prefrontal cortex, particularly its dorsolateral portion (Goldman-Rakic, 1996; Smith & Jonides, 1999). This Updating function requires monitoring and coding incoming information for relevance to the task at hand and then appropriately revising the items held in working memory by replacing old, no longer relevant information with newer, more relevant information (Morris & Jones, 1990). Jonides and Smith (1997) have suggested that this Updating process may involve ''temporal tagging'' to keep track of which information is old and no longer relevant.

Importantly, this Updating function goes beyond the simple maintenance of task-relevant information in its requirement to dynamically manipulate the contents of working memory (Lehto, 1996; Morris & Jones, 1990). That is, the essence of Updating lies in the requirement to actively manipulate relevant information in working memory, rather than passively store information. Consistent with this distinction, recent neuroimaging studies have shown dissociations in the areas required for relatively passive storage and active updating: Whereas the simple storage and maintenance of information has been associated with premotor areas of the frontal cortex and the parietal lobes, the Updating function, as measured by a complex task like the *N*-back task, has been linked to the dorsolateral prefrontal cortex (Jonides & Smith, 1997). In addition, a proposed component of Updating, namely, temporal sequencing and monitoring, has also been associated with the frontal lobes (see Stuss, Eskes, & Foster, 1994, for a review).

The tasks we chose to tap the Updating function are the keep track task (Yntema, 1963), the letter memory task (Morris & Jones, 1990), and the tone monitoring task. All three involve constantly monitoring and updating information in working memory, although the nature of the information that needs to be updated as well as the goals of the tasks is rather different. To our knowledge, no studies have linked the keep track and tone monitoring tasks to the frontal lobes, but a recent PET study has indicated that the updating component of the letter memory task (with the influence of the storage component subtracted) is associated most strongly with the left frontopolar cortex (Van der Linden et al., 1999).

Inhibition of prepotent responses (''Inhibition''). The third executive function examined in this study concerns one's ability to deliberately inhibit dominant, automatic, or prepotent responses when necessary (''Inhibition'' for short). A prototypical Inhibition task is the Stroop task, in which one needs to inhibit or override the tendency to produce a more dominant or automatic response (i.e., name the color word). This type of Inhibition is commonly labeled an executive function—for example, Logan (1994) has called it an ''internally generated act of control'' (p. 190)—and linked to the frontal lobes (e.g., Jahanshahi et al., 1998; Kiefer, Marzinzik, Weisbrod, Scherg, & Spitzer, 1998).

Given that the term *inhibition* is commonly used to describe a wide variety of functions at a number of levels of complexity (Kok, 1999), it is important to note that the conception of Inhibition used here is constrained to the deliberate, controlled suppression of prepotent responses. Thus, by Inhibition, we do not mean inhibition that takes place in typical spreading activation models or connectionist networks. That type of inhibition usually refers to a decrease in activation levels due to negative activation (e.g., a result of negative connection weights) and is not necessarily a deliberate, controlled process. Nor do we mean ''reactive inhibition,'' such as that seen with phenomena like negative priming or inhibition of return. Reactive inhibition seems to be a residual aftereffect of processing that is not usually intended (Logan, 1994), whereas the Inhibition we focus on is a process that is actually intended. Although these two types of inhibition may share some underlying commonality and may be correlated with one another, they are conceptually separable, and we restricted the notion of Inhibition in this study to deliberate, intended inhibition of prepotent responses.[3]

The tasks used to tap the Inhibition ability are the Stroop task (Stroop, 1935), the antisaccade task (Hallett, 1978), and the stop-signal task (Logan, 1994). All require deliberately stopping a response that is relatively automatic, although the specific response that needs to be inhibited differs across tasks. Previous research has shown that both the Stroop task (e.g., Perret, 1974) and the antisaccade task (e.g., Everling & Fischer, 1998) are sensitive to lesions to the frontal lobes and other types of frontal lobe dysfunction. Although the stop-signal task has not been examined in a neuropsychological context, a simpler yet similar ''go–no-go'' task has been shown to strongly implicate the prefrontal cortex among both children (Casey et al., 1997) and adults (Kiefer et al., 1998).

Two Central Goals of the Present Study

Previous individual differences studies of executive functions tend to suggest some level of fractionation of executive functions, but, as we reviewed earlier, there are several serious problems with interpreting the results of typical correlational and EFA studies. The present study was designed to go beyond previous individual differences studies and provide a stronger assessment of the relationships among the three frequently postulated executive functions of Shifting, Updating, and Inhibition. More specifically, the study had two main goals. The first major goal was to specify the extent to which the three target executive functions are unitary or separable. To the extent that the three functions represent distinguishable executive functions,

[3] One alternative conceptualization of inhibition is in terms of actively ''boosting'' activation (or maintaining a high level of activation) for the weaker, to-be-selected process, rather than directly ''suppressing'' the dominant, prepotent process (e.g., Kimberg & Farah, 1993). Both conceptualizations seem plausible at this point as a mechanism involved in the Inhibition process and are compatible with the results of the present study. Thus, we do not strongly endorse one conceptualization over the other, although we discuss the construct of Inhibition in terms of active suppression in this article for the sake of simplicity.

the second major goal was to specify their relative contributions to more complex tests commonly used to assess executive functioning.

With respect to the first goal, we used confirmatory factor analysis (CFA) to specify the degree to which the three postulated executive functions are separable or share the same underlying ability or mechanism. CFA is similar to the EFA technique more commonly used in the field (the term ‘‘factor analysis’’ with no modifier typically refers to EFA, rather than CFA). One major difference, however, is that, whereas EFA finds the one underlying factor model that best fits the data, CFA allows researchers to impose a particular factor model and then see how well that statistical model fits the data (Kline, 1998). In other words, with EFA, one lets the observed data determine the underlying factor model *a posteriori* (this characteristic of EFA is part of the reason the factors extracted with this method do not necessarily have clear interpretations), whereas with CFA, one derives a factor model or models *a priori* on the basis of theoretical considerations and then evaluates its fit to the data. Thus, CFA is a highly theory-driven multivariate analysis technique and serves as a valuable tool for specifying the organization of executive functions.

We used CFA to compare models with one, two, or three factors. Figure 1A illustrates the theoretical model that provided the basis for our analysis (called the ‘‘full, three-factor’’ model). Ellipses in the figure represent the three target latent variables (i.e., Shifting, Updating, and Inhibition), whereas rectangles represent the manifest variables (i.e., individual tasks) that were used to tap the specific functions, as indicated by the straight, single-headed arrows. The curved, double-headed arrows represent correlations among the latent variables.

If it is necessary to postulate three separable factors (one for each target executive function), then this full, three-factor model should provide an excellent fit to the data, and the correlations among the three latent variables will provide an estimation of the degree to which the three target functions are related to one another. In contrast, if the three executive functions essentially tap the same underlying construct and hence should be considered unitary, then a model with one factor (created by fixing all of the correlations among the three latent variables to 1.0 so that this alternative model is ‘‘nested’’ within the full, three-factor model) should provide an excellent fit to the data, a fit statistically *no worse* than the full, three-factor model (because the full, three-factor model has more free parameters than the one-factor model, the absolute fit of the one-factor model cannot exceed that of the three-factor model). Similarly, if two of the target executive functions (but not the third one) tap a common underlying ability, then a model with two factors (i.e., a model that fixes the correlation between the two unitary executive functions to 1.0 but lets the other two correlations vary freely) should provide a fit to the data that is statistically as good as the full, three-factor model. Finally, if the executive functions are completely independent,

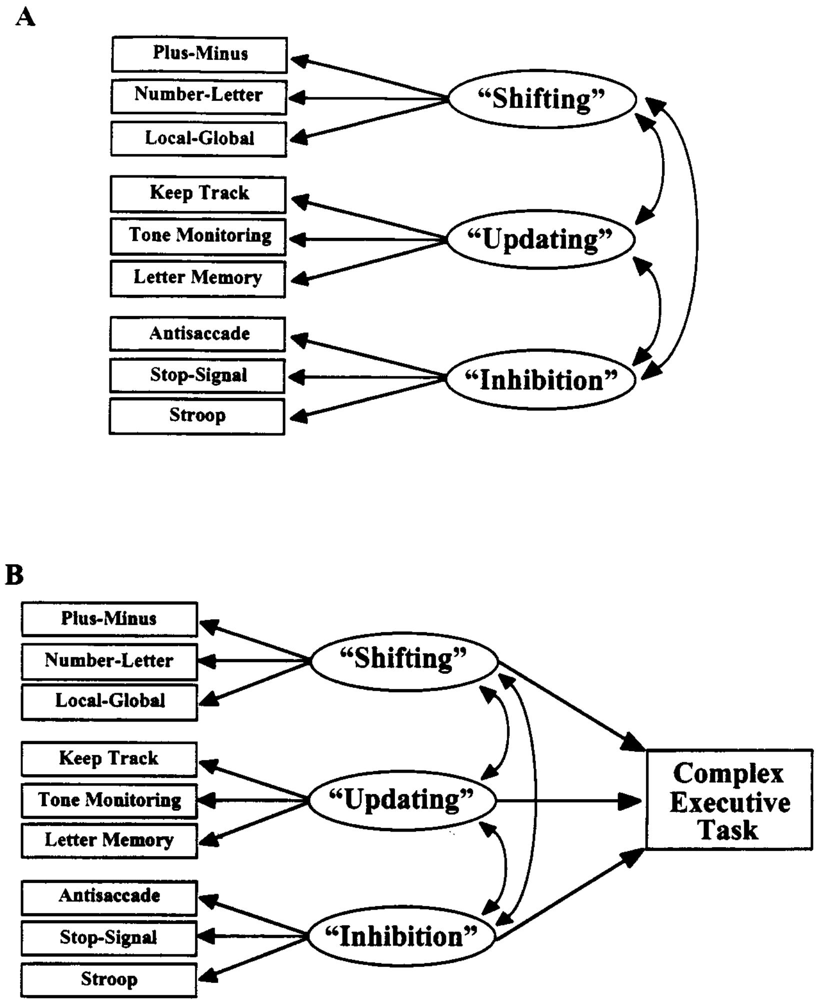

FIG. 1. (A) The theoretical ''full, three-factor'' model used for the confirmatory factor analysis (CFA). The ellipses represent the three executive functions (latent variables), and the rectangles represent the individual tasks (manifest variables) that were chosen to tap the specific executive functions, as indicated by the straight, single-headed arrows. The curved double-headed arrows represent correlations among the latent variables. Both models depict three latent constructs, namely, Shifting, Updating, and Inhibition, which are hypothesized to be correlated but separable. (B) A generic model for the structural equation modeling (SEM) analysis. This model is identical to the CFA model with the addition of a manifest variable on the right side that represents a complex executive function measure. In this particular model (the ''full'' model), the manifest variable on the right has paths from all three latent variables to estimate the contribution of each to performance on the executive task.

then a model with no relationships among the three factors (i.e., a model that fixes the correlations among the factors all to zero) should provide a good fit to the data. Thus, such systematic model comparisons will tell us the degree to which the three executive functions are separable.

For the second goal, we performed a series of structural equation modeling (SEM) analyses to examine how each of the three target executive functions contributes to performance on a number of executive tasks used in cognitive and neuropsychological studies: WCST, TOH, random number generation (RNG), the operation span task, and a dual task. These executive tasks were chosen primarily because they are frequently used as measures of the integrity of executive functioning among frontal lobe patients (i.e., WCST and TOH) or measures of central executive functioning among healthy individuals (i.e., RNG, operation span, and dual tasking).

Figure 1B provides an illustration of the logic behind our SEM analyses. The model is basically the same as the CFA model (Fig. 1A), with the addition of a manifest variable (i.e., an individual executive task) on the right side of the model and potential paths from each latent variable to this new manifest variable. By performing SEM analyses and comparing different alternative models (e.g., models with all three paths, two paths, one path, and zero paths), we sought to determine which path(s) is (are) really necessary to fit the data well or which path(s) can be dropped without significantly worsening the overall data fit. Thus, in this particular application, SEM could be considered a more elaborate version of multiple regression analysis in which latent variables serve as predictor variables. In this context, choosing the best model is analogous to selecting a best-fitting regression model that can parsimoniously account for a significant portion of the variance in the dependent variable with the fewest predictor variables in the equation.

In performing the SEM analyses, we guided our model comparison process by previous proposals in the literature concerning what each of these executive tasks really measures. Specifically, we developed a particular hypothesis (or, in some cases, a particular set of hypotheses) *a priori* and then tested the hypothesis (or hypotheses) with SEM. Thus, the SEM analyses reported in this article provide an independent empirical test of previous proposals regarding the nature of executive function(s) tapped by these complex executive tasks.

METHOD

Participants

The participants were 137 undergraduates from the University of Colorado at Boulder who received partial course credit for taking part in the study. Five additional participants took part in the study, but their data were not complete for the nine target tasks used to tap the three executive functions for the following reasons: Two participants were not native speakers of English and demonstrated marked impairment on certain tasks involving a greater level of language proficiency; one participant was color-blind and had great difficulty performing the Stroop task; and two participants were excluded on the basis of outlier analyses on the nine tasks to be reported later. Thus, the data from these additional participants were not included in the CFA and SEM analyses.

For three of the complex executive tasks (i.e., WCST, TOH, and dual tasking), some observations were lost due to equipment malfunction. Hence, the SEM analyses with these particular tasks relied on slightly fewer observations ($N = 134$ for WCST and dual tasking and $N = 136$ for TOH).

Materials, Design, and Procedure

All participants completed the nine tasks hypothesized to tap one of the three target executive functions of Shifting, Updating, or Inhibition, as well as the five complex tasks commonly used as measures of executive functioning. Task administration was either computerized (Power Macintosh 7200 computers) or paper-and-pencil. A button box with millisecond accuracy was employed for the computerized tasks using reaction time (RT) measures, and a voice key was attached to the button box to record RTs for verbal responses.

The following three tasks were used as the Shifting tasks:

Plus–minus task. The plus–minus task, adapted from Jersild (1927) and Spector and Biederman (1976), consisted of three lists of 30 two-digit numbers (the numbers 10–99 prerandomized without replacement) on a single sheet of paper. On the first list, the participants were instructed to add 3 to each number and write down their answers. On the second list, they were instructed to subtract 3 from each number. Finally, on the third list, the participants were required to alternate between adding 3 to and subtracting 3 from the numbers (i.e., add 3 to the first number, subtract 3 from the second number, and so on). The participants were instructed to complete each list quickly and accurately, and list completion times were measured by a stopwatch. The cost of shifting between the operations of addition and subtraction was then calculated as the difference between the time to complete the alternating list and the average of the times to complete the addition and subtraction lists, and this shift cost served as the dependent measure.

Number–letter task. In the number–letter task, adapted from Rogers and Monsell (1995), a number–letter pair (e.g., 7G) was presented in one of four quadrants on the computer screen. The participants were instructed to indicate whether the number was odd or even (2, 4, 6, and 8 for even; 3, 5, 7, and 9 for odd) when the number–letter pair was presented in either of the top two quadrants and to indicate whether the letter was a consonant or a vowel (G, K, M, and R for consonant; A, E, I, and U for vowel) when the number–letter pair was presented in either of the bottom two quadrants. The number–letter pair was presented only in the top two quadrants for the first block of 32 target trials, only in the bottom two quadrants for the second block of 32 target trials, and in a clockwise rotation around all four quadrants for the third block of 128 target trials. Thus, the trials within the first two blocks required no task switching, whereas half of the trials in the third block required participants to shift between these two types of categorization operations. In all trials (plus 10–12 practice trials in each block), the participants responded by button press, and the next stimulus was presented 150 ms after the response. Similar to the plus–minus task, the shift cost for this task was the difference between the average RTs of the trials in the third block that required a mental shift (trials from the upper left and lower right quadrants) and the average RTs of the trials from the first two blocks in which no shift was necessary.

Local–global task. In the local–global task, a geometric figure often called a Navon figure (Navon, 1977), in which the lines of the ‘‘global’’ figure (e.g., a triangle) were composed of much smaller, ‘‘local’’ figures (e.g., squares), was presented on the computer screen. Depending on the color of the figure (either blue or black), participants were instructed to say out loud the number of lines (i.e., 1 for a circle, 2 for an X, 3 for a triangle, and 4 for a square) in the global, overall figure (blue) or the local, smaller figures (black). Thus, when the colors of the stimuli changed across successive trials, the participants had to shift from examining the local features to the global features or vice versa. A voice key was used to measure RTs. After 36 practice trials (of which 24 served as voice-key calibration trials), participants performed one block of 96 target trials, each separated by a 500-ms response-to-stimulus interval. The target trials were prerandomized, with the constraint that half of the trials require a switch from local to global features or from global to local features, and the shift cost was then calculated as the difference between the average RTs for the trials requiring a shift in mental set (i.e., color of stimulus changed) and the trials in which no shift was required (i.e., the color of stimulus remained the same).

The following three tasks were used as the Updating tasks:

Keep track task. In each trial of the keep track task (adapted from Yntema, 1963), participants were first shown several target categories at the bottom of the computer screen. Fifteen words, including 2 or 3 exemplars from each of six possible categories (animals, colors, countries, distances, metals, and relatives), were then presented serially and in random order for 1500 ms apiece, with the target categories remaining at the bottom of the screen. The task was to remember the last word presented in each of the target categories and then write down these words at the end of the trial. For example, if the target categories were metals, relatives, and countries, then, at the end of the trial, participants recalled the last metal, the last relative, and the last country presented in the list. Thus, participants had to closely monitor the words presented and update their working memory representations for the appropriate categories when the presented word was a member of one of the target categories. Before this task began, participants saw all six categories and the exemplars in each to ensure that they knew to which category each word belonged and then practiced on a single trial with three target categories. Participants then performed three trials with four target categories and three with five target categories, recalling a total of 27 words. The proportion of words recalled correctly was the dependent measure.

Tone monitoring task. In the tone monitoring task (substantially modified from the Mental Counters task developed by Larson, Merritt, & Williams, 1988), participants were presented with four trial blocks, each consisting of a series of 25 tones presented for 500 ms apiece, with an interstimulus interval of 2500 ms. Each block included a mixed order of 8 high-pitched tones (880 Hz), 8 medium-pitched tones (440 Hz), 8 low-pitched tones (220 Hz), and 1 tone randomly selected from the three pitches (for a total of 25 tones). The task was to respond when the 4th tone of each particular pitch was presented (e.g., after hearing the 4th low tone, the 4th medium tone, or the 4th high tone), which required participants to monitor and keep track of the number of times each pitch had been presented. For example, if the sequence was ‘‘low, high, medium, high, high, low, medium, *high,* low, high,’’ then the participant should have responded to the 4th high tone (italicized) and, if asked at the end of the sequence, should also have been aware that his or her mental counters contained 3 low tones, 2 medium tones, and 1 high tone. In order for momentary mental lapses to have less impact on task performance, the tone count for each pitch automatically reset to 0 if participants made an incorrect button press for that pitch (e.g., responding after the 3rd high tone), and participants were informed of this feature before starting the task. Prior to completing the four trial blocks, participants received a guided training session with a shortened block of 14 tones as well as a practice block of 25 tones. With four trial blocks and six potential correct responses per block, the participants could respond correctly a maximum of 24 times. The proportion of correct responses of this total served as the primary measure.

Letter memory task. In the letter memory task (adapted from Morris & Jones, 1990), several letters from a list were presented serially for 2000 ms per letter. The task was simply to recall the last 4 letters presented in the list. To ensure that the task required continuous updating, the instructions required the participants to rehearse out loud the last 4 letters by mentally adding the most recent letter and dropping the 5th letter back and then saying the new string of 4 letters out loud. For example, if the letters presented were ‘‘T, H, G, B, S, K, R,’’ the participants should have said, ‘‘T . . . TH . . . THG . . . THGB . . . HGBS . . . GBSK . . . BSKR’’ and then recalled ‘‘BSKR’’ at the end of the trial. The number of letters presented (5, 7, 9, or 11) was varied randomly across trials to ensure that participants would follow the instructed strategy and continuously update their working memory representations until the end of each trial. After practicing on 2 trials with 5 and 7 letters, respectively, the participants performed 12 trials for a total of 48 letters recalled. The dependent measure was the proportion of letters recalled correctly.

The following three tasks were used as the Inhibition tasks:

Antisaccade task. During each trial of the antisaccade task (adapted from Roberts, Hager, & Heron, 1994), a fixation point was first presented in the middle of the computer screen for a variable amount of time (one of nine times between 1500 and 3500 ms in 250-ms intervals). A visual cue (0.4°) was then presented on one side of the screen (e.g., left) for 225 ms, followed by the presentation of a target stimulus (2.0°) on the opposite side (e.g., right) for 150 ms before being masked by gray cross-hatching. The visual cue was a black square, and the target stimulus consisted of an arrow inside an open square. The participants' task was to indicate the direction of the arrow (left, up, or right) with a button press response. Given that the arrow appeared for only 150 ms before being masked, participants were required to inhibit the reflexive response of looking at the initial cue (a small black square) because doing so would make it difficult to correctly identify the direction of the arrow. The cues and targets were both presented 3.4 in. away from the fixation point (on opposite sides) and the participants were seated 18 in. from the computer monitor (thus, the total subtended visual angle from cue to target was approximately 21.4°). The participants practiced on 22 trials and then received 90 target trials. The proportion of target trials answered correctly served as the dependent measure.

Stop-signal task. The stop-signal task (based on Logan, 1994) consisted of two blocks of trials. On each trial in the first block of 48 trials, used to build up a prepotent categorization response, participants were presented with 1 of 24 words (e.g., duck, gun), balanced for both length and frequency, and were instructed to categorize it as either an animal or nonanimal as quickly as possible without making mistakes. Then, in the second block of 192 trials, participants were instructed *not* to respond (i.e., to inhibit the categorization response) when they heard a computer-emitted tone on 48 randomly selected trials, but otherwise to keep performing the same categorization task as before. As recommended by Logan (1994), the instructions emphasized that the participants should not slow down to wait for possible signals, and whenever slowing was detected, the experimenter reminded them to continue responding as quickly as possible. The time at which the signal occurred during the stop trial was adjusted for each participant by taking the mean response time from the first block of trials and subtracting 225 ms. In all trials (including 34 practice trials), the participants viewed a fixation point for 500 ms and were then allowed up to 1500 ms to categorize the target word by button press. The dependent variable for this task was the proportion of categorization responses for the stop trials.

Stroop task. In the Stroop task (Stroop, 1935), adapted for computer administration, participants were instructed to verbally name the color of a stimulus as quickly as possible in each trial, with RTs measured by voice key. The task included 72 trials with a string of asterisks printed in one of six colors (red, green, blue, orange, yellow, or purple), 60 trials with a color word printed in a different color (e.g., BLUE printed in red color), and 12 trials with a color word printed in the same color (e.g., BLUE in blue color), with the different trial types mixed (i.e., nonblocked). The participants also received three short blocks of approximately 10 trials apiece for voice-key calibration and practice. The dependent measure was the RT difference between the trials in which the word and the color were incongruent and the trials that consisted of asterisks.

We also administered the following five complex executive tasks:

Wisconsin Card Sorting Test. We used a computerized, speeded version of the WCST developed by Kimberg, D'Esposito, and Farah (1997). The task required participants to match a series of target cards presented individually in the middle of the screen with any one of four reference cards shown near the top of the screen. Participants were instructed to sort the target cards into piles under the reference cards according to one of three categories or stimulus attributes—color (red, green, blue, or yellow), number (1, 2, 3, or 4), or shape (circle, cross, star, or square)—and were also told that only one attribute was correct for each target card.

Each target card appeared until a response was given or for a maximum of 3 s, at which point the next trial commenced immediately and the participants received visual feedback (i.e., RIGHT or WRONG appeared below the sorted target card). If the participant did not categorize the target card within this time constraint, the phrase TIME OUT appeared to the right of the target card in the ensuing trial. The category (e.g., ''color'') stayed the same until the participant correctly performed eight consecutive sorts, at which point the sorting criterion changed (e.g., to ''number''). The participants were aware that the sorting criterion would change, but they were not explicitly told the exact number of correctly sorted cards to be achieved before the criterion shifted. After practicing on 30 cards, the main task began and continued until either the participant had successfully achieved 15 sorting categories or the total number of target cards exceeded 288. The main dependent measure was the number of classical perseverative errors, which was the number of times participants failed to change sorting principles when the category changed and kept sorting the cards according to the previous, no longer correct sorting principle.

Tower of Hanoi. In this computerized version of the TOH task, participants were first shown an ending configuration on a piece of paper, consisting of four disks of varying size positioned on three pegs, and were given as much time as necessary to study the configuration. When ready, the participants were shown a different starting configuration on the computer screen and were instructed to make the starting configuration look like the ending configuration by moving the on-screen disks with the computer mouse. The instructions emphasized that the participants were to minimize both the number of moves and the time necessary to accomplish this reconfiguration. When moving the disks, the participants were required to follow a set of rules commonly imposed on the TOH task (i.e., only one disk can be moved at a time, each disk must be placed on one of the pegs, and a larger disk can never be placed on top of a smaller disk). Prior to completing the two target problems, the participants practiced on an easy two-disk problem and then on two four-disk problems that each took a minimum of 11 moves to complete (1 tower-ending and 1 flat-ending). The participants then performed two target four-disk problems that each required a minimum of 15 moves to complete (1 tower-ending and 1 flat-ending). All problems were taken from Humes, Welsh, Retzlaff, and Cookson (1997). The dependent measure for this task was the total number of moves taken to complete the two target problems.

Random Number Generation. In the RNG task, participants heard a computer-generated beep every 800 ms. Their task was to say aloud a number from 1 to 9 for each beep such that the string of numbers would be in as random an order as possible. As an illustration of the concept of randomness (with replacement), the participants were given the analogy of picking a number out of a hat, reading it out loud, putting it back, and then picking another. The importance of maintaining a consistent response rhythm was also emphasized during the instructions, and participants received a brief practice period consisting of 10 beeps.

The valid responses generated during 162 beeps were analyzed using Towse and Neil's (1998) RgCalc program, which produces many different indices that have been commonly used in the analysis of ''randomness.'' The measures we initially derived from the data were the turning point index (TPI), total adjacency (A), runs, Evan's random number generation score (RNG), Guttman's null-score quotient (NSQ), redundancy (R), coupon score, mean repetition gap (mean RG), median repetition gap (med RG), mode repetition gap (mode RG; when there were multiple modes, the smallest was used), phi indices 2 through 7 (phi2–phi7), and analysis of interleaved digrams (RNG2) (see Towse and Neil for full descriptions of these measures). Because Towse and Neil argue that these measures tap different aspects of randomness, we used a principal components analysis to reduce the data (with a Promax rotation to allow for correlated factors). More information about the dependent measures that went into the SEM analyses for this task is provided under Results.

Operation span task. In each trial of the operation span task (adapted from Turner & Engle, 1989), participants received a set of equation–word pairs on the computer screen. For each pair, participants read aloud and verified a simple math equation (e.g., for $(3 * 4) - 6 = 5$, participants said ''three times four minus six equals five . . . false'') and then read aloud a single presented word (e.g., ‘‘king’’). At the end of the trial, the participants recalled all of the words from the entire set of equation–word pairs, with the instructions stipulating that the word from the last pair presented should not be recalled first. For example, if there were four sets of equation–word pairs in the trial, the participants would alternately verify the equation and say the word for each pair and then recall four words at the end of the trial. Each equation remained onscreen until either a verification response was given, at which point the experimenter immediately pressed a response button, or for a maximum of 8 s. Once the equation disappeared, the word was presented for 750 ms before the next equation was displayed. The participants were instructed to begin reading aloud each equation as soon as it appeared and were not allowed any additional time beyond that needed to solve the equation so that the time for idiosyncratic strategies such as rehearsal was minimized. After practicing on three trials at set size 2 (i.e., two equation–word pairs), participants performed four target trials at each set size from 2 to 5. The total number of words recalled correctly (maximum of 56) served as the dependent measure.

Dual task. The dual task required the simultaneous performance of a spatial scanning task (the Maze Tracing Speed Test, developed by Ekstrom, French, Harman, & Dermen, 1976) and a verbal task (word generation). Participants first completed as many mazes as possible in 3 min, with instructions to avoid retracing any lines or removing the pencil from the paper. Next, participants completed the word generation task for 3 min. In this task, participants were auditorily presented with a letter every 20 s and instructed to generate as many words as possible that began with that letter, avoiding proper nouns and function words. In the final dual task condition, participants performed the maze tracing and word generation tasks simultaneously for 3 min. The letters used for the word generation task in the individual and dual task conditions were approximately balanced for the total number of dictionary pages for the letters. Following Baddeley et al. (1997), we used the average proportion of decrement observed in performance from the individual tasks to the dual task, calculated by the following equation:

$$\frac{\left[\left(\dfrac{\text{Maze}_{\text{single}} - \text{Maze}_{\text{dual}}}{\text{Maze}_{\text{single}}}\right) + \left(\dfrac{\text{Word Generation}_{\text{single}} - \text{Word Generation}_{\text{dual}}}{\text{Word Generation}_{\text{single}}}\right)\right]}{2}.$$

General Procedure

Testing took place in two sessions, administered individually during a 2-week period. Each session lasted approximately 1.5 h, for a total of 3 h. The stimuli in each of the tasks were balanced for relevant parameters (e.g., an equal number of true/false answers) when appropriate, and the order of the trials within each task was prerandomized and then fixed for all participants. Also, the order of task administration was fixed for all participants (with the constraint that no two tasks that were supposed to tap the same executive function occurred consecutively) to minimize any error due to participant by order interaction. The tasks administered in Session 1 were (in the order of administration) antisaccade, number–letter, keep track, stop-signal, local–global, Stroop, and letter memory. Those administered in Session 2 were (again in the order of administration) plus–minus, tone monitoring, operation span, RNG, TOH, dual task, and WCST.

Transformations and Outlier Analysis

The distributions of the RT and proportion correct measures for the nine tasks designed to tap the three target executive functions were skewed and/or kurtotic, requiring transformations to achieve normality. For RT measures with multiple trials (only correct trials longer than 200 ms were analyzed), we performed a two-stage trimming procedure. First, upper and lower criteria were determined on the basis of overall, between-subjects RT distributions, and any extreme outliers were replaced with those criterion values. The lower and upper criteria values used in this first stage of trimming were 300 and 3500 ms for the number–letter task, 500 and 3500 ms for the local–global task, and 400 and 2000 ms for the Stroop task, respectively. Next, the within-subject RT distributions were examined for any RTs that were more than 3 standard deviations (*SD*s) away from each individual's mean RT for that task, and these observations were replaced with RTs that were 3 *SD*s away. Because this trimming procedure could not be applied to the plus–minus task (only one trial per condition), the RTs in each condition of this task were trimmed by examining the entire between-subjects distribution and then replacing observations farther than 3 *SD*s from the mean with a value that was 3 *SD*s from the mean. No more than 2.2% of the observations were affected by these trimming procedures in any of the RT-based tasks. For proportion correct measures, we applied an arcsine transformation, which is useful for creating more dispersion in ceiling and floor effects, while having little influence for accuracy scores in the range of .20–.80 (Judd & McClelland, 1989). All of the RT and proportion correct measures achieved a satisfactory level of normality after this trimming/transformation process (see Table 1 for the skewness and kurtosis statistics).

Because the CFA and SEM techniques are sensitive to extreme outliers and careful screening is recommended (Kline, 1998), we performed bivariate outlier analyses on the correlations among the nine tasks designed to tap the three target executive functions.[4] Specifically, outliers were identified by computing leverage, studentized *t*, and Cook's *D* values, which assess how much influence a single observation has on the overall results (Judd & McClelland, 1989). The effects of removal for any participants with very large values for these statistics (i.e., levers greater than .05, *t* values greater than |3.00|, or Cook's *D* values that were much larger than those for the rest of the observations) were determined for each within-construct correlation, and the data for that participant were excluded from analysis only if removal substantially changed the magnitude of the correlations. Only two participants were removed due to these analyses, both of whom greatly affected the correlations within the Inhibition construct.

Statistical Analysis

All of the CFA and SEM analyses reported below were performed with the CALIS procedure (SAS Institute, 1996), a program that uses the maximum likelihood estimation technique to estimate the specified latent variable loadings, based on the *covariance* matrix. For ease of interpretation, the directionality of the dependent measures was adjusted so that larger numbers always indicated better performance.

In both CFA and SEM, we evaluated the fit of each model to the data by examining multiple fit indices: the χ^2 statistic, Akaike's Information Criterion (AIC), the standardized root mean-squared residual (SRMR), Bentler's Comparative Fit Index (CFI), and Bollen's Incremental Fit Index (IFI, also referred to as BL89). We selected these fit indices because they represent different types: absolute fit indices (AIC and SRMR) as well as Type 2 (IFI) and Type 3 (CFI) incremental fit indices. Within these classes of fit indices, the ones we selected were those considered sensitive to model misspecification (i.e., models that lack necessary parameters or cluster the variables inappropriately) while at the same time being relatively insensitive to small sample sizes (i.e., $N < 150$; Hu & Bentler, 1995, 1998).

The most common index of fit is the χ^2 statistic, which measures the ''badness of fit'' of the model compared to a saturated model. Because the χ^2 statistic measures the degree to which the covariances predicted by the specified model differ from the observed covariances,

[4] Because such bivariate outlier analyses were not possible for the five complex executive tasks used in the SEM analyses (one cannot calculate such statistics with latent predictor variables), we examined these additional variables for univariate outliers and replaced observations farther than 3 *SD*s from the mean with a value that was 3 *SD*s from the mean. This procedure affected no more than 2.3% of the observations for each task.

a small value for the χ^2 statistic indicates no statistically meaningful difference between the covariance matrix generated by the model and the observed matrix, suggesting a satisfactory fit. The SRMR also assesses ‘‘badness of fit,’’ as it is the square root of the averaged squared residuals (i.e., differences between the observed and predicted covariances). Lower values of the SRMR indicate a closer fit, with values less than .08 indicating a relatively close fit to the data (Hu & Bentler, 1998). The other fit indices (AIC, CFI, and IFI) are typically used to measure ‘‘goodness of fit.’’ AIC is a modified version of the χ^2 statistic that takes into consideration the ‘‘complexity’’ of the evaluated model (in terms of degrees of freedom) and penalizes more complex models (i.e., models with fewer degrees of freedom). Lower values of AIC (including negative values) indicate better fit. In contrast, for CFI and IFI, higher values indicate better fit, as these indices quantify the extent to which the tested model is better than a baseline model (e.g., one with all covariances set to zero). Typically, IFI and CFI values that exceed .90 or .95 are considered good fits (the values of IFI can exceed 1.0). In addition to these commonly used indices, we also examined specific indications of fit, such as the magnitudes of asymptotically standardized residuals, in comparing different alternative models. None of the models we endorse in our discussion of the CFA and SEM results had large residuals, according to the criteria recommended by Jöreskog and Sörbom (1989).

To examine if one model was significantly better than another, we conducted χ^2 difference tests on ‘‘nested’’ models. This test entails subtracting the χ^2 for the fuller model from the χ^2 for the nested model with a fewer number of free parameters or larger degrees of freedom (degrees of freedom are calculated with an analogous subtraction). If the resulting χ^2 is statistically significant, then the fuller model provides a significantly better fit. For these and all other statistical tests reported here, we used an alpha level of .05.

RESULTS AND DISCUSSION

Preliminary Data Analysis

A summary of descriptive statistics for the nine measures used to tap the three target executive functions (i.e., Shifting, Updating, and Inhibition) is presented in Table 1. All of the measures had relatively low skewness and kurtosis coefficients, and the normalized multivariate kurtosis (Mardia, 1970) for all nine measures was also quite low: .19 (the normalized multivariate kurtosis did not exceed .95 for any of the SEM analyses). Internal reliability estimates for the tasks used in the CFA were calculated using either Cronbach’s alpha or the split-half (odd–even) correlation adjusted by the Spearman-Brown prophecy formula. As seen in Table 1, the reliability estimates for the tasks were relatively low (except for number–letter and stop-signal), a characteristic often reported for executive tasks (Denckla, 1996; Rabbitt, 1997b).

The zero-order correlations among the nine measures, provided in Appendix A, were generally low (.34 or lower), consistent with the results from previous individual differences studies of executive functions. It is important to point out, however, that the correlations among the nine measures were not uniformly low; rather, the tasks considered to tap the same executive function tended to show significant correlations with one another, while correlating not as strongly with the tasks considered to tap the other executive functions, thus showing some signs of convergent and discriminant validity.

TABLE 1
Descriptive Statistics for the Dependent Measures Used in the Confirmatory Factor Analysis and Structural Equation Models ($N = 137$)

Task	Mean (*SD*)	Range	Skewness	Kurtosis	Reliability
Plus–minus	15.5 s (10.8)	−7.2 to 51.7	.60	.57	N/A[a]
Number–letter	546 ms (250)	−45 to 1303	.29	.29	.91[b]
Local–global	210 ms (160)	−289 to 709	.56	1.06	.59[b]
Keep track	.63 (.14) [.58 (.11)]	.22 to .95 [.22 to .81]	.06	−0.21	.31[c]
Tone monitoring	.70 (.26) [.62 (.18)]	.17 to 1.57 [.17 to 1.00]	.38	.36	.63[c]
Letter memory	.99 (.13) [.83 (.07)]	.65 to 1.37 [.60 to .98]	.35	−0.07	.42[c]
Antisaccade	1.16 (.16) [.91 (.07)]	.69 to 1.57 [.63 to 1.00]	−0.24	.27	.77[b]
Stop-signal	.78 (.29) [.67 (.20)]	.02 to 1.57 [.02 to 1.00]	−0.08	−0.27	.92[b]
Stroop	166 ms (60)	50 to 315	.27	−0.65	.72[b]

Note. The data analyses used arcsine-transformed proportion measures and trimmed RTs. For the proportion data, the raw proportion statistics are in brackets.
[a] Reliability could not be calculated for this task because there was only one RT per condition.
[b] Reliability was calculated by adjusting split-half (odd–even) correlations with the Spearman-Brown prophecy formula.
[c] Reliability was calculated using Cronbach's alpha.

This pattern suggests that the measures used to tap each target executive function may have indeed tapped a common underlying ability or function.

To What Extent Are the Three Target Executive Functions Separable?

The first main question we asked in the study was: Are the three target executive functions (i.e., Shifting, Updating, and Inhibition) distinguishable, or do they essentially tap the same underlying construct? We addressed this question with CFA.

The logic of the analysis was as follows: If the three target executive functions are distinguishable constructs, then the full three-factor model depicted in Fig. 1A should provide a significantly better fit to the data than either the model that assumes the unity of all three executive functions (called the ‘‘one-factor'' model) or the models that assume the unity of two of the executive functions (called the ‘‘two-factor'' models). If the three executive functions actually are completely unitary and essentially the same construct, then the one-factor model should provide a fit to the data that is statistically no worse than the more complex three-factor or two-factor models. Finally, if the three functions are entirely separate, then the ‘‘three inde-

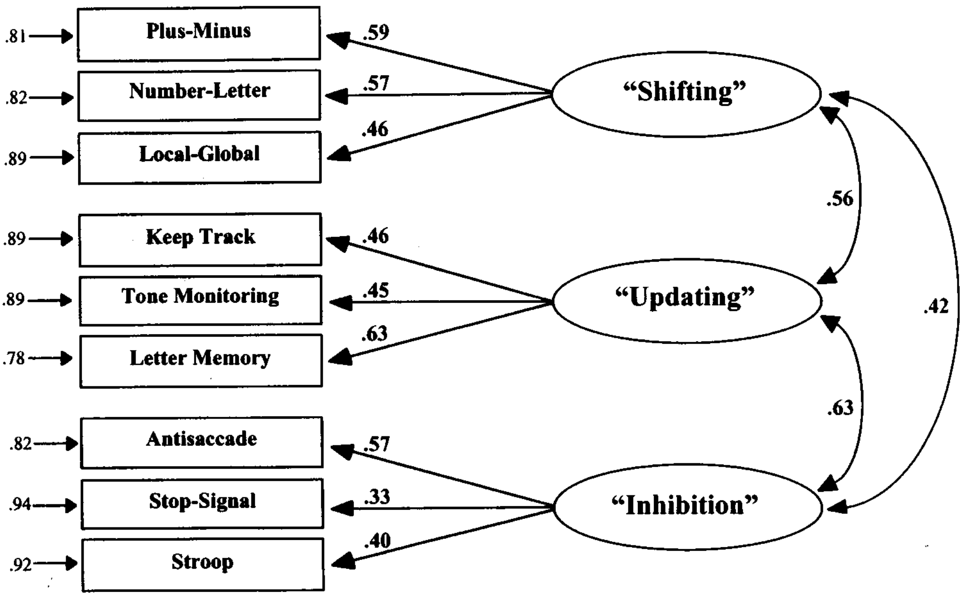

FIG. 2. The estimated three-factor model. Single-headed arrows have standardized factor loadings next to them. The loadings, all significant at the .05 level, are equivalent to standardized regression coefficients (beta weights) estimated with maximum likelihood estimation. The numbers at the ends of the smaller arrows are error terms. Squaring these terms gives an estimate of the variance for each task that is not accounted for by the latent construct. The curved, double-headed arrows have correlation coefficients next to them and indicate significant correlations between the latent variables.

pendent factors'' model, in which all the interfactor correlations are set to zero, should provide a fit to the data similar to that of the model in which the correlations are allowed to vary freely.

The full three-factor model, complete with the estimated factor loadings, is illustrated in Fig. 2. The numbers next to the straight, single-headed arrows are the standardized factor loadings, and those next to the curved, double-headed arrows are the correlations between the factors. In addition, the numbers at the ends of the smaller, single-headed arrows represent the error terms. Squaring these error terms gives an estimate of the unexplained variance for each task, which could be attributed to idiosyncratic task demands and measurement error. Note that all the factor loadings listed in Fig. 2 are equivalent to standardized regression coefficients and can be interpreted accordingly.

The fit indices for this full three-factor model, summarized in Table 2 (Model 1), were all excellent. Specifically, this model produced a nonsignificant $\chi^2(24, N = 137) = 20.29$, $p > .65$, indicating that the model's predic-

TABLE 2
Fit Indices for the Full Confirmatory Factor Analysis Model and Reduced Models ($N = 137$)

Model	*df*	χ^{2a}	AIC[b]	SRMR[b]	CFI[c]	IFI[c]
1. Full three-factor	**24**	**20.29**	**−27.71**	**.047**	**1.00**	**1.04**
2. One-factor	27	36.17	−17.83	.065	.89	.90
Two-factor models						
3. Shifting = Updating	25	29.35	−20.65	.057	.95	.95
4. Shifting = Inhibition	25	29.17	−20.83	.060	.95	.96
5. Updating = Inhibition	25	24.29	−25.71	.052	1.00	1.01
6. Independent three factors	27	47.03*	−6.97	.115	.76	.79

Note. The endorsed model is indicated in bold. AIC, Akaike's Information Criterion; SRMR, standardized root mean-squared residual; CFI, Bentler's Comparative Fit Index; IFI, Bollen's Incremental Fit Index.

[a] χ^2 that were not significant at the .05 level indicate that those models provided reasonable fits; however, all χ^2 difference tests indicated that the reduced models (2–6) provided significantly worse fits than the full model (1).

[b] Lower values of AIC and SRMR indicate better fit, with SRMR $< .08$ indicating a close fit to the data.

[c] Values above .95 for CFI and IFI indicate good fit.

* $p < .05$.

tions did not significantly deviate from the actual data pattern. In addition, the values of the AIC and SRMR were quite low (–27.71 and .047, respectively), whereas the IFI and CFI were well above .95 (1.04 and 1.00, respectively). Thus, this full three-factor model seems to fit the overall data quite well.

One important issue is whether the three latent variable factors could actually be considered to be measuring the same underlying ability. As shown in Fig. 2, the estimates of the correlations among the three latent variables were moderate, ranging from .42 to .63. The 95% confidence intervals for the correlations were [.29, .84] for the Updating and Shifting factors, [.30, .96] for the Updating and Inhibition factors, and [.09, .76] for the Shifting and Inhibition factors, respectively. Because none of these intervals contain 1.0, we can reject the hypothesis that any pair of the three latent variable factors is in fact the same construct.

This conclusion was further supported by the direct statistical comparison of alternative models. We first tested the hypothesis that the three executive functions are not completely unitary by creating a one-factor model that assumes complete unity of the three target executive functions and comparing it against the full three-factor model depicted in Fig. 2. Table 2 summarizes the fit indices for this one-factor model (Model 2), which we created by fixing the correlations among the three latent variable factors at 1.0 (i.e., perfect correlation). The values of the indices were all poorer than the full three-factor model. The AIC and SRMR were relatively high (–18 and .065, respectively), and the IFI and CFI were lower than .95 (.89, and .90, respectively). In addition, the χ^2 difference test produced a significant result, $\chi^2(3) = 15.88$, $p < .01$, suggesting that the one-factor model fit the data significantly worse than the three-factor model did and hence must be rejected.

We also estimated three nested two-factor models in which two of the three executive functions were assumed to be the same. In these models, two of the correlations among the three latent variable factors were allowed to vary and the remaining correlation was set to 1.0. Even though all the fit indices were respectable for these two-factor models (Models 3–5 in Table 2), the χ^2 difference tests showed that the full three-factor model provided a significantly better fit than any of the three two-factor models, all $\chi^2(1) \geq 4.00$, $p < .05$. In other words, none of the correlations among the latent variables could be set to 1.0 without significantly worsening the fit of the model. These findings further support the notion that the three hypothesized constructs are indeed separable.

In addition to the above comparisons, we also compared the full three-factor model to a reduced ‘‘three independent factors’’ model in which all of the correlations among the latent variables were set to zero (i.e., the model in which the three target executive functions are assumed to be completely independent of one another). The resulting fit indices for this model, shown in Table 2 (Model 6), were poor, including a significant χ^2 ($p < .05$, indicating an unsatisfactory overall fit). The χ^2 difference test also indicated that the three independent factors model provided a significantly worse fit than the full three-factor model, χ^2(3 $= 26.74$, $p < .001$, suggesting that the three executive functions share at least some commonality and cannot be considered completely independent.

Taken together, these CFA results suggest that, even though they are clearly distinguishable, the three latent variables share some underlying commonality. Thus, the three target executive functions show signs of both unity and diversity, a point that we consider in more detail under General Discussion.

Which Executive Function(s) Do Complex Executive Tasks Really Tap?

After establishing some separability of the three target executive functions (i.e., Shifting, Updating, and Inhibition) with CFA, we examined the extent to which these functions contribute to performance on more complex executive tasks—WCST, TOH, RNG, operation span, and dual tasking—by performing a series of SEM analyses. The descriptive statistics for these complex tasks are listed in Table 3. Although different proposals have been made regarding what each of these complex tasks really tap, such proposals have not been independently tested in previous neuropsychological or individual differences studies of executive functions and remain highly speculative. In the SEM analyses, we explicitly tested the previously suggested accounts of what these executive tasks really measure.

The logic of the analysis is similar to that used in the CFA and centers

TABLE 3
Descriptive Statistics for the Dependent Measures Used in the Structural Equation Models (N = 137 Unless Noted)

Task	Mean (*SD*)	Range	Skewness	Kurtosis
WCST Perseveration[a]	32 (12)	15 to 67	.91	.22
Tower of Hanoi[b]	46 moves (12)	30 to 86	1.40	1.94
RNG				
Component 1	0 (1)	−3.22 to 2.06	−0.88	1.07
Component 2	0 (1)	−3.02 to 1.66	−0.73	.34
Component 3	0 (1)	−2.10 to 3.22	.74	.76
Operation span	43 words (6)	30 to 55	.01	−0.70
Dual task[a]	.89 (.13)	.52 to 1.27	.22	.25

Note. Extreme observations for each task were trimmed to be 3 *SD*s from the mean.
[a] N = 134.
[b] N = 136.

around the comparisons of alternative models. On the basis of previous proposals, we first selected, for each executive task, models that included specific paths from only one or, in some cases, two latent variables. These *a priori* models were then compared against a ''full'' model that included paths from all three latent variables (illustrated in Fig. 1B). As was the case with the CFA, a hypothesized ''reduced'' model is considered good if the fit indices meet the standard criteria and if a χ^2 difference test indicates that the model's fit is not statistically worse than the fit of the full model. If multiple hypothesized models are considered good for any given task, the more parsimonious model should be preferred. In addition to the full and hypothesized reduced models, we also estimated a ''no-path'' model, which included no paths from the latent variables to the executive task of interest. The preferred hypothesized model should provide a significantly better fit than this no-path model if the executive task is related to any of the three executive functions.

For every model tested, we allowed all of the factor loadings into the latent variables and interfactor correlations to vary (Anderson & Gerbing, 1988). Hence, the estimated parameters could differ from the values found for the original CFA (i.e., those presented in Fig. 2). We allowed these parameters to vary (rather than fixing them at the values obtained in the CFA) because, in addition to examining the path coefficients from the three latent variables to each target executive task, we also wanted to test the stability of the three-factor structure supported by the CFA. If the three-factor CFA model we endorsed earlier is somehow misspecified and the underlying factor structure is rather unstable, then adding an extra executive task in the model may cause major distortions to the original factor structure (i.e., major changes in the factor loadings of the individual executive tasks and/or in the correlations among the three latent variables). In turn, if the paths from the latent variables to the added complex executive tasks are grossly misspecified, then systematic changes from the original factor structure may reveal the nature of the misspecification, as we will discuss in the case of the operation span task results. We should emphasize, however, that, for the structural model we endorse for each executive task, the changes in the parameters were generally quite small. In fact, across the endorsed SEM models, the individual factor loadings and the latent variable correlations showed an average change of only .02 and .03, respectively. Such stability in the parameters across different SEM models suggests that the factor structure depicted in Fig. 2 was highly reliable, giving further credence to the results of the CFA.

Below, we describe the results of our SEM analyses for each complex executive task. For each task, we first provide a brief description of the proposals made previously about what executive function(s) that task may tap and then discuss the results of the SEM analyses that specifically tested these proposals. In the tables that summarize the results for individual complex executive tasks (Tables 4–8), we report the standardized path coefficients as well as the χ^2 statistic, the SRMR, and the IFI for each model we tested. We reduced the number of fit indices in the tables for brevity, but the other indices reported for the CFAs all showed similar results. The model we endorse for each executive task is highlighted in boldface.

Wisconsin Card Sorting Test. Despite the finding that it is sensitive to some impairments that do not necessarily implicate the frontal lobes (Anderson, Damasio, Jones, & Tranel, 1991; Dunbar & Sussman, 1995; Reitan & Wolfson, 1994), the WCST is perhaps the most frequently used test of executive functions in the neuropsychological populations (particularly patients with frontal lobe damage). It has also been successfully used (often with some slight modifications) among normal populations (e.g., Kimberg et al., 1997; Lehto, 1996; Levin et al., 1991). In the literature, the WCST is often conceptualized as a set shifting task (e.g., Berg, 1948) because of its requirement to shift sorting categories after a certain number of successful trials, although this idea has not been independently tested before to the best of our knowledge. In addition, several researchers have considered the hypothesis that the task requires inhibitory control to suppress the current sorting category and switch to a new one (e.g., Ozonoff & Strayer, 1997).

For these reasons, we evaluated the hypothesis that Shifting or Inhibition or both would predict WCST performance by testing the two-path model with paths from Shifting and Inhibition as well as two one-path models with the path from either Shifting or Inhibition. Given the emphasis in the literature on the set shifting requirement of this task as well as the fact that the task does not build up a strong prepotent response before a shift is required, we expected that either a two-path model with both paths from Shifting and Inhibition or a model with only a path from Shifting would provide the best fit.

We performed our SEM analyses on the number of perseverative errors,

TABLE 4
Fit Indices and Standardized Regression Coefficients for Structural Equation Models with Wisconsin Card Sorting Test Classic Perseverations ($N = 134$)

					Coefficients for specified paths		
Model	*df*	χ^2	SRMR	IFI	Shifting	Updating	Inhibition
1. Full three paths	30	25.02	.048	1.05	.33$^+$	.01	.09
2. Two paths from Shifting and Inhibition	31	25.02	.048	1.06	.33*	—	.10
3. One path from Shifting	**32**	**25.45**	**.048**	**1.06**	**.38***	—	—
4. One path from Inhibition	32	30.59	.056	1.01	—	—	.33*
5. No paths	33	37.27	.076	.96	—	—	—

Note. The endorsed model is indicated in bold.
$^+$ $p < .10$.
* $p < .05$.

the measure often considered most sensitive to frontal lobe dysfunction.[5] The results are summarized in Table 4. The χ^2 difference tests indicated that the two-path model with both paths from Shifting and Inhibition (Model 2) provided as good a fit as the full three-path model (Model 1), $\chi^2(1) = 0.00$, $p > .10$. It also produced a significantly better fit than both the no-path model (Model 5), $\chi^2(2) = 12.25$, $p < .01$, and the one-path model with a path from Inhibition (Model 4), $\chi^2(1) = 5.57$, $p < .05$, suggesting that this two-path model provided a good overall fit to the data. However, this model with paths from both Shifting and Inhibition (Model 2) was not statistically better than the one-path model with only a single path from Shifting (Model 3), $\chi^2(1) = .43$, $p > .10$, indicating that the path from Inhibition was really not making much contribution to the prediction of WCST perseverations once Shifting ability had been taken into account. This conclusion is also corroborated by the fact that the full three-path model had a marginally significant coefficient for Shifting (.33), but a much lower one for Inhibition (.09). Thus, taken together, the results from the perseveration measure suggest that the one-path Shifting model is the most parsimonious one and support the conclusion that the Shifting ability is a crucial component of perseverative errors in the WCST, at least in this sample.

As the existing literature suggests (e.g., Anderson et al., 1991; Reitan & Wolfson, 1994), the WCST is clearly a complicated task that taps various

[5] We also examined another dependent measure, the total number of trials necessary to achieve 15 categories, which is analogous to a standard clinical measure of the number of categories achieved within a fixed number of cards. Because the correlation between this measure and the perseveration measure was high ($r = .79$) and showed essentially the same pattern of results, we report only the SEM analyses from the perseveration measure in this article.

cognitive processes and hence cannot be considered selectively sensitive to frontal lobe impairments per se. For example, Dunbar and Sussman (1995) have shown that an impairment in the phonological loop, which could arise from posterior lesions in the left hemisphere, may also lead to perseverations on the WCST by making it difficult to keep the current category highly accessible in memory. Despite the complexity of the WCST as a task, the current results from the SEM analysis demonstrate that the WCST indeed taps, at least in part, one aspect of executive functioning, Shifting, suggesting that it may still serve as a useful executive task if proper caution is taken.

Tower of Hanoi. The TOH puzzle, along with the similar Tower of London puzzle, is frequently described as tapping a ‘‘planning’’ ability (e.g., Arnett et al., 1997), an ability that involves mapping out a sequence of moves in preparation for the task (Morris, Miotto, Feigenbaum, Bullock, & Polkey, 1997). Despite this prevalent conception, the extent to which participants actually do careful planning in this task is unclear, at least when it is administered without specification of what strategies to use (as is usually the case with neuropsychological testing of patients). Indeed, a detailed analysis of strategies has shown that multiple strategies can be used in solving the TOH puzzle (Simon, 1975).

According to this analysis, the strategy that is perhaps most closely related to the notion of ‘‘planning’’ and the one actually guaranteed to solve the puzzle in the minimum number of moves is the goal-recursion strategy, used in some previous studies of the TOH puzzle (Carpenter, Just, & Shell, 1990). This strategy involves extensive goal management and requires setting up a series of subgoals (which, in essence, constitute multiple smaller TOH puzzles with fewer disks) to achieve the superordinate goal. Despite the elegance of this strategy, it is highly demanding, as it requires maintaining a stack of subgoals in working memory. An alternative strategy that is used more prevalently is the so-called perceptual strategy, which involves simply making a next move that will bring the current state perceptually closer to the goal state. This perceptual strategy is much less demanding, and studies of the TOH have demonstrated that most people tend to favor and spontaneously adopt the perceptual strategy in the usual implementation and administration of this task (Goel & Grafman, 1995).

On the basis of this evidence, our main prediction for the TOH puzzle performance was that the Inhibition factor may play an important role because, when one is using the perceptual strategy, the major difficulty seems to come from moves that involve ‘‘goal–subgoal conflicts.’’ These conflicts occur when the optimal action requires moves that take the disk configuration temporarily further away from the goal state—namely, moves that require one to transfer a disk in the opposite direction as the end goal (Morris et al., 1997) and/or to block the goal peg with a disk that must later be cleared (Goel & Grafman, 1995; Simon, 1975). Making these counterintuitive ‘‘conflict moves’’ likely involves overcoming the natural tendency to make more

TABLE 5
Fit Indices and Standardized Regression Coefficients for Structural Equation Models with Tower of Hanoi: Total Number of Moves for the 15-Move Problems ($N = 136$)

Model	*df*	χ^2	SRMR	IFI	Coefficients for specified paths: Shifting	Updating	Inhibition
1. Full three paths	30	24.10	.048	1.06	−0.14	.14	.33
2. One path from Inhibition	**32**	**24.78**	**.049**	**1.07**	—	—	**.37***
3. No paths	33	34.87	.071	.98	—	—	—

Note. The endorsed model is indicated in bold.
* $p < .05$.

obvious, perceptually congruent moves, hence requiring the Inhibition ability (Goel & Grafman, 1995). Given that we intentionally avoided constraining the participants' strategies in the present study's implementation of the TOH puzzle (to simulate typical neuropsychological administrations of this task), we expected that most participants would use the perceptual strategy and hence that the Inhibition factor would play some role in predicting the number of moves they took to solve the target problems.

We tested this hypothesis by estimating a one-path model with a path from the Inhibition factor, along with the three-path and no-path models. The results, summarized in Table 5, indicated that the Inhibition path model (Model 2) provided as good a fit to the data as the three-path model (Model 1), $\chi^2(2) = .68$, $p > .10$, and a significantly better fit than the no-path model (Model 3), $\chi^2(1) = 10.09$, $p < .01$. In addition, the other one-path models we tested (i.e., the ones that include the path from Shifting or Updating, respectively) were not as good as the Inhibition model, a point corroborated by the path coefficients in the full, three-path model (Model 1; −.14 for Shifting, .14 for Updating, and .33 for Inhibition). Thus, we found evidence to support the hypothesis that Inhibition contributes to performance on the TOH puzzle.

To examine the extent to which ''conflict moves'' were indeed responsible for TOH performance, we analyzed the optimal solution paths for the two target problems used in this study. The main finding was that, across both problems, 26% of the moves involved moving a disk in the opposite spatial direction from its ultimate goal and 33% required blocking a goal peg (these two types of conflict moves were not completely independent). Thus, conflict moves constituted a substantial proportion of the moves required for the optimal solution. In addition, we also analyzed each participant's solution paths for the two target problems to assess where the first deviation from the optimal solution occurred, if any deviations did occur. Across both problems, 55% of the first errors occurred on moves requiring moving a disk spatially away from its ultimate goal, and 68% occurred on moves that required blocking a goal peg. These results indicate that the first moves to ''trip up'' our participants (and, hence, to cause longer solution paths) were most often conflict moves. Although this analysis is rough and only assessed the first error each participant made on each problem (after the first deviation from the optimal sequence, the solution paths differed, making it difficult to compare across participants), the results nonetheless support the view that the ability to inhibit the tendency to make perceptually congruent yet incorrect moves is a crucial component of TOH performance.

The important role of the Inhibition ability in solving the TOH puzzle suggests that, at least in its typical method of administration that encourages the perceptual strategy, the TOH should not be conceptualized as a ''planning'' task (Goel & Grafman, 1995). According to recent research (Murji & DeLuca, 1998), this conclusion may also generalize to an analogous Tower of London task, which has also been widely used as a ''planning'' task. We should emphasize, however, that if participants were to use a more demanding strategy that requires more extensive goal management (as in the case of the goal-recursion strategy; Carpenter et al., 1990), then the TOH task might be related less strongly to the Inhibition factor and more strongly to the Updating factor, to the extent that Updating also applies to the management of goal information in working memory.

Random number generation. Within the framework of Baddeley's (1986) multicomponent model of working memory, the RNG task has been one of the most frequently used tasks to examine the functioning of the central executive component. Although a systematic investigation of the underlying processes for this task has begun only recently (e.g., Baddeley, Emslie, Kolodny, & Duncan, 1998; Towse, 1998), several proposals have been made regarding what abilities or functions it really taps.

One common proposal emphasizes the importance of suppressing stereotyped sequences like counting (e.g., 1–2–3–4) to make the produced sequence as random as possible (e.g., Baddeley, 1996; Baddeley et al., 1998), suggesting that the Inhibition factor may play an important role. Another proposal suggests that keeping track of recent responses and comparing them to a conception of randomness is a central aspect of RNG (e.g., Jahanshahi et al., 1998), thus pointing to a role for the Updating factor.

These explanations are not mutually exclusive, and it is entirely possible that both processes contribute to performance on the RNG task. In fact, a recent analysis suggests that this might indeed be the case: Towse and Neil (1998) performed a principal components analysis (PCA) on a set of randomness indices and found that the indices loaded on multiple components. One of the components had high loadings for the randomness indices that seem to be sensitive to the degree to which stereotyped sequences are produced, whereas another component had high loadings for the indices that seem to assess the degree to which each number is produced equally frequently. Towse and Neil interpreted these components as the ''prepotent

TABLE 6
Fit Indices and Standardized Regression Coefficients for Structural Equation Models with Random Number Generation ($N = 137$)

					Coefficients for specified paths		
Model	*df*	χ^2	SRMR	IFI	Shifting	Updating	Inhibition
A. Component 1 (‘‘prepotent associates’’)							
1. Full three paths	30	25.15	.047	1.05	.12	−0.05	.35
2. One path from Inhibition	**32**	**25.60**	**.048**	**1.06**	—	—	**.39***
3. No paths	33	36.97	.073	.96	—	—	—
B. Component 2 (‘‘equality of response usage’’)							
1. Full three paths	30	34.71	.058	.96	−0.08	.52$^+$	−0.17
2. One path from Updating	**32**	**35.52**	**.059**	**.97**	—	**.33***	—
3. No paths	33	44.21$^+$	.075	.90	—	—	—

Note. The endorsed models are indicated in bold.
$^+ p < .10$.
$* p < .05$.

associates’’ component and the ‘‘equality of response usage’’ component, respectively.

We tested Towse and Neil’s (1998) interpretations of these randomness components. Specifically, we first performed a PCA (with an oblique Promax rotation to allow for the possibility that these components might be correlated) on 15 indices of randomness we derived from the data.[6] A three-component solution, shown in Appendix B, was obtained that generally replicated Towse and Neil’s results (the three components accounted for 63% of the total variance). Component 1 was similar to Towse and Neil’s ‘‘prepotent associates’’ component, and Component 2 was similar to their ‘‘equality of response usage’’ component. On the basis of their interpretations of these components, we hypothesized that the present study’s Component 1 should be related to the Inhibition factor, whereas Component 2 should be related to the Updating factor.[7]

We evaluated these predictions by estimating two sets of structural models, one for each of the two RNG components derived from PCA. For our Component 1 (prepotent associates), the main model we tested had the hypothesized path from the Inhibition factor. As Table 6A indicates, this one-

[6] We reversed the direction of all the measures except TPI, mean RG, mode RG, and the phi indices so that higher numbers indicate better performance (i.e., more suppression of habitual counting, more equality of response usage, and less repetition avoidance). We excluded med RG from analysis because it had only three values, thereby violating the normality assumption. We also excluded NSQ because of its high correlation with Evan’s RNG index, $r(135) = .90$.

[7] The prediction for Component 3, comprised primarily of the phi indices, is less clear. The phi indices are complex measures that seem to quantify how reticent participants are to repeat a number at given distances (e.g., the phi4 measure indicates how much participants avoided repeating the same number after three intervening numbers, compared to what would be observed in a random sequence). We did not have a strong *a priori* hypothesis as to the underlying

path model (Model 2) produced as good a fit to the data as the full three-path model (Model 1), $\chi^2(2) = .45$, $p > .10$, and a significantly better fit than the no-path model (Model 3), $\chi^2(1) = 11.37$, $p < .001$. As can be inferred from the three coefficients in the full three-path model (Model 1) in the table, the other single-path models (i.e., the ones with the path from the Shifting or Updating factor, respectively) did not produce as good a fit. These results confirm Towse and Neil's (1998) interpretation that a set of randomness indices that load highly on this component are indeed sensitive to one's ability to inhibit prepotent responses.

For Component 2 (equality of response usage), the main tested model had the hypothesized path from the Updating factor. As indicated in Table 6B, this one-path model (Model 2) produced as good a fit as the three-path model (Model 1), $\chi^2(2) = .81$, $p > .10$, and a significantly better fit than the no-path model (Model 3), $\chi^2(1) = 8.69$, $p < .01$. Taken together with the finding that the other single-path models (i.e., the one with the path from the Shifting factor or the Inhibition factor) did not provide satisfactory fits, these results suggest that the randomness indices that load highly on the equality of response usage component are indeed sensitive to one's ability to update and monitor information in working memory.

These results provide supporting evidence for the previously postulated accounts of the processes underlying RNG. Specifically, RNG draws on multiple executive functions and requires the Inhibition ability to suppress habitual and stereotyped responses as well as the Updating ability to monitor response distribution.[8] This multidimensionality of the RNG task highlights the necessity of using multiple randomness indices to evaluate performance on this task, particularly depending on what aspects of executive functioning one wishes to examine.

These conclusions are corroborated by a recent neuropsychological study on RNG (Jahanshahi et al., 1998). This study found that transcranial magnetic stimulation over the left dorsolateral prefrontal cortex increased the tendency to produce sequences of numbers adjacent on the number line (an indication of habitual counting similar to the Adjacency measure used in the present study, which loaded on the prepotent associates component), without having any effect on a measure of repetition distance (an indication of cycling similar to the repetition gap measures used in the current study, which loaded on the equality of response usage component). This dissociation sug-

target executive function(s) for this measure because others have noted that repetition avoidance is relatively automatic and does not rely on a limited capacity resource (Baddeley et al., 1998).

[8] It has also been suggested that RNG may involve shifting retrieval strategies (Baddeley, 1996). To the extent that the Shifting factor taps this notion, however, the relatively low Shifting coefficients in the full, three-factor models for both RNG factors indicate that this function may not play a major role in performance on the RNG task.

gests that there might be some neural basis for the separability between the Inhibition and Updating factors, as far as performance on the RNG task is concerned.

Operation span. Along with an analogous reading span task (Daneman & Carpenter, 1980), the operation span task (Turner & Engle, 1989) has been used as a measure of working memory capacity that strongly implicates the operations of the central executive (Engle, Tuholski, Laughlin, & Conway, 1999b) and is predictive of performance on complex cognitive tasks, such as reading comprehension tests (Daneman & Merikle, 1996) and complex fluid intelligence tests (Engle et al., 1999b). Although there is no clear consensus as to what this task (or other analogous working memory span tasks) really measures (Miyake & Shah, 1999), there are a number of different proposals that relate to the target executive functions we examined.

The first, perhaps most straightforward possibility is that the operation span test assesses participants' abilities to temporarily store and update incoming information and, hence, should be related to the Updating factor. A second possibility is that the operation span scores are related to the Shifting factor (either instead of or in addition to the Updating factor). Several researchers have pointed out that complex working memory span tasks like the operation span test may require participants to constantly shift back and forth between the processing and storage requirements of the task (i.e., verifying equations and remembering target words). They further suggest that the ability to efficiently shift between these requirements may be crucial for, or at least play an important role in, performance on these tasks (e.g., Conway & Engle, 1996; Towse, Hitch, & Hutton, 1998). According to this view, the model that includes a path from the Shifting factor to the operation span scores should provide a good fit to the data.

We tested these hypotheses with a two-path model (paths from Shifting and Updating) as well as two models with paths from only one of these functions. The results are presented in Table 7, along with the results of the three-path and no-path models for comparison. The χ^2 difference tests indicated that the two-path model (Model 2) provided as good a fit to the data as the full three-path model (Model 1), $\chi^2(1) = .08$, $p > .10$, and a better fit than the no-path model (Model 5), $\chi^2(2) = 38.49$, $p < .001$. Note, however, that the path coefficient from Shifting in this two-path model is negative (i.e., poorer Shifting ability is associated with better operation span scores). A close examination of the models revealed that this weak negative relation was likely to reflect a statistical accommodation of the fact that the relations among the Shifting tasks and operation span scores were weaker than would be expected on purely statistical grounds.[9]

[9] Our statistical explanation of this negative coefficient is as follows: Because the Updating and Shifting latent variables are moderately correlated with each other, even the model with only a path from Updating would expect at least some slight (but not necessarily significant)

TABLE 7
Fit Indices and Standardized Regression Coefficients for Structural Equation Models with Operation Span ($N = 137$)

Model	*df*	χ^2	SRMR	IFI	Coefficients for specified paths: Shifting	Updating	Inhibition
1. Full three paths	30	25.89	.048	1.03	-0.43^{+}	.97*	-0.08
2. Two paths from Updating and Shifting	31	25.97	.049	1.04	-0.42^{+}	.91*	—
3. One path from Shifting[a]	32	49.61*	.073	.87	.51*	—	—
4. One path from Updating	**32**	**32.03**	**.056**	**1.00**	—	**.61***	—
5. No paths	33	64.46*	.101	.76	—	—	—

Note. The endorsed model is indicated in bold.
$^{+}$ $p < .10$.
* $p < .05$.
[a] This model caused two of the paths from the Shifting tasks (number–letter and local–global) to become nonsignificant and resulted in a Heywood case (i.e., a correlation > 1), indicating model misspecification.

This ''statistical accommodation'' interpretation is supported by the observation that the model with only a path from Shifting (Model 3) provided a much worse fit than both the full three-path model (Model 1), $\chi^2(2) = 23.72$, $p < .001$, and the two-path model (Model 2), $\chi^2(1) = 23.64$, $p < .001$, and caused major distortions to the Shifting latent variable (i.e., the loadings for the number–letter and local–global tasks dropped below significance, and the correlation between Updating and Shifting went beyond the upper limit of 1.0, indicating model misspecification). In contrast, the model with only a path from Updating (Model 4), while statistically worse than the three-path model (Model 1), $\chi^2(2) = 6.14$, $p < .05$, and the two-path model (Model 2), $\chi^2(1) = 6.06$, $p < .05$, did not cause such major distortions to the factor structure[10] and was also clearly better than the one-path Shifting model (Model 3) in terms of the fit indices (see Table 7). Thus, based on these statistical reasons as well as the fact that no existing theoretical proposals

correlations between the Shifting tasks and operation span scores. As the correlation matrix presented in Appendix A indicates, however, the actual correlations were essentially 0 (i.e., only in the $-.04$ to .09 range), thus causing the path coefficient from the Shifting variable to be negative to accommodate this lack of expected correlations. We are not sure why the correlations between operation span scores and the Shifting tasks were lower than statistically expected.

[10] It should be noted that this one-path Updating model (Model 4) did cause some modest distortion to the factor structure to accommodate the lower than expected correlations between operation span scores and the Shifting tasks. Specifically, the interfactor correlation between Shifting and Updating dropped from a CFA estimate of .56 (see Fig. 2) to .40 in the endorsed one-path Updating model. This distortion, however, was much more modest in magnitude than that observed for the one-path Shifting model (Model 3) and was not accompanied by major changes in the pattern of loadings for the nine tasks used to tap the target executive functions.

have postulated a negative relationship between operation span performance and set shifting abilities, we endorse the one-path Updating model (Model 4) as the best one for operation span scores.

These results support the hypothesis that the operation span task primarily involves the ability to continuously update and monitor incoming information. This conclusion is also consistent with the findings of other studies that have found significant correlations between working memory span tasks and the letter memory task (Lehto, 1996) as well as the keep track task (Engle et al., 1999b), two of the Updating measures we used in this study. In contrast, as the poor fit of the one-path model from Shifting suggests, we found no evidence for the proposal that the ability to efficiently switch back and forth between the processing component (equation verification) and the storage component (word span) is a crucial aspect of the operation span task, at least to the extent to which the Shifting factor captures the ability to make such switches.

Dual task. The last complex executive task we examined was dual tasking, which has been considered a prime example of the type of situation that implicates the central executive component of working memory (Baddeley, 1996) and has been widely used to study the functioning of the central executive (e.g., Baddeley et al., 1997; Baddeley & Logie, 1999; Bourke, Duncan, & Nimmo-Smith, 1996; Hegarty, Shah, & Miyake, in press). Supporting this claim, a neuroimaging study has shown that simultaneously performing a verbal task and a visuospatial task activates the prefrontal cortex in addition to the areas involved in processing verbal and visuospatial information (D'Esposito et al., 1995). In addition, when compared to performance decrements on individual tasks, neuropsychological studies have found disproportionately larger dual task decrements in various patients with suspected executive function deficits, including traumatic brain injury, frontal lobe lesions, and Alzheimer's and Parkinson's diseases (see Baddeley et al., 1997, for a review).

Despite this general agreement that dual tasking involves executive control processes, there is still no clear consensus on what abilities or specific executive functions are implicated in dual task performance (Miyake & Shah, 1999). One common proposal is that dual tasking involves constantly and rapidly shifting mental set between tasks (e.g., Duncan, 1995). This conception would predict that the Shifting factor would contribute to dual task performance.

We tested this hypothesis by comparing a model with only a path from Shifting to dual task performance against the three-path and no-path models. As shown in Table 8, there was no evidence that Shifting contributed to dual task performance: The model with only a path from Shifting (Model 2) was no better than the no-path model (Model 3), $\chi^2(1) = 0.0$, $p > .10$. Further, the three-path model (Model 1) was also no better than the no-path model, $\chi^2(3) = 1.60$, $p > .10$, indicating that none of our factors significantly pre-

TABLE 8
Fit Indices and Standardized Regression Coefficients for Structural Equation Models with Dual Task Performance (N = 134)

					Coefficients for specified paths		
Model	*df*	χ^2	SRMR	IFI	Shifting	Updating	Inhibition
1. Full three paths	30	27.41	.052	1.03	−0.02	.24	−0.27
2. One path from Shifting	32	29.01	.054	1.03	−0.01	—	—
3. No paths	**33**	**29.01**	**.054**	**1.04**	—	—	—

Note. The endorsed model is indicated in bold.

dicted dual task performance.[11] These results suggest that dual tasking may tap an executive function that is somewhat independent of the three target functions examined in this study, although null results such as these need to be interpreted cautiously.

Summary. The results for the SEM analyses indicate that the three target executive functions contribute differentially to performance on the more complex executive tests. Specifically, Shifting seems to contribute to WCST performance, Inhibition to TOH performance (at least in the way it is typically administered), Inhibition and Updating to RNG performance, and Updating to operation span performance. Dual task performance did not seem to be related to any of the three executive functions examined in this study, although this result is difficult to interpret.

Alternative Explanations

Although the statistical models evaluated in this article are relatively simple, the interpretations of the CFA and SEM results critically hinge on our assumption that the three latent variables in these models indeed succeeded in tapping the three target executive functions (i.e., Shifting, Updating, and Inhibition, respectively). Any violations of this assumption would seriously challenge the conclusions we drew from the CFA and SEM results. Therefore, it is important to rule out alternative explanations that question the validity of that assumption.

One such alternative explanation is that different models that were not explicitly tested and reported in this article (e.g., a different classification of

[11] We also performed the same SEM analyses on another dependent measure, namely, the average standardized residuals obtained by separately regressing the maze completion and word generation scores in the dual task condition on their respective scores from the single task condition, standardizing both sets of residuals from these analyses, and averaging the two sets. Although less widely used, this measure has a number of statistical advantages over the proportion decrement score reported in the main text, such as less error and higher reliability (Cohen & Cohen, 1983). The SEM results, however, remained virtually identical, suggesting that the null results might not be due solely to the statistical characteristics of the proportion decrement scores.

the nine manifest variables into three latent variables) might produce equivalent or even better fits to the data than the CFA model presented in Fig. 2. Although it is difficult (as well as unwise) to test all the CFA models possible and show that the model of our choice is indeed superior to those models, it seems clear that a substantially different CFA model would not produce as good a fit for several reasons. First, an inappropriate clustering of the task would be apparent from a number of sources, such as the magnitude and significance of the specified paths and the magnitudes of the residuals in the fitted covariance matrix. The CFA model endorsed here (Fig. 2) as well as the SEM model selected for each complex executive task was free from problems indicated by such sources. Second, the results of an EFA performed on the same data set, albeit not as clear-cut as the CFA results, largely conformed to the factor structure that we postulated (see Appendix C for the results).[12] Because EFA maximizes fit to the data without any constraints on which tasks should load on which factors, this finding suggests that a substantially different structure would not fit the data better.

Closely related is the alternative explanation based on the types of dependent measures used for different latent variables: Because the three tasks loading on the Shifting factor were all RT-based measures and the three tasks loading on the Updating factor were all accuracy measures, the separability of the factors might be due to this methodological artifact. Two lines of evidence argue against this account. First, if the Shifting and Updating factors were separable primarily because they were composed of RT-based and accuracy measures, respectively, then we would expect to see some signs of unsatisfactory model fit, such as large residuals in the fitted covariance matrix, particularly for the Inhibition measures that included both RT-based and accuracy measures. There were no such signs in the data, however. Second, the SEM results did not conform to the pattern that the Shifting factor predicted performance on RT-based executive tasks and the Updating factor predicted performance on accuracy-based executive tasks. Instead, the three latent variables showed differential contributions to performance on the complex executive tasks in a manner consistent with our *a priori* predictions.

[12] An examination of the EFA results, presented in Appendix C, indicates that the tone monitoring task may be related not just to the Updating factor but also to the Inhibition factor, suggesting that it may not be considered a relatively pure Updating task. In retrospect, the tone monitoring task does seem to involve an Inhibition ability in that it requires participants not only to monitor counters for the three tones but also to reset those counters every time the fourth tone for each pitch occurs. Subjectively, resetting the counter to 0 is quite difficult and may require an Inhibition ability to overcome the tendency to keep counting. To make sure that this impurity of the tone monitoring task did not distort the conclusions we reached, we also estimated the same CFA and SEM models without the tone monitoring task (i.e., only two, rather than three, tasks used for the Updating factor) to examine the impact of this task being included. The results of these analyses indicated that this three-factor model without the tone monitoring task also provided an excellent fit to the data, and the qualitative conclusions for the model comparisons (both CFA and SEM) remained identical.

These results suggest that the separability of the three executive functions is not due to an artifact resulting from the RT–accuracy distinction in the dependent measures.

Another alternative explanation is that some nonexecutive task requirements that were common within the three tasks chosen to tap each target executive function might have driven the underlying factor structure, rather than the presence of separable executive functions, as we contend. Although this alternative explanation cannot be completely ruled out on the basis of available data, we attempted to minimize the influence of idiosyncratic task requirements by deliberately choosing, for each latent variable, tasks that required the same executive function but involved quite different specific requirements (e.g., stopping a prepotent eye movement for antisaccade and stopping a prepotent categorization response for stop-signal). Furthermore, it is not the case that we simply picked tasks for each function that essentially were the same tasks with minor parametric variations. Thus, it seems almost impossible to explain the obtained pattern of CFA and SEM results purely in terms of the commonality and separability of task requirements other than the three postulated executive functions.

In summary, the arguments against these alternative explanations are strong; there is little evidence in the data that suggests a violation of the assumption that the three latent variables in the CFA and SEM analyses tapped the intended target executive functions. Although it is a mistake to take for granted (or consider it proven) that the latent variables fully captured the intended underlying functions or abilities (Kline, 1998), these considerations provide strong support for the view that the latent variables indeed captured the respective target executive functions.

GENERAL DISCUSSION

In this article, we reported an individual differences study that examined the organization and roles of three often-postulated executive functions—shifting between mental sets or tasks (Shifting), updating and monitoring of working memory contents (Updating), and inhibition of prepotent responses (Inhibition)—at the level of latent variables, rather than at the level of manifest variables (i.e., individual tasks). One primary goal of the study was to specify the degree of relationship among the three target functions and thereby contribute to the understanding of the unitary versus nonunitary nature of executive functions. The second main goal was to examine how the three target executive functions contribute to performance on more complex executive tasks. The study yielded clear results with respect to both of these goals.

Regarding the first main goal, the results from the CFA indicated that the three target functions (i.e., Shifting, Updating, and Inhibition) are clearly distinguishable. The full three-factor model in which the correlations among the three latent variables were allowed to vary freely produced a significantly better fit to the data than any other models that assumed complete unity among two or all three of the latent variables. The three target executive functions are not completely independent, however, and do seem to share some underlying commonality. In the full three-factor model (Fig. 2), the estimates of the correlations among the three latent variables were moderately high (ranging from .42 to .63). In addition, this full model provided a far better fit to the data than the three-factor model that assumed complete independence among the three latent variables. These results suggest that the three often postulated executive functions of Shifting, Updating, and Inhibition are separable but moderately correlated constructs, thus indicating both unity and diversity of executive functions.

As for the second goal, the results of the SEM analyses showed that the executive tasks often used in cognitive and neuropsychological studies are not completely homogeneous in the sense that different executive functions contribute differentially to performance on these tasks. Specifically, we found that the Shifting ability contributes to performance on the WCST. The Inhibition ability seems to play an important role in solving the TOH puzzle, at least when no specific instructions for strategies are given and many people are likely to use the perceptual strategy to perform the task. Producing random sequences of numbers in the RNG task seems to depend on multiple abilities, particularly the Inhibition ability and the Updating ability, which appear to be tapped by different sets of randomness indices. Finally, the operation span task, a prevalent measure of verbal working memory capacity, seems to primarily implicate the Updating ability. These results indicate that the Shifting, Updating, and Inhibition abilities contribute differentially to performance on commonly used executive tasks, even though they are moderately correlated with one another. Moreover, the results offer a clear, independent confirmation of some previously proposed accounts of what these tasks really measure, at least in a sample of young, healthy college students.

The only complex executive task that did not relate clearly to the three target executive functions was the dual task. Although such null results need to be interpreted cautiously, one possibility is that the simultaneous coordination of multiple tasks is an ability that is somewhat distinct from the three executive functions examined in this study.

It is important to point out that the current data are based on a restricted sample of young college students. Therefore, the results may not be completely generalizable to more cognitively diverse samples, such as those that include noncollege students, young children, elderly adults, or neurologically impaired participants. For example, the degree of separability of different executive functions may be less pronounced among such less restricted samples (e.g., Legree, Pifer, & Grafton, 1996). It could also be the case that different factors than the ones we reported here contribute more strongly to performance on the complex executive tasks, possibly reflecting different strategies adopted by participants or some specific patterns of age-related changes or neurological impairments in executive functions. Although such limitations in generalizability across samples are possible, there is also a good chance of similar patterns of results emerging across different samples, given that the overall pattern of zero-order correlations we found in this study is fairly analogous to the patterns obtained from previous individual differences studies that tested a wide range of target populations (e.g., college students, young children, elderly adults, and brain-damaged patients).

The Unity and Diversity of Executive Functions Revisited

The main results from the CFA analyses indicate that executive functions may be characterized as separable but related functions that share some underlying commonality. Thus, as Teuber (1972) suggested in his review of frontal lobe functions more than a quarter of a century ago, the results point to both unity and diversity of executive functions and indicate that both of these aspects need to be taken into consideration in developing a theory of executive functions (see also Duncan et al., 1997).

Concerning the unity of executive functions, the results of the present study are compatible with a fair number of theoretical proposals that note some ‘‘family resemblance’’ or common mechanisms across different executive functions or functions putatively performed by the frontal lobes (e.g., Duncan et al., 1996, 1997; Engle et al., 1999a; Kimberg & Farah, 1993). The moderately high intercorrelations among the three target executive functions raises one important theoretical question, however. What might the source(s) of the commonality be? Although precisely specifying the nature of the underlying commonality is beyond the scope of this article and awaits future research, at least two explanations seem possible.

First, although the nine tasks used in the CFA were each chosen to tap one target executive function, it is likely that they did share some common task requirements, particularly the maintenance of goal and context information in working memory. Working memory plays a prominent role in several existing theoretical accounts of executive functions, in which the crucial role of the frontal lobes is hypothesized to be the active maintenance of goals, plans, and other task-relevant information in working memory (Engle et al., 1999a, 1999b; Kimberg & Farah, 1993; O’Reilly, Braver, & Cohen, 1999; Pennington, Bennetto, McAleer, & Roberts, 1996). For example, Engle et al. (1999a, 1999b) recently proposed that a crucial component of working memory capacity is ‘‘controlled attention,’’ which is a domain-free attentional capacity to actively maintain or in some cases suppress working memory representations. In their account, any situations that involve controlled processes (such as goal maintenance, conflict resolution, resistance to or suppression of distracting information, error monitoring, and effortful memory search) would require this ‘‘controlled attention’’ capacity, regardless of the specifics of the tasks to be performed. Thus, the ability to keep goal-related and other task-relevant information active in working memory during controlled processing could be the basis for the observed commonality among the three executive functions.

Another possible explanation is that the three target executive functions all involve some sort of inhibitory processes to operate properly. For example, one could argue that the Updating function may require ignoring irrelevant incoming information and also suppressing no longer relevant information. Similarly, the Shifting function may require deactivating or suppressing an old mental set to switch to the new set. Although conceptually separable, this type of inhibition (what one might call inhibition or suppression of irrelevant or unnecessary information or mental sets) may be related to the deliberate, controlled inhibition of prepotent responses that we focused on in the reported study. Thus, all three target functions may share some inhibitory process, which in turn might have led to the moderate correlations among the three executive functions. Although this account is vague in terms of what the notion of ''inhibition'' really means, it deserves further investigation, given that the theoretical proposals that emphasize inhibition as a basic unit of working memory and executive control processes have become increasingly popular in the literature (e.g., Dempster & Corkill, 1999; Zacks & Hasher, 1994).

As for the diversity of executive functions, the results of the present study are also quite compatible with a substantial body of research noting the apparent neuropsychological or correlational dissociations of executive functions reviewed earlier. Moreover, the results also support recent theoretical attempts to fractionate the central executive (e.g., Baddeley, 1996; Baddeley & Logie, 1999) or the SAS (e.g., Stuss, Shallice, Alexander, & Picton, 1995), both of which tended to have a unitary flavor in their earlier conceptualizations.

One important question that needs to be considered regarding the diversity of executive functions is how best to classify separable executive functions. In this article, we have taken a rather pragmatic approach, focusing on three of the most frequently postulated functions in the literature. Our choice of the three functions was not arbitrary, however. We chose these functions because they seemed relatively basic (or at least more basic than prevalently mentioned higher level concepts like ''planning'') and have often been used to explain performance on complex executive tasks like the ones we examined in this study. The CFA and SEM results demonstrate that our strategy was successful and that examining the organization of executive functions at this level of analysis has merit, at least at this early stage of executive function research.

Despite this success, we are not claiming that the three investigated executive functions are the only executive functions, nor would we suggest that they are anything like the fundamental units or primitives of cognition. Our exploration of the diversity of executive functions is only a first step, and there are a number of important issues that need to be addressed in future research to better characterize the nature of separability or diversity of executive functions.

First, although our choice of the three target functions in this study seemed a reasonable one, it is certainly not exhaustive and there are other important relatively basic functions that need to be added to the current list. One such function, suggested by the SEM results for dual task performance, is the coordination of multiple tasks (Baddeley, 1996; Emerson, Miyake, & Rettinger, 1999), which may be somewhat separable from the three functions examined in this study. Second, the relationship between these relatively basic executive functions and more complex concepts like ''planning'' needs to be examined. If a combination of relatively basic functions can account for more complex executive functions (e.g., a combination of Shifting and Updating), then it helps make the classification of executive functions less chaotic. Finally, although the current level of analysis might be the most useful at this moment, it is also possible that the target functions we considered here can be decomposed into more basic component processes. Such a finer level of analysis seems to have been adopted by Stuss et al. (1995) in their effort to fractionate the SAS. Although this finer level of analysis faces the difficulty of selecting tasks that primarily tap one target process and hence may not lend itself readily to individual differences analyses, it appears a theoretically worthwhile approach to pursue.

In summary, although there are many more issues that need to be explored with respect to the organization of executive functions, the current results, together with some recent theoretical proposals (Duncan et al., 1996, 1997), help reconcile the controversy regarding the ''unitary versus nonunitary'' nature (or ''unity versus diversity'') of executive functions. A simple dichotomy will not suffice, and both aspects must be taken into account.

Implications of the Latent Variable Approach for Studying Executive Functions

In the reported study, we examined the organization and roles of three often postulated executive functions at the level of latent variables, rather than at the level of manifest variables. This latent variable approach has several important advantages over a more common approach of relying on zero-order correlations and EFA, particularly in the context of studying executive functions.

First, the latent variable approach can alleviate the task impurity problem, namely, that commonly used executive tasks are highly complex and typically place heavy demands on not just executive processes of interest, but also nonexecutive processes within which the executive processing requirement is embedded. The latent variable approach circumvents this problem by statistically ''extracting'' what is common across multiple tasks that all involve the same target processing requirement and analyzing the relationships among different executive functions in terms of these ''purer'' factors. Second, this approach may also help resolve another common problem in executive function research, namely the construct validity problem. Most proposals of what each executive task measures have tended to be speculative and not independently tested, but, as we demonstrated in our SEM analyses, the latent variable approach can provide a useful way to characterize the nature of specific executive functions implicated in complex executive tasks. These advantages are particularly important, given that the executive tasks are generally associated with low reliability (Denckla, 1996; Rabbitt, 1997b) and are thereby constrained to yield low intercorrelations. There may be too much error to test specific hypotheses or theoretical proposals if one analyzes the correlations at the level of individual tasks.

Our success in being able to ''extract'' common factors for each of the three latent variables is encouraging for executive function research, particularly in the context of a recent remark by Rabbitt (1997b): ''In our laboratory, we have been unable to find any commonality of individual differences in 'inhibition' between each of a wide variety of logically identical but superficially dissimilar Stroop-like tasks. That is, we can find no evidence that the ability to inhibit responses across a range of different tasks is consistently greater in some individuals than in others'' (pp. 12–13). At the level of zero-order correlations, the underlying commonality might not be obvious, but analyzing the data at the level of latent variables may increase the chances of revealing the common structure if it exists.

In addition to these advantages for individual differences studies of executive functions, the latent variable approach provides important implications for other lines of research on executive functions, including testing of brain-damaged patients in neuropsychological settings and neuroimaging studies of executive functions.

First, the results of this study suggest that it is important to systematically administer multiple executive tasks to understand the nature of sparing and impairments in a patient's executive functioning. Given that executive functions are separable and that different executive functions contribute differentially to various executive tasks, simply relying on prevalently used tasks like the WCST and TOH as general measures of executive functioning does not suffice. Although the generalizability of the current results to neuropsychological populations needs to be carefully evaluated first, it is important to be aware of the underlying separable functions and assess the patient's profile of executive functioning by taking into consideration which task taps which executive function(s) (see Miyake, Emerson, & Friedman, in press, for further discussion of the implications of the latent variable approach for clinical assessment).

Second, the current results also have interesting implications for neuroimaging studies of executive functions. So far, such studies have focused on complex executive tasks like the WCST, TOH, and RNG to examine the neural basis of executive functions, particularly the involvement of the frontal lobes in performance on those tasks. In almost all these cases, each study reports the pattern of brain activation observed for one executive task, localizing a set of specific brain regions that are considered important for certain processes (e.g., ‘‘planning’’ for TOH). However, one major danger of relying on just one task to infer the neural implementation of a specific target executive function is that, even though a clever subtraction method is used to isolate a target process, it is still possible that the isolated process includes other nonexecutive processes specific to that particular task. An interesting alternative is to consider multiple tasks (two or more) that are known to share the same underlying target process (perhaps as a result of an independent latent variable analysis study) and then examine the common regions of activation across these tasks. Although it may be more costly and time-consuming, this latent variable approach to neuroimaging may also help illuminate the degree of commonality or separability of different executive functions at the level of brain implementation or functioning.

APPENDIX A

Pearson Correlation Coefficients for the 15 Measures (N = 137 Unless Noted)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Plus–minus	—														
2. Number–letter	.32*	—													
3. Local–global	.23*	.32*	—												
4. Keep track	.23*	.08	.12	—											
5. Tone monitoring	.22*	.19*	.00	.15	—										
6. Letter memory	.24*	.11	.21*	.34*	.27*	—									
7. Antisaccade	.15	.17	.11	.12	.26*	.22*	—								
8. Stop-signal	.11	.13	.06	.10	.09	.04	.19*	—							
9. Stroop	.07	.09	−0.05	.11	.16	.18*	.20*	.18*	—						
10. WCST perseveration[a]	.26*	.13	.18*	.09	.19*	.14	.15	−0.01	.10	—					
11. TOH[b]	.08	.10	−0.09	.13	.18*	.14	.21*	.08	.17	−0.02	—				
12. RNG Component 1	.20*	.13	.01	.03	.11	.19*	.24*	.12	.11	.13	.10	—			
13. RNG Component 2	.20*	−0.07	.07	.29*	.06	.19*	.02	.18*	.01	−0.08	.12	.02	—		
14. Operation span	.09	.08	−0.04	.41*	.28*	.34*	.16	.13	.20*	.16	.04	.17*	.13	—	
15. Dual task[a]	−0.03	−0.02	.05	−0.09	−0.03	.12	−0.08	−0.16	.06	.06	−0.18*	−0.05	−0.09	−0.14	—

Note. WCST, Wisconsin Card Sorting Test; TOH, Tower of Hanoi; RNG, random number generation.
[a] $N = 134$.
[b] $N = 136$.
* $p < .05$.

APPENDIX B

Loadings for the Principal Components Analysis of 15 RNG Measures ($N = 137$)

	Component		
Measure	1	2	3
TPI	.92	−0.06	−0.15
A	.89	−0.14	−0.21
Runs	.86	−0.16	−0.01
RNG	.85	.16	.13
R	.06	.86	.22
Coupon score	−0.03	.81	.02
Mean RG	−0.05	.65	−0.19
Mode RG	−0.01	.53	−0.32
Phi6	.02	−0.52	.37
Phi7	.18	−0.48	.32
Phi3	.02	.01	.84
Phi4	−0.04	−0.22	.78
Phi2	−0.24	.00	.71
Phi5	−0.04	−0.34	.63
RNG2	.49	.36	.50
Correlations			
Component 1	—		
Component 2	.02	—	
Component 3	.04	−0.06	—

Note: An oblique Promax rotation was used to obtain estimates of correlations among the components. These three components were the only ones with eigenvalues larger than 1, and the Scree plot also suggested the three-component solution. TPI, turning point index; A, total adjacency; RNG, Evan's random number generation score; R, redundancy; mean RG, mean repetition gap; mode RG, mode repetition gap; phi2–7, phi indices; RNG2, analysis of interleaved digrams.

APPENDIX C

Factor Loadings for the Exploratory Principal Factor Analysis of the Shifting, Updating, and Inhibition Tasks ($N = 137$)

	Factor		
Measure	1	2	3
Plus–minus	.40	.22	.05
Number–letter	.58	−0.11	.16
Local–global	.58	.07	−0.17
Keep track	−0.01	.58	−0.05
Tone monitoring	.01	.22	.35
Letter memory	.04	.57	.05
Antisaccade	.08	.07	.44
Stop–signal	.09	−0.09	.38
Stroop	−0.12	.09	.43
	Correlations		
Factor 1	—		
Factor 2	.39	—	
Factor 3	.30	.42	—

Note: An oblique Promax rotation was used to obtain estimates of interfactor correlations. These three factors were the only ones with eigenvalues larger than 1 in the unreduced correlation matrix, and the Scree plot also suggested the three-factor solution.

REFERENCES

Allport, A., & Wylie, G. Task switching, stimulus-response bindings, and negative priming. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII.* Cambridge, MA: MIT Press. [in press]

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin,* **103,** 411–423.

Anderson, S. W., Damasio, H., Jones, R. D., & Tranel, D. (1991). Wisconsin card sorting performance as a measure of frontal lobe damage. *Journal of Clinical and Experimental Neuropsychology,* **13,** 909–922.

Arnett, P. A., Rao, S. M., Grafman, J., Bernardin, L., Luchetta, T., Binder, J. R., & Lobeck, L. (1997). Executive functions in multiple sclerosis: An analysis of temporal ordering, semantic encoding, and planning abilities. *Neuropsychology,* **11,** 535–544.

Baddeley, A. D. (1986). *Working memory.* New York: Oxford Univ. Press.

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

Baddeley, A., Della Sala, S., Gray, C., Papagno, C., & Spinnler, H. (1997). Testing central executive functioning with a pencil-and-paper test. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 61–80). Hove, UK: Psychology Press.

Baddeley, A., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *Quarterly Journal of Experimental Psychology,* **51A,** 819–852.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multicomponent model. In

A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). New York: Cambridge Univ. Press.

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

Bourke, P. A., Duncan, J., & Nimmo-Smith, I. (1996). A general factor involved in dual-task performance decrement. *Quarterly Journal of Experimental Psychology,* **49A,** 525–545.

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

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

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review,* **97,** 404–431.

Casey, B. J., Trainor, R. J., Orendi, J. L., Schubert, A. B., Nystrom, L. E., Giedd, J. N., Castellanos, F. X., Haxby, J. V., Noll, D. C., Cohen, J. D., Forman, S. D., Dahl, R. E., & Rapoport, J. L. (1997). A developmental functional MRI study of prefrontal activation during performance of a go-no-go task. *Journal of Cognitive Neuroscience,* **9,** 835–847.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Conway, A. R. A., & Engle, R. W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory,* **4,** 577–590.

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

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior,* **19,** 450–466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review,* **3,** 422–433.

Dempster, F. N., & Corkill, A. J. (1999). Individual differences in susceptibility to interference and general cognitive ability. *Acta Psychologica,* **101,** 395–416.

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

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature,* **378,** 279–281.

Dunbar, K., & Sussman, D. (1995). Toward a cognitive account of the frontal lobe function: Simulating frontal lobe deficits in normal subjects. *Annals of the New York Academy of Sciences,* **769,** 289–304.

Duncan, J. (1995). Attention, intelligence, and the frontal lobes. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 721–733). Cambridge, MA: MIT Press.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology,* **30,** 257–303.

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

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Emerson, M. J., Miyake, A., & Rettinger, D. A. (1999). Individual differences in integrating and coordinating multiple sources of information. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **25,** 1300–1321.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999a). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). New York: Cambridge Univ. Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999b). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General,* **125,** 309–331.

Everling, S., & Fischer, B. (1998). The antisaccade: A review of basic research and clinical studies. *Neuropsychologia,* **36,** 885–899.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods,* **4,** 272–299.

Godefroy, O., Cabaret, M., Petit-Chenal, V., Pruvo, J.-P., & Rousseaux, M. (1999). Control functions of the frontal lobes: Modularity of the central-supervisory system? *Cortex,* **35,** 1–20.

Goel, V., & Grafman, J. (1995). Are the frontal lobes implicated in ‘‘planning’’ functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia,* **33,** 623–642.

Goldman-Rakic, P. S. (1996). The prefrontal landscape: Implications of functional architecture for understanding human mentation and the central executive. *Philosophical Transactions of the Royal Society of London,* **351,** 1445–1453.

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research,* **18,** 1279-1296.

Hegarty, M., Shah, P., & Miyake, A. Constraints on using the dual-task methodology to specify the degree of central executive involvement in cognitive tasks. *Memory & Cognition.* [in press]

Hu, L.-T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Thousand Oaks, CA: Sage.

Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods,* **3,** 424–453.

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

Jahanshahi, M., Profice, P., Brown, R. G., Ridding, M. C., Dirnberger, G., & Rothwell, J. C. (1998). The effects of transcranial magnetic stimulation over the dorsolateral prefrontal cortex on suppression of habitual counting during random number generation. *Brain,* **121,** 1533–1544.

Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology,* Whole No. 89.

Jonides, J., & Smith, E. E. (1997). The architecture of working memory. In M. D. Rugg (Ed.), *Cognitive neuroscience* (pp. 243–276). Cambridge, MA: MIT Press.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications* (2nd ed.). Chicago: SPSS.

Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model-comparison approach.* San Diego, CA: Harcourt Brace Jovanovich.

Kiefer, M., Marzinzik, F., Weisbrod, M., Scherg, M., & Spitzer, M. (1998). The time course of brain activations during response inhibition: Evidence from event-related potentials in a go/no go task. *NeuroReport,* **9,** 765–770.

Kimberg, D. Y., D’Esposito, M., & Farah, M. J. (1997). Effects of bromocriptine on human subjects depend on working memory capacity. *NeuroReport,* **6,** 3581–3585.

Kimberg, D. Y., & Farah, M. J. (1993). A unified account of cognitive impairments following frontal lobe damage: The role of working memory in complex, organized behavior. *Journal of Experimental Psychology: General,* **122,** 411–428.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford.

Kok, A. (1999). Varieties of inhibition: Manifestations in cognition, event-related potentials and aging. *Acta Psychologica,* **101,** 129–158.

Larson, G. E., Merritt, C. R., & Williams, S. E. (1988). Information processing and intelligence: Some implications of task complexity. *Intelligence,* **12,** 131–147.

Lehto, J. (1996). Are executive function tests dependent on working memory capacity? *Quarterly Journal of Experimental Psychology,* **49A,** 29–50.

Legree, P. J., Pifer, M. E., & Grafton, F. C. (1996). Correlations among cognitive abilities are lower for higher ability groups. *Intelligence,* **23,** 45–57.

Levin, H. S., Culhane, K. A., Hartmann, J., Evankovich, K., Mattson, A. J., Harward, H., Ringholz, G., Ewing-Cobbs, L., & Fletcher, J. M. (1991). Developmental changes in performance on tests of purported frontal lobe functioning. *Developmental Neuropsychology,* **7,** 377–395.

Levin, H. S., Fletcher, J. M., Kufera, J. A., Harward, H., Lilly, M. A., Mendelsohn, D., Bruce, D., & Eisenberg, H. M. (1996). Dimensions of cognition measured by the Tower of London and other cognitive tasks in head-injured children and adolescents. *Developmental Neuropsychology,* **12,** 17–34.

Logan, G. D. (1985). Executive control of thought and action. *Acta Psychologica,* **60,** 193–210.

Logan, G. D. (1994). On the ability to inhibit thought and action: A user’s guide to the stop signal paradigm. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 189–239). San Diego, CA: Academic Press.

Lowe, C., & Rabbitt, P. (1997). Cognitive models of aging and frontal lobe deficits. In P. Rabbitt (Ed.), *Methodology of frontal and executive functions* (pp. 39–59). Hove, UK: Psychology Press.

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

Lyon, G. R., & Krasnegor, N. A. (Eds.). (1996). *Attention, memory, and executive function.* Baltimore: Brookes.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika,* **57,** 519–530.

Miyake, A., Emerson, M. J., & Friedman, N. P. Assessment of executive functions in clinical settings: Problems and recommendations. *Seminars in Speech and Language.* [in press]

Miyake, A., & Shah, P. (1999). Toward unified theories of working memory: Emerging general consensus, unresolved theoretical issues, and future research directions. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 442–481). New York: Cambridge Univ. Press.

Monsell, S. (1996). Control of mental processes. In V. Bruce (Ed.), *Unsolved mysteries of the mind: Tutorial essays in cognition* (pp. 93–148). Hove, UK: Erlbaum.

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology,* **81,** 111–121.

Morris, R. G., Miotto, E. C., Feigenbaum, J. D., Bullock, P., & Polkey, C. E. (1997). The effect of goal-subgoal conflict on planning ability after frontal- and temporal-lobe lesions in humans. *Neuropsychologia,* **35,** 1147–1157.

Moulden, D. J. A., Picton, T. W., Meiran, N., Stuss, D. T., Riera, J. J., & Valdes-Sosa, P. (1998). Event-related potentials when switching attention between task-sets. *Brain and Cognition,* **37,** 186-190.

Murji, S., & DeLuca, J. W. (1998). Preliminary validity of the cognitive function checklist: Prediction of Tower of London performance. *Clinical Neuropsychologist,* **12,** 358–364.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology,* **9,** 353–383.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol. 4, pp. 1–18). New York: Plenum.

O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically based computational model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 375–411). New York: Cambridge Univ. Press.

Ozonoff, S., & Strayer, D. L. (1997). Inhibitory function in nonretarded children with autism. *Journal of Autism and Developmental Disorders,* **27,** 59–77.

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

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

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

Posner, M. I., & Raichle, M. E. (1994). *Images of mind.* New York: Sci. Am.

Rabbitt, P. (Ed.). (1997a). *Methodology of frontal and executive function.* Hove, UK: Psychology Press.

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

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

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., Lawrence, A. D., McInnes, L., & Rabbitt, P. M. A. (1998). A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: Implications for theories of executive functioning and cognitive aging. *Journal of the International Neuropsychological Society,* **4,** 474–490.

Roberts, R. J., Hager, L. D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General,* **123,** 374–393.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General,* **124,** 207–231.

Rogers, R. D., Sahakian, B. J., Hodges, J. R., Polkey, C. E., Kennard, C., & Robbins, T. W. (1998). Dissociating executive mechanisms of task control following frontal lobe damage and Parkinson's disease. *Brain,* **121,** 815-842.

SAS Institute (1996). *SAS/STAT software: Changes and enhancements through release 6.11.* Cary, NC: SAS Inst.

Schachar, R. J., Tannock, R., & Logan, G. (1993). Inhibitory control, impulsiveness, and attention deficit hyperactivity disorder. *Clinical Psychology Review,* **13,** 721–739.

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

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

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology,* **7,** 268–288.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science,* **283,** 1657–1661.

Spector, A., & Biederman, I. (1976). Mental set and mental shift revisited. *American Journal of Psychology,* **89,** 669–679.

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

Stuss, D. T., & Benson, D. F. (1986). *The frontal lobes.* New York: Raven Press.

Stuss, D. T., Eskes, G. A., & Foster, J. K. (1994). Experimental neuropsychological studies of frontal lobe functions. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 9, pp. 149–185). Amsterdam: Elsevier Science.

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

Teuber, H.-L. (1972). Unity and diversity of frontal lobe functions. *Acta Neurobiologiae Experimentalis,* **32,** 615–656.

Towse, J. N. (1998). On random generation and the central executive of working memory. *British Journal of Psychology,* **89,** 77–101.

Towse, J. N., Hitch, G. J., & Hutton, U. (1998). A reevaluation of working memory capacity in children. *Journal of Memory and Language,* **39,** 195–217.

Towse, J. N., & Neil, D. (1998). Analyzing human random generation behavior: A review of methods used and a computer program for describing performance. *Behavior Research Methods, Instruments, & Computers,* **30,** 583–591.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language,* **28,** 127–154.

Van der Linden, M., Collette, F., Salmon, E., Delfiore, G., Degueldre, C., Luxen, A., & Franck, G. (1999). The neural correlates of updating information in verbal working memory. *Memory,* **7,** 549–560.

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology,* **7,** 131–149.

Yntema, D. B. (1963). Keeping track of several things at once. *Human Factors,* **5,** 7–17.

Zacks, R. T., & Hasher, L. (1994). Directed ignoring: Inhibitory regulation of working memory. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 241–264). San Diego, CA: Academic Press.

Accepted December 2, 1999