1990

Validation of Executive Function Tasks Through a Multitrait-Multimethod Approach

Judith Alison Gnys
University of Rhode Island

Follow this and additional works at: <https://digitalcommons.uri.edu/theses>

Recommended Citation

Gnys, Judith Alison, "Validation of Executive Function Tasks Through a Multitrait-Multimethod Approach" (1990). *Open Access Master's Theses*. Paper 1579.
<https://digitalcommons.uri.edu/theses/1579>

This Thesis is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Master's Theses by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu.

BF432
C48
G53
1990

VALIDATION OF EXECUTIVE FUNCTION TASKS
THROUGH A MULTITRAIT-MULTIMETHOD APPROACH

BY
JUDITH ALISON GNYS

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTERS OF ARTS
IN
PSYCHOLOGY

UNIVERSITY OF RHODE ISLAND

1990

22743-50

Abstract

A multitrait-multimethod paradigm was used to assess the stability, discriminant validity, and convergent validity of two tasks of executive functioning within a sample of 96 normal preschoolers. Memory served as the discriminant construct against which the executive-function construct was compared, and each construct was measured verbally and nonverbally. A matrix of cross-validated, Pearson product-moment correlation coefficients was computed to evaluate the psychometric properties of the executive-functioning tasks. Results did not support the construct validity of the selected tasks, as measures of executive functioning in this preschool population. Instead, the patterns of correlation coefficients derived suggested that the methods and traits explained nearly equal amounts of variance in the measures. Limitations of the current study and implications for future research are discussed.

Acknowledgements

I would like to take this opportunity to thank my thesis committee members, Dr. W. Grant Willis, Major Professor; Dr. Janet Kulberg; Dr. Dominic Valentino; and Dr. James Loy for thier time and effort in helping me complete this study. Special thanks to Dr. Grant Willis for giving me so much of his time and support. I also would like to thank Donald Round and Susan Helm for assisting me with the data collection. Also, a special thanks to all the children who participated and made this study possible.

Table of Contents

	Page
Abstract.....	ii
Acknowledgements.....	iii
Table of Contents.....	iv
List of Tables.....	vi
Introduction.....	1
Definition.....	1
Research Evidence.....	1
Adults.....	2
Children.....	4
Measurement of Executive Functioning.....	6
Psychometric Issues and Clinical Implications.....	6
Tasks.....	8
Tower of Hanoi.....	8
Verbal Fluency.....	8
Multitrait-Multimethod Validation.....	9
Hypotheses.....	10
Summary.....	11
Method.....	13
Subjects.....	13
Tasks.....	13
Procedure.....	18

Results.....	1 8
Discussion.....	2 4
Evaluation of the Matrix.....	2 4
Reliability.....	2 4
Validity.....	2 5
Convergent Validity.....	2 5
Discriminant Validity.....	2 5
Alternative Propositions.....	2 6
Criticisms.....	2 9
Implications.....	3 0
Summary.....	3 1
Appendix	3 3
Bibliography.....	3 6

List of Tables

	Page
1. Two by Two Task Matrix.....	13
2. Means and Standard Deviations for Selected Tasks.....	19
3. Multitrait- Multimethod Matrix.....	19
4. Cross-Validated, Adjusted, Multitrait-Multimethod Matrix	21
5. Pair-wise Comparisons.....	23

In this section, the construct executive function is defined and research evidence supporting executive functioning within normal and clinical, adult and child samples is reviewed. Next, selected tasks that have been used as measures of executive functioning are discussed along with implications for reliable and valid measurement. Finally, the purpose of the study is explained and hypotheses and predictions are made.

Definition

Executive function involves maintaining an appropriate set in order to achieve a future goal (Luria, 1973; Shallice, 1982). This involves strategic planning, impulse control, and organized search, as well as flexibility of thought and action (Welsh, Pennington, & Grossier, in press). Clinical evidence suggests that executive functions may be subserved by prefrontal cortical regions of the brain, given the relationship of these functions to the modulation of behavior through selective attention, organization, and the synthesis of sensory with motor functions (Becker, Isaac, & Hynd, 1987).

Research Evidence

Despite a substantial amount of neuropsychological and cognitive research about executive functions within adult samples, there is relatively little information available about executive functioning within child samples. Until recently, the nature of prefrontally-mediated function during childhood has eluded investigation. The majority of information about executive function in children, therefore, has been derived from clinical case studies of adults who have sustained frontal lobe damage (Luria, 1973). There have been some recent attempts to operationalize prefrontal

functioning in children, but these attempts have concentrated on adult neuropsychological measures with little adaptation for the child's developmental level (Fletcher & Taylor, 1984).

Adults

The majority of the information regarding executive functioning in humans derives from clinical case studies of adults who have sustained prefrontal lobe damage (Luria, 1973). Thus, specific knowledge of prefrontal-lobe functions has remained relatively limited and hypotheses concerning these functions remain controversial (Stuss & Benson, 1984). Some authors have credited prefrontal-lobe association cortex with the highest intellectual functions, whereas others have not been able to confirm a role for this region of the brain (Teuber, 1964). Behavioral investigations of prefrontal-damaged adults (as well as experimentally-lesioned animals) converge on the notion that the cognitive domain of executive function is most adversely impaired. For example, adults who sustained damage to the prefrontal regions of the brain exhibited impairments in planning, self-monitoring, achieving future goals, and inflexibility of thoughts and actions (Damasio, 1985; Luria, 1973). Similarly, monkeys with prefrontal damage had difficulty maintaining a goal-oriented set on delayed-response tasks (Goldman-Rakic, 1987). These kinds of studies have led many investigators to speculate that the prefrontal regions of the adult brain subserve a variety of executive-function cognitive behaviors (e.g., Kolb & Whishaw, 1980).

For example, neuroanatomical investigations have shown that the prefrontal cortex is bidirectionally linked with the limbic and

reticular activating systems, the posterior cortex, and the motor regions within the frontal lobes themselves (Goldman-Rakic, 1987; Stuss & Benson, 1984). Anatomically, such an arrangement implies a system that may provide regulatory control over the perceptual and attentional functions subserved by subcortical structures. The importance of the frontal lobes derives from their rich afferent and efferent interconnections with most other aspects of the central nervous system (Stuss & Benson, 1984). Thus, given its abundant interconnections, it is probably inappropriate to assign specific behaviors to localized sites within the prefrontal lobes (Passler, Isaac, & Hynd, 1985). Here, Luria's (1966) concept of a system specialized for the programming, regulation, and verification aspects of cognition is widely used clinically to explain disorders that can arise from prefrontal lobe lesions (Shallice, 1982).

The control of motor responses, of course, also primarily is a prefrontal-lobe function. Thus, lesions in the prefrontal cortex can lead to perseveration and disinhibition (Luria, 1973). Luria (1973) studied motor impairments in patients who sustained prefrontal damage. Using a variety of motor tasks, he postulated two major kinds of motor disorders following prefrontal-lobe damage. The first, associated with lesions in pre-motor zones, resulted in problems executing dextrous movements such as drawing and disinhibition. The second resulted in inflexibility, that is, an inability to shift from one action to another. Luria observed that, subsequent to damage to the prefrontal cortex, patients were unable to form action plans and could not match response outcomes with the original intentions formulated by those plans. Moreover, these

patients performed poorly on tasks which required organization of a series of actions in response to verbal directives (Drewe, 1975). Basically, patients understood task requirements, but simply were unable to perform (Passler et al., 1985). Luria observed that these patients no longer could control their behavior through the speech of either themselves or of others.

Children

Normal samples. Research in developmental psychology has demonstrated that executive function skills emerge in infancy and childhood. For example, Bruner (1973) found evidence of anticipation and planning in infants, as well as impulse control, self-monitoring, and set maintenance in toddlers. In contrast, Luria (1966) stated that the prefrontal regions of the brain do not develop in children until around age four through seven years. Similarly, Golden (1981) suggested that the frontal lobes do not mature until around age 12.

Passler, Isaac, and Hynd (1985) were among the first to attempt to study this issue empirically. These investigators adapted adult neuropsychological measures of putative prefrontal functions (see Luria, 1973) for use with school-age children. Results of their cross-sectional research showed that children between the ages of 6 and 12 were able to perform behavioral tasks putatively subserved by the prefrontal lobes, with varying degrees of success. Here, age-related changes were observed in executive-function tasks.

Similarly, Becker, Hynd, and Isaac (1988) investigated the development of children's nonverbal abilities to regulate and to inhibit motor action on various neuropsychological measures.

Consistent with Passler et al. (1985) these investigators also observed age-related changes in behaviors putatively subserved by prefrontal lobes.

Welsh, Pennington, and Grossier (in press) also investigated performance on a battery of executive function tasks as a function of age. These investigators hypothesized that given developmentally-appropriate behavioral measures, rudimentary forms of prefrontal skills would be exhibited in young children. Welsh et al. (in press) selected a battery of six tasks that had documented sensitivity to frontal lobe dysfunction in adults (Visual Search, Verbal Fluency, Wisconsin Card Sorting, and Motor Sequencing). From the literature in developmental psychology, they also selected the Tower of Hanoi and the Matching Familiar Figures Test. Results were consistent with Becker et al. (1987) and with Passler et al. (1985) in supporting the hypothesis that early emerging prefrontal skills solidified in a stagelike fashion throughout childhood.

Clinical samples. Executive functioning also has been investigated in a number of child clinical samples. For example, Welsh, Pennington, Ozonoff, Rouse, and McCabe (1987) hypothesized that children with early-treated phenylketonuria (PKU) would exhibit impairments on a battery of tasks designed to measure executive functioning. Here, it was proposed that characteristic, mild elevations in phenylalanine would lead to lower central levels of biogenic amines, including dopamine (DA). Given the localization of the dopamine pathways, this mild DA depletion would be expected to be associated with subtle prefrontal dysfunctions, which in turn might affect executive functions such as set maintenance, planning,

and organized search. Welsh et al. evaluated 11, early-treated PKU preschoolers and a sample of matched, unaffected peers. Results supported the hypothesized specific deficit in executive functions. The PKU children demonstrated significant impairments on the tasks purported to measure these skills (Visual Search, Verbal Fluency, Motor Planning, and Tower of Hanoi).

Behaviors associated with prefrontal-lobe functioning also have been found in children with Attention-deficit Disorder (ADD). Chelune, Ferguson, Koon, and Dickey (1986), for example, investigated the similarities and differences in neuropsychological performance between a group of children diagnosed as ADD and a group of matched controls. In particular, these investigators were interested in whether the ADD children would differ on tasks (Wisconsin Card Sorting, Progressive Figures, and Color Forms) used in the evaluation of adult patients with frontal lobe lesions. Results revealed deficits for the ADD children, but not the matched controls on tests presumed to measure prefrontal-lobe inhibitory control. In addition, age trends were noted on several variables. These investigators suggested a possible maturational lag in prefrontal-lobe functioning among the ADD subjects.

Measurement of Executive Functioning

Psychometric Issues and Clinical Implications

Although a number of neuropsychological studies have used tasks with documented sensitivity to prefrontal-lobe dysfunction in adult samples, as well as in some child samples, the validity and reliability of these tasks as measures of executive functioning in children has not yet been established (Welsh & Pennington, 1988).

Clinically, neuropsychologists typically evaluate child patients with standardized measures originally validated using brain damaged adults. Thus, most of these measures have not been designed in consideration of potential developmental discontinuities. Moreover, normative data also are lacking (Welsh et al. 1988; Fletcher & Taylor, 1984).

There are a number of problems in selecting executive-function tasks for children that are based on pathognomonic signs yielded by adult performances (Fletcher & Taylor, 1984). For example, the assumption that procedures employed with adults are differentially sensitive to brain disorders in children is not necessarily valid. Additionally, tests developed for adults do not necessarily measure the same abilities in children. Presently, neuropsychological organization in children remains unclear. Some advances in child neuropsychology (cf. Hynd & Willis, 1988; Tramontana & Hooper, 1988; Welsh & Pennington, 1988), however, suggest important potential differences in brain-behavior relationships for children versus adults. Clearly, basing assessments of children on scaled-down versions of adult assessment tools is, at best, a tenuous practice. Unfortunately, many currently available assessment tools in child neuropsychology largely represent downward extensions of adult neuropsychological tasks, many of which lack any clear linkage with the neuropsychology of the developing brain. Consequently, there is a limited range of well-validated assessment methods and tasks designed specifically for children (Tramontana, 1988).

Tasks

There are a number of tasks, however, that seem particularly promising for child neuropsychological assessments, given their theoretical foundation. Two of these are the Tower of Hanoi and Verbal Fluency.

Tower of Hanoi. The Tower of Hanoi has been used in a number of studies for a variety of reasons. For example, Shallice (1982) found that adults with prefrontal-lobe damage were deficient on this task. Simon (1975) used the Tower of Hanoi to study the relationship between cognitive strategies and perceptual/memory demands. Borys, Spitz, and Dorans (1982) used the Tower of Hanoi to study problem-solving abilities in nonretarded children and mentally deficient young adults. Similarly, Byrnes, and Spitz (1977) used the Tower of Hanoi to study problem-solving abilities in retarded and nonretarded children.

More recently, the Tower of Hanoi has been used to measure executive functioning. For example, Welsh et al. (in press) studied normative performance of children on a number of prefrontal measures, including the Tower of Hanoi. Welsh et al. found that normal children's performance on the Tower of Hanoi was indistinguishable from adult performance by age six. Finally, Welsh et al. (1987) also used the Tower of Hanoi as a measure of prefrontal functioning to see if PKU children would exhibit specific impairments on this task. Here, the performance of the PKU children was significantly more impaired than that of the matched controls.

Verbal Fluency. The second task, although often adapted, also has been used in previous research. For example, Milner (1963)

found that adults with prefrontal-lobe damage exhibited deficits on the Thurstone Word Fluency Test. Similarly, Jones-Goitman and Milner (1977) tested 100 patients with cortical excisions on a nonverbal design fluency task. Results showed that the patients with right prefrontal lesions were significantly impaired on these tasks. More recently, verbal fluency tasks adapted from the McCarthy Scales for Children's Abilities (1972) have been used to measure executive functioning in normal and early-treated PKU children (Welsh et al., 1987; Welsh, et al., in press). Welsh et al. (1987) found that PKU children exhibited more set-maintenance deficits (i.e., perseveration) on the Verbal Fluency task than their matched controls. Moreover, for the normal children, adult-level performance on the Verbal Fluency task continued to improve throughout adolescence.

Multitrait-Multimethod Validation

The purpose of the present study was to help establish the reliability, (i.e., stability) convergent validity, and discriminant validity of these selected tasks that have been used as measures of executive functioning for children. A multitrait-multimethod design (Campbell & Fiske, 1959) was used for this purpose. This kind of design involves correlating performance on tasks that measure various traits through various methods. Reliability can be assessed through correlations of tasks of the same trait measured with the same method (homotrait-homomethod). Convergent validity can be assessed through correlations of tasks of the same trait measured with different methods (homotrait-heteromethod). Discriminant validity can be assessed through two sets of correlations: first,

through correlations of tasks of different traits measured with the same method (heterotrait-homomethod) and second, through correlation of tasks of different traits measured with different methods (heterotrait-heteromethod). Thus, this kind of design yields separate sets of correlations which represent (a) reliability; (b) convergent validity; (c) discriminant validity, heterotrait-homomethod; and (d) discriminant validity, heterotrait-heteromethod.

Several aspects of the multitrait-multimethod matrix address issues of validity (Campbell & Fiske, 1959). First, validation is typically convergent, that is, it is established through confirmation by independent measures. Second, discriminant validity is required to justify novel trait measures, to validate test interpretation, and to establish construct validity. Third, each task used for measurement purposes is a trait-method unit which represents a union of a particular trait content with measurement procedures not specific to that content. Thus, error variance associated with a task can be due to the method of measurement as well as to the trait which was measured. Finally, in order to examine discriminant validity and to estimate the contribution of trait and method variance, Campbell and Fiske recommended that more than one trait and more than one method must be used in the validation process.

Hypotheses

Three basic assumptions were made about the multitrait-multimethod design. First, it was assumed that the multitrait-multimethod approach of assessing reliability and validity was a valid one. Second, it was assumed that the constructs of executive

function and memory were relatively independent. Finally, it was assumed that verbal and nonverbal measurement methods were independent.

The major hypothesis of this investigation was that the stability and construct validity of Verbal Fluency and Tower of Hanoi as measures of executive functioning in preschoolers could be evaluated through a multitrait-multimethod paradigm. Therefore, it was predicted that support for the validity of the selected tasks of executive functioning would be provided by a rank ordering of correlation coefficients as follows: reliability > convergent validity > discriminant validity, heterotrait-homomethod > discriminant validity, heterotrait-heteromethod (with convergent validity significantly greater than zero). Homotrait-heteromethod correlations that were significantly greater than zero would provide evidence for convergent validity. Homotrait-heteromethod correlations that were significantly greater than their corresponding heterotrait-heteromethod correlations would provide evidence for discriminant validity. Heterotrait-homomethod correlations would provide evidence of method variance within component measures (Campbell & Fiske, 1959). Deviations from this order would suggest different degrees of poor validity for the tasks (e.g., shared method variances).

Summary

In summary, the construct executive functioning describes the cognitive ability to maintain an appropriate problem-solving set for attaining a future goal. This involves strategic planning, impulse control, and organized search, as well as flexibility of thought and

action. Despite a substantial amount of neuropsychological and cognitive research about executive functioning within adult samples, there is relatively little information currently available about executive functioning within child samples. In addition, despite its importance as a psychological construct, there have been few investigations which have evaluated reliability and validity of tasks used to assess executive functions in children. The present study addressed this issue. A multitrait-multimethod design was used to investigate the reliability and construct validity of verbal and nonverbal tasks of executive function.

Method

Subjects

Participants were 57 white males and 39 white females enrolled in preschool and kindergarten classes in Rhode Island (M age = 63 months; SD age = 3.9 months). None of the children were experiencing any learning difficulties, as reported by their teachers. Written permission from each child's parent or guardian was secured prior to participation (see Appendix A). All participants were treated in accordance with ethical standards adopted by the American Psychological Association (APA, 1981) and by the University of Rhode Island Institutional Review Board (1989).

Task

Table 1 shows the task matrix used. The tasks comprised by this two-by-two (i.e., trait-by-method) matrix were selected based on empirical and theoretical criteria. For example, previous research on mean performance levels for different age groups, reliability and validity estimates for the memory tasks, and factor loadings for the nonverbal executive-function tasks were examined.

Table 1

Two-by-two Task Matrix

	Method 1 (Verbal)	Method 2 (Nonverbal)
Trait 1	Tower of Hanoi	Verbal Fluency
Trait 2	Memory for Sentences	BeadMemory

Verbal Fluency (executive functioning, verbal). This task, previously adopted as a measure of planning (Welsh, Pennington, & Grossier in press), required a systematic search of the semantic network. A semantic category (e.g., "colors") was spoken to the child who was then required to say as many words within that semantic category as possible within a 40 s time interval. The categories, adapted from the McCarthy Scales for Children's Abilities (MSCA) (McCarthy, 1972), were "food," "clothing," "animals," and "things to ride." The dependent variable was the total number of correct words said across four nonpractice trials.

Welsh et al. (in press) found that this task loaded most highly on a factor which was labeled, "speeded response." There was no information available about reliability or validity for this task. The average split-half reliability of the General Cognitive Index (GCI) from the MSCA as reported in the test manual, however, is .93, and the average split-half reliability for the Verbal Scale is reported as .88. The predictive validity of the GCI as indicated by correlations with performances on various achievement tests is reported as around .66, and the concurrent validity of the GCI with IQ scores from the Stanford-Binet Intelligence Scale, form L-M (Terman & Merrill, 1973), is reported as .81.

Tower of Hanoi (executive functioning, nonverbal). This ring-transfer task, previously adopted as a measure of planning capacities, required the child to plan a sequence of moves in order to duplicate a model (Borys, Spitz, & Dorans, 1982; Simon, 1975; Welsh et al., in press). An initial configuration of rings was translocated to

duplicate a goal configuration. Identical Tower of Hanoi apparatuses were assembled in front of the child and the investigator who sat facing each other on opposite sides of the table. Each apparatus consisted of a 48 cm by 14 cm rectangular base with three, 18 cm vertical dowels spaced 11.5 cm apart. Three, donut-shaped rings (8, 10, and 12 cm, colored red, yellow, and blue, respectively) were placed on the dowels.

The investigator's rings were arranged on the investigator's right-most dowel to form an inverted V-shaped tower. This arrangement represented the goal configuration that the child was required to achieve on all trials. Rings could only be moved according to the following rules: (a) a larger ring could not be placed on a smaller ring, (b) only one ring could be moved at a time, and (c) each ring either had to be on a dowel or in the child's hand at all times. In order to simplify the task explanation, the child was told a story about a family of monkeys (daddy = large ring, mommy = midsized ring, baby = small ring) jumping on trees (three dowels) (Khlar & Robinson, 1981). All the participants demonstrated their understanding of the rules by completing both legal and illegal translocations, and by successfully completing two practice trials. Individual translocations were recorded by the investigator, who documented an exact record of the sequence of moves generated by each participant on each trial.

The dependent variable was a quality-of-planning score which reflected the number of trials required for two consecutive correct solutions of problems of varying levels of difficulty (Borys, Spitz, & Dorans, 1982; Welsh et al., in press). For example, a score of six was

assigned when a problem was solved correctly on Trials 1 and 2; a score of five for correct solutions on Trials 2 and 3; a score of four for Trials 3 and 4; a score of three for Trials 4 and 5; a score of two for Trials 5 and 6; and a score of zero for Trial 5 or no trials correct. (Consistent with Welsh et al., scores of one were not assigned.) Thus, the total score across the six problems ranged from 0 through 36.

Welsh et al. (in press) found that Tower of Hanoi loaded most highly on a factor labeled, "planning." There was no information available about the reliability or validity for this task.

Bead Memory (memory, nonverbal). This task, adapted from the Stanford-Binet Intelligence Scale, fourth edition (Thorndike, Hagen, & Sattler, 1986), is a nonverbal task of short-term memory (Sattler, 1988). The task required the child to remember and to duplicate a bead configuration. On Items 1 through 10, the child was shown one or two beads for 3 s, and then was required to point to the correct bead(s) on a photograph. On Items 11 through 42, the child was shown a photograph of a vertical stick on which from two to six variously colored and shaped beads had been placed. The photograph was removed after a 5 s exposure, and the child was required to reconstruct the model with the actual beads. Two successful practice trials (i.e., Items B and C) were required before the actual test.

All children began with Item 11 and were required to achieve a basal of four items at two consecutive levels. If the child failed to achieve the basal, the investigator administered earlier items (i.e., Items 10 and under) until a basal was attained. Once the basal was established, the child continued until three of four or all four items

at two consecutive levels were failed. The dependent variable was the total number of correct reconstructions across the nonpractice trials.

The average internal-consistency reliability estimate for Bead Memory, as reported in the test manual, is .87. Moreover, factor analytic findings suggest that Bead Memory contributes moderately to a factor labeled, "Nonverbal/Visualization," (Thorndike et al., 1986) at all ages (median factor loading = .36).

Memory for Sentences (memory, verbal). This subtest, also adapted from the Stanford-Binet Intelligence Scale, fourth edition (Thorndike, Hagen, & Sattler, 1986), is a verbal, short-term memory task (Sattler, 1988). The task required the child to repeat a sentence read by the investigator. Sentences ranged in length from 2 through 22 words. Two successful practice trials (i.e., Items S2 and S3) were required prior to the actual trials. All children began with Item 11 and were required to attain a basal of four items at two consecutive levels. If the child failed to achieve the basal, the investigator administered earlier items (i.e., Items 10 and under) until a basal was attained. Once the basal was established, the child continued until three of four or all four items at two consecutive levels were failed. The dependent variable was the total number of correct sentences across the nonpractice trials.

The average internal-consistency reliability estimate for Memory for Sentences, as reported in the test manual, is .89. Moreover, factor analytic findings suggest that Memory for Sentences contributes moderately to a factor labeled, "Verbal

Comprehension," (Thorndike et al., 1986) at all ages (median factor loading = .58).

Procedure

The standardized battery of four tasks was administered, in a counterbalanced order, twice to all children. Each child was individually tested by one of three investigators in one session lasting approximately 50 minutes. The four tasks, comprising two traits and two methods, were selected according to recommended guidelines for multitrait-multimethod construct validation paradigms (Campbell & Fiske, 1959). Thus, the several methods used to measure each trait were appropriate to the trait as conceptualized. Moreover, there was no prior evidence to suggest common method variance shared by the tasks selected

Results

This investigation comprised eight dependent variables which corresponded to performance on the the first and second administrations of the four tasks: Verbal Fluency, first administration (VF1) and Verbal Fluency, second administration (VF2); Tower of Hanoi, first administration (TOH1) and Tower of Hanoi second administration (TOH2); Bead Memory, first administration (BM1) and Bead Memory, second administration (BM2); and Memory for Sentences, first administration (MS1) and Memory for Sentences, second administration (MS2). Means and standard deviation of each of these variables are presented in Table 2.

Table 2

Descriptive Statistics

<u>First Administration</u>			<u>Second Administration</u>	
Task	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
TOH	21	6.31	28	5.51
VF	23	6.45	23	6.76
BM	12	5.29	14	5.0
MS	17	3.6	18	3.9

Interrelationships among all variables were analyzed using an matrix of Pearson product-moment correlation coefficients. This matrix is presented in Table 3.

Table 3

Multitrait-Multimethod Matrix

	TOH1	VF1	BM1	MS1	TOH2	VF2	BM	MS2
TOH1	1.0	.36***	.29**	.33**	.74***	.37***	.31**	.31**
VF1	.36***	1.0	.11	.30**	.31**	.74***	.21*	.36***
BM1	.29**	.10	1.0	.35***	.18	.26**	.68**	.46***
MS	.34***	.30**	.36***	1.0	.26**	.31*	.34***	.81***

* $p < .05$

** $p < .01$

*** $p < .001$

Next, all convergent and divergent validity coefficients were subjected to a cross-validation procedure. Typically, two independent sets of scores are used for this purpose: one for the screening sample, and the other for the calibration sample (Lord & Novick, 1968). In the present investigation, however, two dependent sets of scores were used: the first administration of the tasks (i.e., VF1, TOH1, BM1, and MS1) for the screening sample, and the second administration of the tasks (i.e., VF2, TOH2, BM2, and MS2) for the calibration sample.

Cross-validated coefficients were calculated by correlating calibration scores with scores predicted from simple linear-regression equations generated from screening scores. Predictors in the regression equations were selected in order to yield the lowest standard error of estimate for predictions. Here, the cross-validated, convergent-validity (homotrait-heteromethod) estimate for VF and TOH was calculated by correlating TOH2 with a prediction of TOH1 from VF1. The cross-validated, convergent-validity (homotrait-heteromethod) estimate for BM and MS was calculated by correlating MS2 with a prediction of MS1 from BM1. The cross-validated, discriminant validity (heterotrait-homomethod) estimate for VF and MS was calculated by correlating MS2 with a prediction of MS1 from VF1. The cross-validated, discriminant validity (heterotrait-homomethod) estimate for TOH and BM was calculated by correlating BM2 with a prediction of BM1 from TOH1. The cross-validated, discriminant validity (heterotrait-heteromethod) estimate for TOH and MS was calculated by correlating MS2 with a prediction of MS1 from TOH1. Finally, the cross-validated, discriminant

validity (heterotrait-heteromethod) estimate for VF and BM was calculated by correlating BM2 with a prediction of BM1 from VF1.

Next, in order to adjust these cross-validated validity estimates for unreliabilities associated with criterion scores, each coefficient was subjected to an adjustment procedure (Thorndike, 1982, p. 222). Adjusted coefficients were calculated by dividing cross-validated measures by the square roots of the associated test-retest reliability coefficients for their respective criterion measures. Table 4 shows the resulting matrix of cross-validated, adjusted correlation coefficients.

Table 4

Cross Validated, Adjusted, Multitrait-Multimethod Matrix

	VF	MS	TOH	BM
VF	(.74)			
MS	.44	(.81)		
TOH	.45	.31	(.72)	
BM	.30	.58	.35	(.68)

All coefficients are significantly different from zero ($p < .001$).

Finally, multiple pair-wise comparisons were made. This was done in order to assess the significance of the differences between the following estimates: (a) test-retest reliabilities for VF and for TOH versus convergent validity between VF and TOH, (b) convergent validity between VF and TOH versus discriminant validity (heterotrait-homomethod) between VF and MS and between BM and TOH, (c) convergent validity between VF and TOH versus discriminant (heterotrait-heteromethod) validity between TOH and MS and between VF and BM, (d) discriminant validity between BM and TOH and between VF and MS versus discriminant (heterotrait-heteromethod) validity between TOH and MS and between VF and BM, and (e) discriminant validity between TOH and MS and between VF and BM versus zero. In order to address the issue of family-wise type-I error rate (i.e., multiple significance tests), inferential t and z statistics for comparing elements of a correlation matrix (Steiger, 1980) were used. The overall (one-tailed) alpha level adopted was .05. Results indicated that test-retest reliability estimates significantly exceeded convergent validity estimates ($t_s = 3.56$ and 4.14 , $p < .001$ in both cases), but that convergent and discriminant validity estimates did not differ significantly ($t_s = .10$ to 1.44 , $p > .10$ in all cases). Moreover, heterotrait-heteromethod discriminant validity coefficients were significantly greater than zero ($z_s = 2.92$ and 3.02 , $p < .001$ in both cases) indicating that all reliability and validity estimates significantly exceeded zero (i.e., reliability > convergent validity = heterotrait-homomethod discriminant validity = heterotrait-heteromethod discriminant validity > 0). Table 5

presents the 12 comparisons made, inferential statistics calculated, and levels of significance.

Table 5

Pair-Wise Comparisons

Reliability vs convergent validity:

VF/VF vs. VF/TOH $t=4.14, p<.001$

TOH/TOH vs. VF/TOH $t=3.56, p<.001$

Convergent validity vs. discriminant (hetero-homo) validity:

VF/TOH vs. VF/MS $t=.10, NS$

VF/TOH vs. BM/TOH $t=.84, NS$

Convergent validity vs. discriminant (hetero-hetero) validity:

VF/ TOH vs. TOH/MS $t=.134, NS$

VF/TOH vs. VF/BM $t=.133, NS$

Discriminant (hetero-homo) validity vs discriminant (hetero-hetero) validity:

VF/MS vs. TOH/MS $t=1.44, NS$

VF/MS vs. VF/BM $t=1.39, NS$

TOH/BM vs. TOH/MS $t=.46, NS$

TOH/BM vs. VF/BM $t=.49, NS$

Discriminant (hetero-hetero) validity vs. 0:

TOH/MS vs. 0 $z=3.02, p<.001$

VF/BM vs. 0 $z=2.92, p<.001$

Discussion

The purpose of the present study was to help establish the reliability, convergent validity, and discriminant validity of two tasks that have been used as measures of executive functioning for children. A multitrait-multimethod design (Campbell & Fiske, 1959) was used for this purpose.

The major hypothesis of this investigation was that the stability and construct validity of Verbal Fluency and Tower of Hanoi as measures of executive functioning in preschoolers could be evaluated through a multitrait-multimethod paradigm. It was predicted that support for the construct validity of the selected tasks of executive functioning would be provided by a rank ordering of correlation coefficients as follows: reliability > convergent validity > heterotrait-homomethod discriminant validity > heterotrait-heteromethod discriminant validity > 0). It further was predicted that deviations from this order would suggest different degrees of poor validity for the tasks (e.g., shared method variances).

Evaluation of the Matrix

Reliability

Initially, reliability estimates were calculated for the executive function tasks. Here, the stability of Verbal Fluency over a brief time interval (approximately 25 min) within this preschool population was assessed as .74. This value signifies that 74% of the variance in the Verbal Fluency measure depended on true variance in the trait measured; conversely 26% depended on error variance (Anastasi, 1988). Similarly, the stability of Tower of Hanoi over the same brief time interval within this preschool population was

assessed as .71. Again, this value signifies that 71% of the variance in the Tower of Hanoi measure depended on true variance in the trait measured; conversely 29% depended on error variance.

Validity

In order to minimize the influence of psychometric error on the results of the present study, all validity coefficients were subjected to a cross-validation procedure as well as an adjustment for the unreliabilities associated with criterion scores (Campbell & Fiske, 1959; Lord & Novick, 1968; Pedhazur, 1982; Thorndike, 1982).

Convergent validity. The convergent validity of the executive-function construct was addressed by evaluating the relationship between component measures obtained by different methods, that is, the cross-validated, adjusted correlation coefficient for VF with TOH. The value of this coefficient was .45 ($p < .001$), indicating that approximately 20% of the variance was shared by these two measures.

Discriminant validity. The discriminant validity of the executive-function construct was assessed by evaluating the relationship between a trait measure with a similar measure of a different trait. According to Campbell and Fiske (1959) a variable used as a trait measure should correlate higher with an independent effort to measure the same trait than with measures of different traits using the same methods. In the present study, these comparisons were made by comparing the convergent validity coefficient (i.e., VF with TOH = .45) with the corresponding heterotrait-homomethod coefficient (i.e., BM with TOH = .35, and VF with MS = .44). Results of the present investigation indicated that

these comparisons were not significantly different, and therefore did not support the construct validity of VF and TOH as measures of executive functioning in this preschool sample.

Other evidence bearing on the discriminant validity of the executive-function construct was assessed by evaluating the relationship between a trait measure with a different measure of a different trait. Again, according to Campbell and Fiske (1959) a variable used as a trait measure should correlate higher with an independent effort to measure the same trait than with measures of different traits using different methods. In the present study, these comparisons were made by comparing the convergent validity coefficient (i.e., VF with TOH = .45) with the corresponding heterotrait-heteromethod coefficient (i.e., TOH with MS = .31, and VF with BM = .30). Again, results of the present investigation indicated that these comparisons were not significantly different, and therefore did not support the construct validity of VF and TOH as measures of executive functioning in this preschool sample. Instead, these results suggested that method of measurement accounted for a significant proportion of score variance (nearly equal to trait variance) in these measures.

Alternative Propositions

The present study did not support the construct validity of VF and TOH as measures of executive functioning in preschoolers. In these kinds of situations, Campbell and Fiske (1959) suggested that several alternative propositions be considered. First, perhaps neither the verbal nor the nonverbal method used was adequate for measuring the executive-function construct. In the present study,

however, there was adequate research to indicate that the memory tasks selected could be measured adequately through both verbal and nonverbal channels. Here, it has been found that BM contributed moderately to a Nonverbal/Visualization factor at all ages and that MS contributed moderately to a Verbal Comprehension factor at all ages (Sattler, 1988). Similarly, VF appeared as if it could be measured adequately by the method selected for the study. For example, VF loads on the Verbal Scale of the McCarthy Scales of Children's Abilities (McCarthy, 1972). Although TOH has been used as a nonverbal measure in previous research, there currently is no factor analytic research to support a nonverbal loading for this task (cf. Borys et al., 1982).

A second alternative proposition proposed by Campbell and Fiske (1959) is that one of the two methods does not really measure the trait. Given what is known about the memory tasks, it seems probable that they were adequately measured by both methods. It is possible, though, that for TOH and BM, one of the methods (perhaps the nonverbal one) was not adequate. For example, it is possible that the children could have used a verbal strategy to solve TOH and BM. Thus, the nonverbal method might not have been adequate for one or both of these tasks.

A third alternative is that the trait is not a functional unity. As noted previously, an assumption of the present study was that the two traits (i.e., memory and executive function) were independent constructs. It is possible, that these constructs may be interrelated to some extent. For example, both traits may be associated with a common neuropsychological substrate, perhaps involving prefrontal

cortical functions. A primary feature of prefrontal-lobe functioning is the ability to maintain a set for future oriented and goal directed activity. Fuster (1985) suggested that for a set to guide adaptive problem solving, it must rely on a working memory function that maintains the necessary stimulus information. For many tasks, a child must maintain an appropriate set over time. For many memory tasks, the child is required to remember intentionally, and to plan in order to accomplish a goal. It is likely that these intentional, goal-directed behaviors represent a form of executive function that transcend any one, underlying cognitive process such as memory.

Moreover, organization in memory refers to the process by which the individual attempts systematically to encode and to retrieve the information presented in order to maximize performance (Pelligrino & Ingram, 1979). This definition of memory reflects a bias toward viewing memory paradigms as problem-solving situations. In this instance, particular kinds of memory tasks can be considered in terms of problem solving. Here, the individual typically is given a goal, general rules typically are specified, and the individual is then required to attain the goal independently (Puff, 1979). In order to attain such a goal, the individual must have access to internalized strategies, which can ameliorate limitations on the storage and retrieval (i.e., memory) of information.

Another issue explored by researchers in problem solving has been the underlying mechanism of developmental advances (Welsh & Pennington, 1988). It has been proposed that the acquisition and execution of increasingly advanced executive strategies results from

a gradual increase in the size of the child's working memory (Case, 1974). Thus, the concept of working memory has to do with the on-line mental representation of information necessary to solve problems. In order for encoded information to be useful in obtaining a goal, it should be processed in a strategic and meaningful manner (Welsh & Pennington, 1988). It is this component of memory that younger children have problems with, thus limiting their effectiveness as problem solvers. Although young children often have the capabilities to do so, their performance in deliberate memory tasks frequently reflects a poor strategic deployment of these skills (Ornstein & Corsale, 1979).

Thus, although the traits of executive function and memory were assumed to be independent constructs at the outset of this research, results of the current investigation suggest that these constructs may be closely related. It is possible, for example, that the constructs of memory and executive function may not be separable neuropsychological traits.

Criticisms

There are a number of methodological issues that should be considered in interpreting this research. First, the selection of the executive function tasks could have affected the outcome. For example, it is possible that VF and TOH measured different aspects of executive functioning. Welsh et al. (in press) found that VF and TOH loaded on different factors, with TOH loading on a factor labeled, "Planning," and VF loading on a factor labeled, "Fluid and Speeded Response." Tasks loading on the latter factor required the child to maintain a set to guide behavior that was less complex and had

fewer distracting response alternatives than TOH. Thus, it is possible that the two tasks selected measured different aspects of executive functioning.

A second criticism of this study was the selection of the discriminant trait. As discussed, it is possible that memory and executive function are associated, thus confounding results, and leading to the rejection of the assumption that these constructs were independent.

A final criticism is related to the selection of the nonverbal tasks. It is possible that BM and TOH were solved through verbal strategies by the children. Other tasks might have minimized the use of verbal strategies to a greater extent than those used in the present study.

Implications

The results of this study did not support the convergent validity of VF and TOH as measures of executive functioning in preschoolers. In fact, the patterns of correlation coefficients indicated that the method and trait variances associated with these tasks were approximately equal. Thus, previous findings from research using these tasks should be interpreted cautiously. Moreover, future research should consider the role of memory when evaluating performance on these kinds of tasks. For example, if memory is a component of executive functioning, and if the acquisition and execution of increasingly advanced executive strategies results from a gradual increase in the size of working memory, then past research (e.g., Becker et al., 1988; Passler et al., 1985) documenting observed age-related changes on tasks designed

to measure executive functioning in children may be confounded with mnestic aptitudes. Previous studies have used VF and TOH as measures that putatively are sensitive to prefrontal lobe dysfunction in children (e.g., Welsh et al., in press). Clearly, results of the present investigation suggested that VF and TOH may not be robust measures, and therefore caution should be exercised in interpreting results in this manner.

Clinically, with the adoption of Public Law 99-457 (i.e., The Education of the Handicapped Act Amendments of 1986), there has arisen a concomitant need to expand the repertoire of reliable and valid assessment tasks available to psychologists and other psychoeducational specialists for preschoolers. Moreover, the clinical relevance of executive functioning to early psychoeducational treatment interventions suggests that this construct deserves attention at the preschool level. The present study was the first to assess systematically the test-retest reliability and construct validity of tasks designed to measure executive functioning in children. Future research needs to evaluate further these and other psychometric properties of such tasks. In addition, future research should address whether the components of executive function are separable from other cognitive and neuropsychological processes such as memory.

Summary

The primary goal of this study was to ascertain test-retest reliability and construct validity estimates for two tasks designed to measure executive functioning in preschoolers. This was accomplished through a multitrait-multimethod analysis. Four tasks

were used, two for each construct (i.e., executive functioning and memory) and two for each method (i.e., verbal and nonverbal). Results did not support the construct validity of the tasks selected. Instead, the patterns of correlation coefficients derived suggested that the methods and traits explained nearly equal amounts of variance in the measures. Thus, caution should be exercised in using these tasks for research, assessment, and intervention purposes.

Appendix A

Informed Consent

Dear Parent/Guardian:

A study is being conducted by Judith A. Gnys, a graduate student from the University of Rhode Island, supervised by Dr. Grant Willis. The purpose of this study is to improve our understanding of how children solve particular learning tasks.

This project essentially is an investigation of the validity of tasks commonly used to measure planning and problem-solving skills in children. Your child's individual performance will not be evaluated, rather what Judith will be evaluating are the tasks. Specifically, if the tasks actually measure what they are suppose to measure.

In this study, children will be observed as they complete four tasks twice. This will involve about 50 minutes during one day. Children typically enjoy the activities. There are no risks involved and results will be kept confidential.

Your permission for your child to participate would be greatly appreciated. In order to allow your child to participate, please sign and date this letter, and ask your child to bring it to his or her teacher as soon as possible. Please know that your permission is entirely voluntary and you are free to change your mind at any time.

If you have any questions, feel free to contact Judith Gnys at (401) 726-5244 or Dr. Grant Willis at (401) 792- 4245.

(Questions concerning rights as participant may be directed to the Institutional Review Board on Human Subjects, Graduate School, University of Rhode Island).

I have read the above letter and my child and I have agreed to his or her participation in the study described.

I understand that there are no risks involved and although there will be no direct benefit to my child, my child's participation will aide in the understanding of how children solve particular tasks.

I understand that my permission is entirely voluntary and that my child and I are free to change our mind at any time and withdraw from the study.

I understand that all results will be kept confidential and that my child's name will not be associated with his or her performance. I grant permission for _____ to participate in the study as described.

Signature of Parent Guardian

Child's Bithdate

Bibliography

- Anastasi, A. (1988). Psychological testing. New York: Macmillan.
- Becker, M. G., Isaac, W., & Hynd, G. (1988). Neuropsychological development of non-verbal behaviors attributed to "frontal lobe" functioning. Developmental Neuropsychology, 3, 275-298.
- Borys, S. V., Spitz, H. H., & Dorans, B. A. (1982). Tower of Hanoi performance of retarded young adults and nonretarded children as a function of solution length and goal state. Journal of Experimental Child Psychology, 33, 87-110.
- Bruner, J. (1973). Organization of early skilled action. Child Development, 44, 1-11.
- Byrnes, M. A., & Spitz, H. H. (1977). Performance of retarded adolescents and nonretarded children on the Tower of Hanoi problem. American Journal of Mental Deficiencies, 81, 561-569.
- Campbell, J. P., & Fiske, D. W. (1959). Convergent and discriminant validation by the multi-trait multi-method matrix. Psychological Bulletin, 56, 81-105.
- Chelune, G. J., Ferguson, W., Koon, R., & Dickey, T. O. (1986). Frontal lobe disinhibition in attention deficit disorder. Child Psychiatry and Human Development, 16, 221-234.
- Cowart, C. A., & McCallum, R. S. (1988). Multitrait-multimethod investigation of the simultaneous-successive-planning model. Journal of Psychoeducational Assessment, 6, 55-66.
- Damasio, A. R. (1985). The frontal lobes. In K. M. Heilman and E. Valenstein (Eds.), Clinical neuropsychology, (pp. 376-399). New York: Oxford University Press.
- Drewe, E. A. (1975). An experimental investigation of Luria's theory on the effects of frontal lesions in man. Neuropsychologia, 13, 421-429.
- Fletcher, J. M., & Taylor, H. G. (1984). Neuropsychological approaches to children: Towards a developmental psychology. Journal of Clinical Neuropsychology, 6, 39-56.

- Fuster, J. M. (1985). The pre-frontal cortex, mediator of cross-temporal contingencies. Human Neurobiology, 4, 169-179.
- Goitman-Jones, M., & Milner, B. (1976). Design fluency: The invention of nonsense drawings after focal cortical lesions. Neuropsychologia, 15, 653-674.
- Golden, C. J. (1981). The Luria-Nebraska Children's Battery: In G. W. Hynd & J. E. Obrzut (Eds.), Neuropsychological assessment and the school-aged child (pp. 227-302). New York: Grune & Stratton.
- Goldman-Rakic, P. S. (1987). Development of cortical circuitry and cognitive function. Child Development, 58, 601-622.
- Kolb, D., & Whishaw, I. Q. (1980). Fundamentals of human neuropsychology. San Francisco: W. H. Freeman.
- Lee, R., & Malone, M. (1981). Multitrait-multimethod-multirater analysis of performance ratings of law enforcement officials. Journal of Applied Psychology, 66, 625-632.
- Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wellsley.
- Luria, A. R. (1973). The working brain. New York: Basic Books.
- Luria, A. R. (1966). Higher cortical functions in man. London: Tavistock.
- McCarthy, D. A. (1972). Manual for the McCarthy Scales of Children's abilities. San Antonio: Psychological Corporation.
- Milner, B. (1982). Some cognitive effects of frontal-lobe lesions in man. Royal Society of London, B298, 211-226.
- Ornstein, P. A., & Corsale, K. (1979). Organizational factors in children's memory. In R. C. Puff (Ed), Memory organization and structure. (pp. 219-253). New York: Academic Press.
- Passler, M. A., Isaac, W., & Hynd, G. W. (1985). Neuropsychological development of behavior attributed to frontal lobe functioning in children. Developmental Neuropsychology, 1, 349-370.
- Pedazur, E. J. (1982). Multiple regression in behavioral research. New York: CBS College Publishing.
- Pellegrino, J. W., & Ingram, A. L. (1979). Processes, products, and measures of memory organization. In R. C. Puff (Ed), Memory organization and structure. New York: Academic Press.

- Puff, R. C. (1979). Memory organization and structure. New York: Academic Press.
- Sattler, J. M. (1988). Assessment of children. San Diego, CA: Jerome Sattler.
- Shallice, T. (1982). Specific impairments of planning. Royal Society of London, B298, 199-209.
- Simon, H. A. (1975). The functional equivalence of problem solving skills. Cognitive Psychology, 7, 268-288.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. Psychological Bulletin, 87, 245-251.
- Stuss, D. T., & Benson, F. (1984). Neuropsychological studies of frontal lobes. Psychological Bulletin, 95, 3-28.
- Terman, L. M., & Merrill, M. A. (1973). Stanford-binet intelligence scale: Manual for the third revision. Boston, MA: Houghton-Mifflin.
- Teuber, H. L. (1964). The riddle of frontal lobe functioning in man. In J. M. Warren, & K. Akert (Eds.), The frontal granular cortex and behavior. (pp. 410-444). New York: McGraw-Hill.
- Thorndike, R. L. (1982). Applied psychometrics. Boston, MA: Houghton-Mifflin.
- Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). Guide for administering and scoring the Stanford-Binet Intelligence Scale: Fourth Edition. Chicago: Riverside..
- Tramontana, M. G. & Hooper, S. R. (1988). Assessment issues in child neuropsychology. New York: Plenum.
- Tramontana, M. G. (1988). Problems and prospects in child neuropsychological assessment. In M. G. Tramontana & S. R. Hooper (Eds.), Assessment issues in child neuropsychology. New York: Plenum Press.
- Welsh, M. C., & Pennington, B. F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. Developmental Neuropsychology, 4, 199-230.
- Welsh, M. C., Pennington, B. F., & Grossier, B. F. (in press). A normative-developmental study of executive function: A window on prefrontal function in children.

Welsh, M. C., Pennington, B. F., & McCabe, E. (1987). Performance of preschool phenylketonyric (PKU) children on prefrontal measures. (Abstract) Journal of Clinical and Experimental Neuropsychology, 9, 28.

Welsh, M. C., Pennington, B. F., Ozonoff, S., Rouse, B., & McCabe, E. (manuscript submitted) Neuropsychology of early-treated PKU: Specific executive function defecits.