Psychology in the Schools, Vol. 47(7), 2010
Published online in Wiley InterScience (www.interscience.wiley.com)

DOI: 10.1002/pits.20500

HOW DO EXECUTIVE FUNCTIONS FIT WITH THE CATTELL–HORN–CARROLL MODEL? SOME EVIDENCE FROM A JOINT FACTOR ANALYSIS OF THE DELIS–KAPLAN EXECUTIVE FUNCTION SYSTEM AND THE WOODCOCK–JOHNSON III TESTS OF COGNITIVE ABILITIES

RANDY G. FLOYD AND RENEE BERGERON

The University of Memphis

GLORIA HAMILTON

Middle Tennessee State University

GILBERT R. PARRA

The University of Memphis

This study investigated the relations among executive functions and cognitive abilities through a joint exploratory factor analysis and joint confirmatory factor analysis of 25 test scores from the Delis–Kaplan Executive Function System and the Woodcock–Johnson III Tests of Cognitive Abilities. Participants were 100 children and adolescents recruited from general education classrooms. Principal axis factoring followed by an oblique rotation yielded a six-factor solution. The Schmid–Leiman transformation was then used to examine the relations between specific cognitive ability factors and a general factor. A variety of hypothesis-driven models were also tested using confirmatory factor analysis. Results indicated that all tests measure the general factor, and 24 tests measure at least one of five broad cognitive ability factors outlined by the Cattell–Horn–Carroll theory of cognitive abilities. These results, with limitations considered, add to the body of evidence supporting the confluence of measures of executive functions and measures of cognitive abilities derived from individual testing. © 2010 Wiley Periodicals, Inc.

In recent years, a number of assessment instruments measuring executive functions (e.g., Gioia, Isquith, Guy, & Kenworthy, 2000) and books targeting them (e.g., McClosky, Perkins, & Van Diviner, 2008) have been marketed to school psychologists and other professionals engaged in assessment of children and adolescents. Executive functions can be conceived as the set of cognitive processes that promote the organization of thought and behavior, but, like the term *intelligence*, its definitions vary substantially. For example, some equate executive functions with *self-regulation*, which seems to describe something much more general than a cognitive process. Eslinger (1996) offered the following definition of executive functions after reviewing chapters by experts devoted to conceptualizing and measuring them:

> Executive functions are defined as psychological processes that have the purpose of controlling implementation of activation–inhibition response sequences that is guided by diverse neural representations (verbal

---

The Woodcock–Muñoz Foundation provided financial support for the study. The opinions expressed in this manuscript do not necessarily reflect those of the Woodcock–Muñoz Foundation.

We are thankful for Dr. Susan Copeland, The University of Memphis Campus School, and other participating schools for allowing access to participants. We also thank Jillayne Bose-Deakins, Miriam Hancock, Jane Harland, Heather Hill, Tina Hutchison, Amanda Johnson, Allison Margulies, Connie Mazza, Angela Redwine, Samantha Shockley, Lisa Smith, Andrea Starkey, Kevin Steelman, Kimberly Sumara, Michael Wood, and Amy Yerka, who served as test examiners and Maria Haley Leforgee, who providing training for test examiners. Appreciation also goes to Kevin McGrew for a conversation about the nature of executive functions.

Portions of this research were presented at the annual meeting of the National Association of School Psychologists (2005).

Renee Bergeron is now affiliated with the Hawaii Department of Education.

Correspondence to: Randy G. Floyd, 202 Psychology Building, The University of Memphis, Department of Psychology, Memphis, TN 38152. E-mail: rgfloyd@memphis.edu

> rules, biological needs, somatic states, emotions, goals, mental models) for the purpose of meeting a balance of immediate situational, short-term, and long-term future goals that span physical-environmental, cognitive, behavioral, emotional, and social spheres (p. 381).

Certainly, executive functions appear to include an expansive and varied set of component processes that affect adaptive behaviors.

Research has suggested that deficits demonstrated during tests of executive functions are associated with a number of psychological or medical disorders, such as attention–deficit/hyperactivity disorder, oppositional defiant disorder, learning disabilities, Alzheimer's disease, and frontal lobe brain injuries (e.g., Pennington & Ozanoff, 1996; Willcutt et al., 2001; Zelazo & Mueller, 2002). Although this research suggests relations between executive functions and psychopathology, and this discriminative validity evidence is worthwhile, relatively little is known about the relations between executive functions and specific cognitive abilities and their measurement overlap. One may ask the following questions: Are executive functions and cognitive abilities distinct constructs? Do their effects on behavior differ? Does one term subsume the other? Do some methods of assessing these constructs most accurately represent them?

Some researchers have found that the child and adult versions of the Wechsler intelligence scales (e.g., Wechsler, 1974, 1981) contain subtests that measure the same factors as subtests from neuropsychological test batteries designed to assess purported components of executive functions, including working memory (Chittooran, D'Amato, Lassiter, & Dean, 1993; Leonberger, Nicks, Goldfader, & Munz, 1991; Leonberger, Nicks, Larrabee, & Goldfader, 1992), nonverbal or spatial reasoning (Leonberger et al., 1991, 1992; Sherman, Strauss, Spellacy, & Hunter, 1995), and attention and concentration (Leonberger et al., 1992). Although most of these studies did not consider the possibility that *the general factor* (Jensen, 1998) influenced all test scores across batteries, some researchers have found that both tests of executive functions and tests of cognitive abilities tap a general construct (e.g., Kelly, Arceneaux, Dean, & Anderson, 1994). The typically uniform positive relations between measures of cognitive abilities and measures of executive functions as well as the similarities between the descriptions of executive functions, some specific cognitive abilities, and the general factor have led researchers to suggest a confluence of these constructs (Crinella & Yu, 2000; Detterman, 1982; Sternberg & Gardner, 1983) and to question their distinctions. These linkages and distinctions have, however, too rarely been investigated.

It is notable that most of the studies employing factor-analytic techniques cited earlier in text employed the Wechsler intelligence scales to assess cognitive abilities—despite the fact that these scales do not appear to measure a full range of specific cognitive abilities as specified in recent factor-analytic research (Carroll, 1993). Consequently, little is known about the relations between executive functions and specific abilities such as Fluid Reasoning and Long-Term Storage Retrieval. Only recently has research included measures of executive functions and a wide range of ability measures based on recent theory and specified an array of theory-driven ability factors in analyses (Burns, Nettelbeck, & McPherson, 2009; Salthouse, 2005; Salthouse & Davis, 2006).

The central purpose of this study is to understand the relations among executive functions, the general factor, and specific cognitive abilities consistent with the Cattell–Horn–Carroll (CHC) theory of cognitive abilities. To obtain measures of executive functions, we employed the Delis–Kaplan Executive Function System (DKEFS; Delis, Kaplan, & Kramer, 2001), which includes the only conormed executive function tests with a nationally representative normative sample of children and adolescents. Thus, it provided us with some of the most ideal measures of executive functions available. In particular, the DKEFS includes a range of test conditions that require increasing use of executive function, and we selected those conditions that seemed to best measure some central components of executive functions, such as inhibition, flexibility, and switching ability. To assess a broader range of cognitive abilities than were included in many previous factor-analytic studies, we employed the Woodcock–Johnson III Tests of Cognitive Abilities (WJ III; Woodcock, McGrew, & Mather, 2001) to operationalize several of the broad cognitive abilities specified in CHC theory. Thus, based on a sizable body of research focusing on the structure of Woodcock–Johnson tests batteries (e.g., Floyd, Keith, Taub, & McGrew, 2007; McGrew & Woodcock, 2001; Taub & McGrew, 2004) as well as several joint factor-analytic studies including them (e.g., Keith, Kranzler, & Flanagan, 2001; Phelps, McGrew, Knopik, & Ford, 2005), the WJ III tests provide a relatively wide range of marker or reference tests of well-established general and broad ability factors.

Because we believed that there was neither sufficient research nor substantive theory to guide the development of models specifying the relations between measures of executive functions and measures of CHC cognitive abilities, we first employed hierarchical exploratory factor analysis (EFA) to develop a plausible model describing these relations. Because EFA requires experimenter judgment regarding a number of issues—most notably, the number of factors to extract—and do not necessarily allow the "data to speak for themselves," we subsequently engaged in a model-comparison approach using confirmatory factor analysis (CFA). This approach allowed us to test and compare to our initial plausible model alternative models specifying distinctions between measures of cognitive abilities and measures of executive function, more simplistic models specifying only a general factor and only broad factors affecting measures, and integrative models developed to be consistent with prior research. Finally, we focused the study on children and adolescents because most studies examining related issues appear to have focused on adults and because the research informing CHC theory's model of cognitive abilities is based substantially on analysis of results from children and adolescents.

Method

Participants

Participants were 100 children and adolescents (49 boys and 51 girls) ranging in age from 8 to 18 years (mean [M] = 10.7, standard deviation [SD] = 2.5). Approximately 83% of the sample were White, 14% were Black, and 3% were Asian. We used father's education level (when available) and mother's education level (in five cases) as indicators of socioeconomic status: 1% of parents did not complete high school, 17% completed only high school, 35% attended some college, and 47% obtained a college degree or higher. Although all participants were recruited from general education classrooms, some parents reported that their children or adolescents (<10%) had been previously identified as experiencing some educational condition or psychological disorder.

Measures

DKEFS. The DKEFS (Delis et al., 2001) is composed of nine tests—Trail Making, Verbal Fluency, Design Fluency, Color–Word Interference, Sorting, 20 Questions, Word Context, Tower, and Proverb. Proverb was not administered as part of this study because it is inappropriate for the youngest participants in our targeted age range. Delis and colleagues (2001) reported reliability and validity evidence for the DKEFS test scores, and reviewers have noted its large and nationally representative normative sample and its sound development (e.g., Homack, Lee, & Riccio, 2005). Specific conditions of each remaining DKEFS test are described later in this article. Each condition produces an age-based scaled score ($M = 10$, $SD = 3$).

Trail Making Condition 4: Letter Number Switching requires examinees to quickly connect letters and numbers by alternating between the two. Its score represents the total number of correct connections. Across ages 8 to 19, Letter Number Switching yielded a 9- to 74-day test–retest reliability coefficient of .20 (Delis et al., 2001).

Verbal Fluency Condition 1: Letter Fluency requires examinees to rapidly generate words that begin with target letters. The Letter Fluency score represents the number of correct words produced across three trials. An alternate-form reliability coefficient of .83 was reported from a sample ages 16 to 89. Across ages 8 to 19, the Letter Fluency score yielded a 9- to 74-day test–retest reliability coefficient of .67. Condition 2: Category Fluency requires examinees to rapidly generate words that fall within semantic categories, and the resultant score represents the number of correct words produced across two trials. An alternate-form reliability coefficient of .71 and a test–retest reliability coefficient of .70 were reported. Condition 3: Category Switching requires examinees to rapidly switch between saying words from two different semantic categories, and the resultant score represents the number of correct shifts completed between two semantic categories. An alternate-form reliability coefficient of .44 and a test–retest reliability coefficient of .53 were reported (Delis et al., 2001).

Design Fluency Condition 3: Switching requires examinees to draw as many novel designs as quickly as possible by alternating connections between filled and empty dots presented in a standard constellation. The Total Correct score reflects the number of different designs completed. Across ages 8 to 19, Switching: Total Correct yielded a 9- to 74-day test–retest reliability coefficient of .13 (Delis et al., 2001).

Color–Word Interference Condition 3: Inhibition requires examinees to rapidly name the color in which words are printed while inhibiting the more automatic task of reading the words. The Condition 3: Inhibition score represents the seconds required to name the colors in which words are printed. Across ages 8 to 19, Inhibition yielded a 9- to 74-day test–retest reliability coefficient of .90 (Delis et al., 2001).

During the Free Sorting condition of Sorting, examinees must sort six cards into two different categories and describe the rule used to generate the categories. The Free Sorting Confirmed Correct score represents the accuracy of the descriptions of the categorization rules. During the Sort Recognition condition, examinees must describe the rule used by the examiner to generate categories for the cards. The Sort Recognition Description score also represents the accuracy and quality of examinees' descriptions of the categorization rules. Alternate-form reliability coefficients of .57 for Free Sorting Confirmed Correct and .72 for Sort Recognition Description were reported. Across ages 8 to 19, Free Sorting Confirmed Correct yielded a 9- to 74-day test–retest reliability coefficient of .49, and Sort Recognition Description yielded a test–retest reliability coefficient of .56 (Delis et al., 2001).

Word Context requires examinees to determine the meaning of made-up words using clues stemming from the use of the word in sentences. The Total Consecutively Correct score represents the efficiency of providing correct responses to clue sentences and the continued provision of that correct response to the remaining clue sentences. Across ages 8 to 19, the median Spearman–Brown corrected split-half reliability coefficient across the 10 items was .52, and the 9- to 74-day test–retest reliability coefficient was .58 (Delis et al., 2001).

Across four trials of Twenty Questions, examinees must ask the fewest yes or no questions as possible to identify a target object from a set of common objects presented in a visual array. The Total Achievement score represents the total number of questions asked to arrive at the target objects across trials, after adjustment for guessing. An alternate-form reliability coefficient of .37 was reported. Across ages 8 to 19, the median Spearman–Brown corrected split-half reliability coefficient across the four trials of the test was .81, and the 9- to 74-day test–retest reliability coefficient was .06 (Delis et al., 2001).

Tower requires examinees to move disks of varying size across pegs to build towers using the fewest number of moves. The Total Achievement score represents the number of disk moves required to reproduce the towers models. Across ages 8 to 19, the median Spearman–Brown corrected split-half reliability coefficient across the two halves of the test was .61, and the 9- to 74-day test–retest reliability coefficient was .51 (Delis et al., 2001).

WJ III Tests of Cognitive Abilities

A total of 18 tests from the WJ III were initially employed, but the tests Sound Blending and Auditory Attention were omitted from analysis. McGrew and Woodcock (2001) reported reliability and validity evidence for the WJ III test scores, and the WJ III tests have been evaluated favorably by reviewers (e.g., Cizek, 2003). Unless noted in the section that follows, tests demonstrated either median internal consistency or test–rest reliability coefficients of .80 or higher across ages 8 to 18. Each test produces an age-based standard score ($M = 100$, $SD = 15$).

Fluid Reasoning (Gf). Concept Formation requires examinees to identify the rule governing the organization of colored geometric figures when shown instances and non-instances of concepts. Analysis–Synthesis requires examinees to analyze the components of an incomplete logic puzzle and to determine missing components.

Visual Processing (Gv). Spatial Relations requires examinees to select the component parts of a whole shape. Picture Recognition requires examinees to view images for 5 seconds and identify those images within a larger array of images after the initial images have been removed. Planning requires examinees to trace as many lines as possible in increasingly complex patterns without lifting the pencil or retracing lines. Picture Recognition and Planning demonstrated median reliability coefficients of .72 and .75, respectively.

Processing Speed (Gs). Visual Matching requires examinees to rapidly mark pairs of identical numbers in rows. Decision Speed requires examinees to rapidly mark pairs of drawings in rows that are most closely related. Pair Cancellation requires examinees to rapidly mark repeated patterns of drawings.

Long-Term Storage and Retrieval (Glr). Retrieval Fluency requires examinees to rapidly generate words that fall within semantic categories across three trials. Rapid Picture Naming requires examinees to rapidly name pictures of common objects. Visual–Auditory Learning requires examinees to associate new visual symbols with orally presented words to translate combinations of the symbols into words and sentences.

Short-Term Memory (Gsm). Numbers Reversed requires examinees to listen to series of numbers and repeat them backward. Auditory Working Memory requires examinees to listen to a mixed series of words and numbers, say the words in order, and then say the numbers in order. Memory for Words requires examinees to repeat unrelated words in the correct sequence. Memory for Words demonstrated a median reliability coefficient of .77.

Comprehension–Knowledge (Gc). Verbal Comprehension requires examinees to name pictured objects, say words similar in meaning to words presented, say words that are opposites in meaning to words presented, and complete analogies. General Information requires examinees to respond to questions about the typical locations and functions of objects.

Procedures

Participants were recruited from two school districts in western and middle Tennessee during the 2001–2002 and 2002–2003 school years. Letters to parents of children and adolescents enrolled in Grades 3–12 were distributed by classroom teachers. Adult caregivers were asked to return the informed consent and demographic information form to teachers or mail it to the researchers via a postage-paid envelope. In addition, informed consent was obtained from all child participants. No data were collected to determine participation rate. The WJ III and the DKEFS were administered in counterbalanced order by trained graduate students in psychology. Within each test battery, tests were administered in standard order. Testing was typically conducted during two, 2-hour sessions. After testing completion, all participants were awarded a $20.00 gift card to a department store. Norm-based standardized scores were obtained using scoring software.

Analyses

Data-screening procedures were conducted prior to analyses, and assumptions regarding multivariate normality, absence of outliers, linearity, and homogeneity of variance–covariance matrices were met (Tabachnick & Fidell, 2007). Initial analysis revealed that 10 of the 29 variables contained cases with missing values. Because no variable contained more than 5% of cases with missing data, we used the expectation-maximization (EM) algorithm (Schafer, 1997) to estimate missing data. The differences between variable means from the original data and those obtained after applying the EM algorithm were miniscule and nonsignificant (see Table 1).

Table 1
M *and* SD *Values For the WJ III and DKEFS Measures Before and After Application of the EM Algorithm*

	Original Data			After EM		
Variable	*N*	*M*	*SD*	*N*	*M*	*SD*
DKEFS Measures						
Trail Making: Number-Letter Switching	99	10.2	3.5	100	10.2	3.5
Verbal Fluency: Letter Fluency	100	11.6	3.4	100	–	–
Verbal Fluency: Category Fluency	100	12.3	3.2	100	–	–
Verbal Fluency: Category Switching	100	11.7	2.6	100	–	–
Accuracy						
Design Fluency: Switching	97	11.5	3.4	100	11.5	3.4
Color-Word Interference: Inhibition	98	10.3	3.3	100	10.3	3.3
Sorting: Free Sorting Confirmed Correct	99	10.1	2.7	100	10.1	2.7
Sorting: Sort Recognition Description	100	10.2	2.9	100	–	–
Twenty Questions: Total Achievement	100	10.3	3.3	100	–	–
Word Context: Total Consecutively Correct	100	10.6	3.0	100	–	–
Tower	100	10.3	2.5	100	–	–
WJ III Tests						
Concept Formation	100	106.2	12.9	100	–	–
Analysis–Synthesis	99	109.4	11.9	100	109.4	11.8
Spatial Relations	100	103.7	9.4	100	–	–
Picture Recognition	100	106.8	10.3	100	–	–
Planning	96	108.4	10.3	100	108.4	10.2
Visual Matching	100	103.3	14.8	100	–	–
Decision Speed	100	108.3	15.8	100	–	–
Pair Cancellation	100	106.0	11.6	100	–	–
Retrieval Fluency	100	106.3	13.0	100	–	–
Rapid Picture Naming	98	98.0	13.7	100	98.0	13.6
Visual–Auditory Learning	100	100.2	13.4	100	–	–
Numbers Reversed	96	105.5	14.9	100	105.5	14.7
Auditory Working Memory	99	110.5	13.4	100	110.5	13.2
Memory for Words	100	106.6	13.6	100	–	–
Verbal Comprehension	100	107.9	12.9	100	–	–
General Information	98	103.4	14.4	100	103.5	14.4

The correlation matrix among the 27 variables was subjected to principal factor analysis, followed by an oblique rotation (i.e., promax). Because the correlations between factors suggested the existence of a higher-order factor (Gorsuch, 1983; Thompson, 2004), we completed the Schmid–Leiman transformation, which permits the examination of the relations among both first- and second-order factors by allowing a higher-order factor (i.e., the general factor) to account for as much of the variance among observed variables as possible and reducing lower-order factors to residual factors that are uncorrelated with both one another and with the higher-order factor.

Results from the EFA and the Schmid–Leiman transformation were used to construct a "baseline" model—called the final EFA model—using Amos 5.0 (Arbuckle & Wothke, 2004). Parameters from factors to their indicators were included in the CFA model if the factor coefficients from the first-order factors to the test scores were $\geq$.25 in the EFA. One exception was made to this rule when one test score did not demonstrate a factor coefficient of this magnitude, and two parameters from first-order factors to test scores were included in three cases. A single, higher-order general factor was modeled to affect each first-order factor. The final model from the EFA was compared to a variety of competing theoretical models (see Burns et al., 2009).

Maximum-likelihood estimation was employed to estimate free parameters. Correlations and *SD* values were the input for Amos; covariance matrices were analyzed. Nested models were compared using chi-square difference ($\Delta\chi^2$) tests, and non-nested models were compared using the Akaike information criterion (AIC; Loehlin, 2004). Nested models that displayed significantly lower chi-square values, relative to degrees of freedom, and non-nested models with lower AIC values than the final EFA model were deemed more accurate models.

Results

M and *SD* values for each of the 29 variables are reported in Table 1. The correlation matrix for all variables can be obtained by contacting the first author.

EFA

Bartlett's test of sphericity was statistically significant ($p < .05$), and the Kaiser–Meyer–Olkin measure of sampling adequacy exceeded .60 for initial and subsequent analyses. The WJ III Visual–Auditory Learning test and the DKEFS Tower subtest yielded low communalities (.32 and .23, respectively) during the initial analysis, and they were omitted from subsequent analyses. Analysis of the remaining 25 variables indicated the presence of three to six factors. Horn's parallel analysis (Horn, 1965) suggested extraction of three factors, the scree plot suggested extraction of four factors, and the eigenvalues-greater-than-or-equal-to-1 rule suggested extraction of six factors. Thus, we extracted three to seven factors and examined the solutions. Because the six-factor solution was most consistent with contemporary theories (e.g., Carroll, 1993) and because its results made the most psychological sense (Gorsuch, 1983; Thompson, 2004), we extracted six first-order factors and then a second-order factor from the correlations between first-order factors. Table 2 presents the Schmid–Leiman solution. Most of the 25 variables load moderately to highly on the second-order factor despite considerable differences in test content and specific abilities measured. It is notable that the mean loading of the DKEFS measures on the second-order factor ($M = .51$, $SD = .07$) is somewhat higher than the mean loading of the WJ III measures ($M = .49$, $SD = .12$). Thus, it appears that the second-order factor represents a general ability that is common to both cognitive ability and purported executive function measures.

As evident in Table 2, first-order factors one through five comprise tests from both test batteries, and factor six comprises only WJ III tests. Loadings of .25 and higher were considered adequate loadings and are bolded (Thompson, 2004). It is notable that the WJ III Analysis–Synthesis test does

Table 2
Schmid–Leiman Solution for the WJ III and DKEFS Measures

Measure	Second-Order Factor	First-Order Factor I	II	III	IV	V	VI
DKEFS Sorting: Free Sorting Confirmed Correct	.50	**.45**	.07	.10	−.10	−.11	.09
DKEFS Word Context: Total Consecutively Correct	.67	**.37**	.01	−.06	.16	.16	−.27
DKEFS Sorting: Sort Recognition Description	.56	**.36**	.13	.00	−.14	.07	.14
WJ III Verbal Comprehension	.69	**.31**	−.07	.04	.01	.22	.05
DKEFS Twenty Questions: Total Achievement	.47	**.30**	.17	−.18	.13	−.06	.04
WJ III General Information	.63	**.29**	−.02	.14	.05	.16	−.07
WJ III Analysis–Synthesis	.58	.22	.07	.07	.05	.08	.13
WJ III Visual Matching	.42	.00	**.77**	−.02	−.03	−.03	.03
WJ III Pair Cancellation	.34	.10	**.58**	−.10	−.04	−.01	−.09
WJ III Decision Speed	.49	.08	**.50**	.08	−.09	.09	.05
DKEFS Color-Word Interference: Inhibition	.52	.07	**.32**	.14	.19	.00	−.10
DKEFS Verbal Fluency: Category Fluency	.47	.05	−.11	**.75**	−.01	.03	.03
WJ III Retrieval Fluency	.44	−.09	.03	**.72**	.01	.09	.01
DKEFS Verbal Fluency: Letter Fluency	.49	.20	.07	**.53**	.02	−.13	.03
WJ III Numbers Reversed	.52	−.08	−.02	.00	**.63**	−.04	.07
WJ III Auditory Working Memory	.57	.12	−.01	−.02	**.41**	.08	.20
DKEFS Trail Making: Number-Letter Switching	.52	−.11	**.33**	.04	**.38**	−.01	.08
WJ III Memory for Words	.49	.21	−.16	.06	**.34**	−.01	−.12
DKEFS Verbal Fluency: Category Switching Accuracy	.42	.05	−.04	.06	−.11	**.38**	−.01
WJ III Rapid Picture Naming	.47	−.13	**.28**	.10	.01	**.36**	−.06
WJ III Concept Formation	.65	.14	−.08	−.09	.08	**.33**	.17
DKEFS Design Fluency: Switching	.44	.02	**.25**	−.08	.01	**.26**	−.04
WJ III Planning	.31	.06	−.03	−.04	.08	−.07	**.53**
WJ III Picture Recognition	.29	−.03	.00	.13	.00	.02	**.46**
WJ III Spatial Relations	.53	.05	−.07	−.05	.06	.22	**.41**

not load highly on any first-order factor based on the >.25 criterion, and three tests (i.e., DKEFS Trail Making: Number-Letter Switching, WJ III Rapid Picture Naming, and DKEFS Design Fluency: Switching) yielded adequate loadings on two first-order factors. Factor one appears to measure the CHC broad ability Comprehension–Knowledge (*Gc*). Factor two appears to measure the CHC broad ability Processing Speed. Factor three appears to assess the CHC broad ability Long-Term Storage and Retrieval and perhaps more specifically, the CHC narrow ability Naming Facility. Factor four appears to measure the CHC broad ability Short-Term Memory. Factor five appears to measure a factor related to mental flexibility that does not correspond well with existing CHC broad or narrow ability factors. It was tentatively considered to be a broad Executive Functions factor. Factor six, formed from only WJ III tests, appears to measure the CHC broad ability Visual Processing.

CFA

Figure 1 presents the final EFA model. Its measurement model includes (a) the scores from the WJ III and DKEFS tests and conditions, which are measured variables represented by rectangles on the left side of the figure; (b) the error and unique variances of the measured variables, which are represented by circles near the rectangles; (c) the first-order broad ability factors, which are represented by ellipses to the right of the rectangles; and (d) the unique variances of the broad ability

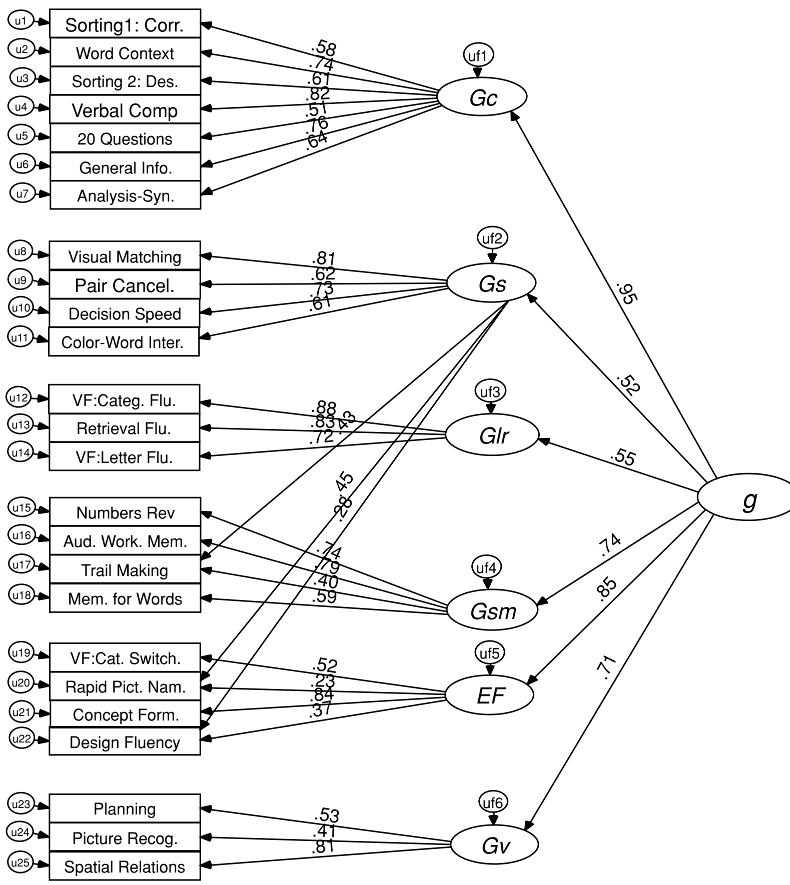

FIGURE 1. Final EFA Model used in model comparison testing via CFA. *Gc:* Comprehension–Knowledge; *Gs:* Processing Speed; *Glr:* Long-Term Storage and Retrieval; *Gsm:* Short-Term Memory; EF: Executive Function; *Gv:* Visual Processing; *g:* General Factor.

factors, which are represented by circles above the broad ability factors. The second-order general factor is represented by a single ellipse on the right side of the figure. Fit statistics for the final EFA model (Model 1) are presented in Table 3. As shown, this model provided a good fit to the data.

Five sets of alternative models were compared to the final EFA model, which is integrative across both test batteries. The first set of alternative models specified (a) the WJ III tests as measuring CHC broad and general abilities—consistent with the test structure and prior publications using the WJ III as well as the final EFA model—and (b) the DKEFS tests as measuring executive functions. In the first model (Model 2 in Table 3), all WJ III tests were specified to be affected (directly) by first-order factors (as listed in the "Measures" subsection) as well as (indirectly) by a second-order general factor. (Only one test, WJ III Rapid Picture Naming, loaded on two first-order factors, Long-Term Storage and Retrieval and Processing Speed.) In addition, all DKEFS tests were specified to be affected by a first-order general Executive Function factor, and the relation between these two general factors was specified. As evident in Table 3, this model produced notable degradation of model fit. Its AIC was notably higher than that in the final EFA model, and other fit statistics were notably inferior to those in the final EFA model. In a second model of this type (Model 3), the portion of the model derived from WJ III tests remained the same as Model 2, but two first-order

Table 3
Fit Indexes and Model Difference Statistics for the Final EFA Model and Alternative Comprehensive Models

Model	χ^2	*df*	$\Delta\chi^{2a}$	$\Delta\, df^a$	*p*	$\Delta\chi^{2b}$	Δdf^b	*p*	CFI	TLI	RMSEA (90% interval)	AIC
1. Final EFA Model	350.70	266							.91	.90	.057 (.039–.072)	468.70
2. CHC Two-Strata Model + EF	451.06	267							.81	.78	.083 (.070–.097)	567.06
3. CHC Two-Strata Model + Verbal EF and Perceptual EF	439.41	265				11.65	2	<.001	.82	.79	.082 (.068–.095)	559.41
4. Integrated CHC General Factor Model	616.76	276	266.06	10	<.001				.64	.61	.112 (.100–.123)	714.76
5. Integrated Verbal and Perceptual	561.21	274				55.55	2	<.001	.70	.67	.103 (.091–.115)	663.21
6. Integrated Wechsler 4-Factor Model	445.06	269							.82	.80	.081 (.068–.095)	557.06
7. Integrated Wechsler 4-Factor Model + *Gf*	435.19	268				9.87	1	.002	.83	.80	.079 (.066–.093)	549.19
8. Integrated Wechsler 4-Factor Model + *Glr*	372.84	268				72.22	1	<.001	.89	.88	.063 (.047–.078)	486.84
9. Integrated CHC Two-Strata Model without First-Order EF	363.08	267							.90	.89	.060 (.044–.075)	479.08

Note. EF: Executive Function; *Gf:* Fluid Reasoning; *Glr:* Long-Term Storage and Retrieval; CFI: comparative fit index; TLI: Tucker Lewis Index; RMSEA: Root Mean Square Error of Approximation; AIC: Akaike Information Criterion.
[a]Compared to final EFA Model.
[b]Compared to more parsimonious nested model.

factors, a Verbal factor and a Perceptual factor (Vernon, 1964; Wechsler, 1974, 1981), were included between a second-order general Executive Function factor and the DKEFS scores. Tests including item content that required verbal output were specified as being affected by the Verbal factor, and tests requiring no verbal output and including item content that was primarily visual in nature were specified as being affected by the Perceptual factor. Model 3 was significantly better fitting than Model 2, $\Delta\chi^2 = 11.65$, $\Delta df = 2$, $p < .001$, but it was notably inferior to the final EFA model. It is also notable that Model 2 and Model 3 yielded correlations between the CHC general factor and the general Executive Function factor that were .99 and 1.0, respectively. These correlations suggest that the two second-order factors are indistinguishable.

A second set of alternative models, which were more basic than the previous set of alternative models but integrated across tests batteries, was compared to the final EFA model. When all WJ III and DKEFS tests were modeled to load on only a single general factor (Model 4), model fit was dismal and significantly worse than that of the final EFA model, $\Delta\chi^2 = 266.06$, $df = 10$, $p < .001$. Next, Model 5 was developed to include correlated Verbal and Perceptual factors affecting all test scores, as described previously. Model fit for Model 5 was significantly better fitting than that for Model 4, $\Delta\chi^2 = 55.55$, $\Delta df = 2$, $p < .001$, but this model was worse fitting than the final EFA model. Thus, both of these integrative models produced degradations in model fit when contrasted with the final EFA model. These models are clearly too simplistic to represent the general and specific abilities measured by the WJ III and DKEFS.

A third set of alternative models reflected the factor structure of the Wechsler intelligence scales that has influenced many factor-analytic studies of executive functions (e.g., Leonberger et al., 1991, 1992). All of these models included a second-order general factor. The first model in this set (Model 6) specified four first-order factors representing Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed. The Working Memory and Processing Speed factors are identical to the Short-Term Memory and Processing Speed factors from the final EFA model. Perceptual Reasoning included both the measures of Visual Processing identified in the final EFA model as well as the WJ III Concept Formation and Analysis–Synthesis tests. The Verbal Comprehension factor included all of the tests measuring Comprehension–Knowledge as well as all of the tests identified in the final EFA model as measuring Long-Term Storage and Retrieval. Model 6 was notably worse fitting than the final EFA model. When the Perceptual Reasoning factor specified in the initial model of this set was divided into a Visual Processing factor (as in the final EFA model) and a Fluid Reasoning factor affecting the WJ III Concept Formation and Analysis–Synthesis tests (Model 7), fit was significantly improved over that in Model 6, $\Delta\chi^2 = 9.87$, $\Delta df = 1$, $p = .002$, but the model remained inferior to the final EFA model. When the Verbal Comprehension factor specified in Model 6 was divided into a Comprehension–Knowledge factor and a Long-Term Storage and Retrieval factor (Model 8), fit was significantly improved over that in Model 6, $\Delta\chi^2 = 72.22$, $\Delta df = 1$, $p < .001$, but it remained inferior to the final EFA model. The Wechsler scale four-factor model and its extensions failed to capture the relations among measures as well as the six-factor model resulting from the EFA.

The fourth set of alternative models tested hypotheses consistent with previous analyses of the WJ III test scores. In Model 9, the broad Executive Function factor from the final EFA model was deleted, and tests loading on it were assigned to other factors consistent with CHC theory. Verbal Fluency: Category Switching Accuracy was assigned to the Long-Term Storage and Retrieval factor, Concept Formation and Analysis–Synthesis were assigned to a Fluid Reasoning factor, and Trail Making and Design Fluency were assigned to only the Processing Speed factor (on which they also had a secondary loading in the final EFA model). As evident in Table 3, Model 9 was worse fitting than the final EFA model (ΔAIC = 10.38). Something appears to be gained when the broad Executive Function factor is specified.

In addition, alternative models were tested that targeted only individual tests (Keith et al., 2001). Across these models, none provide significantly better fit than the final EFA model (see Table 4). First, neither a model in which Analysis–Synthesis loaded on both the Comprehension–Knowledge and the broad Executive Function factors (Model 10a) nor a model in which Analysis–Synthesis loaded on only the broad Executive Function factor (Model 10b) produced significantly better fit than the final EFA model (see Table 4). Second, models based on research indicating that measures from tasks like the DKEFS Trail Making: Number-Letter Switching condition demonstrated sizable relations with measures of Visual Processing—alone and in addition to measures of Processing Speed and Short-Term Memory (Larrabee, 2000; Larrabee & Curtiss, 1995; Sherman et al., 1995)—failed to produce significantly better fit than the final EFA model (see Table 4). These models included (a) a model in which Trail Making loaded on the Processing Speed, Short-Term Memory, and Visual Processing factors (Model 11a); (b) a model in which Trail Making loaded on the Processing Speed and Visual Processing factors (Model 11b); (c) a model in which Trail Making loaded on the Processing Speed and Visual Processing factors (Model 11c); and (d) a model in which Trail Making loaded on only the Visual Processing factor (Model 11d). Only Model 11a produced a slightly better fitting model than the more parsimonious final EFA model, $\Delta\chi^2 = .24$, $\Delta df = 1$, $p = .624$, but it was rejected due to lack of significantly better fit with additional model paths.

Third, based on the logic that DKEFS Design Fluency: Switching measures Visual Processing—alone or in addition to the Processing Speed and broad Executive Function factors—the final EFA model in which Design Fluency loaded on both the Processing Speed and broad Executive Function factors was compared to three alternative models that failed to produce significantly better fit than the final EFA model (see Table 4). These models included (a) a model in which Design Fluency loaded on Processing Speed, Executive Function, and Visual Processing factors (Model 12a); (b) a model in which Design Fluency loaded on Processing Speed and Visual Processing (Model 12b); and (c) a model in which Design Fluency loaded on Executive Function and Visual Processing (Model 12c). Only Model 12a produced a slightly better fitting model than the more parsimonious final EFA model, $\Delta\chi^2 = 0.40$, $\Delta df = 1$, $p = .527$, but it was rejected due to lack of significantly improved model fit with additional model paths.

Fourth, based on research indicating that tests like the DKEFS Color-Word Interference: Inhibition test measure Visual Processing (Sherman et al., 1995), the final EFA model in which Color-Word Interference loaded on Processing Speed alone was compared to two alternative models that failed to produce significantly better fit than the final EFA model (see Table 4). These models included (a) a model in which Color-Word Interference: Inhibition loaded on Processing Speed and Visual Processing (Model 13a) and (b) a model in which it loaded on only Visual Processing (Model 13b). Only Model 13a produced a slightly better fitting model than the more parsimonious final EFA model, $\Delta\chi^2 = 3.64$, $\Delta df = 1$, $p = .056$, but it was rejected due to lack of significantly improved model fit with additional model paths. Fifth, because the DKEFS conditions Sorting: Free Sorting Confirmed Correct, Sorting: Sort Recognition Description, and Twenty Questions: Total Achievement have been hypothesized to involve deductive and inductive reasoning as central processes, and because Salthouse (2005) indicated that one variant of DKEFS Sort Recognition measures a construct similar to Fluid Reasoning, two alternative models were tested for each DKEFS measure. Each alternative model failed to produce significantly better fit than the final EFA model. These alternative models included (a) those in which each DKEFS condition measures both the Comprehension–Knowledge and the broad Executive Function factors (Models 14a, 15a, and 16a) and (b) those in which each measures only Executive Function (Models 14b, 15b, and 16b). Each model in which the DKEFS loaded on both Comprehension–Knowledge and Executive Functions was slightly better fitting than the more parsimonious final EFA model ($p > .05$, see Table 4), but they were rejected due to lack of significantly improved model fit with additional model paths.

Table 4
Fit Indexes and Model Difference Statistics for the Final EFA Model and Test-Focused Alternative Models

Model and Paths	χ^2	*df*	$\Delta\chi^{2a}$	Δdf^a	*p*	$\Delta\chi^{2b}$	Δdf^b	*p*	TLI	CFI	RMSEA (90% interval)	AIC
1. Final EFA Model	350.70	266							.91	.90	.057 (.039–.072)	468.70
10a. EF & *Gc* to Analysis–Synthesis	349.85	265	0.85	1	.357				.91	.90	.057 (.039–.072)	469.85
10b. EF to Analysis–Synthesis	352.85	266				3.00	1	.083	.91	.90	.057 (.040–.073)	470.85
11a. *Gv, Gsm,* & *Gs* to Trail Making: LNS	350.46	265	0.24	1	.624				.91	.90	.057 (.039–073)	470.46
11b. *Gv* & *Gsm* to Trail Making: LNS	368.14	266				17.68	1	<.001	.89	.88	.062 (.046–.077)	486.14
11c. *Gv* & *Gs* to Trail Making: LNS	361.11	266				10.65	1	.001	.90	.89	.060 (.043–.075)	479.11
11d. *Gv* to Trail Making: LNS	382.42	267				21.31	1	<.001	.88	.86	.066 (.051–.081)	498.42
12a. *Gs*, EF, & *Gv* to Design Fluency: Switching	350.30	265	0.40	1	.527				.91	.90	.057 (.039–.073)	470.30
12b. *Gs* & *Gv* to Design Fluency: Switching	357.67	266				7.37	1	.007	.90	.89	.059 (.042–.074)	475.67
12c. EF & *Gv* to Design Fluency: Switching	356.51	266				6.21	1	.013	.91	.89	.059 (.041–.074)	474.51
13a. *Gv* & *Gs* to CWI: Inhibition	347.06	265	3.64	1	.056				.91	.90	.056 (.038–.072)	467.06
13b. *Gv* to CWI: Inhibition	367.85	266				20.79	1	<.001	.89	.88	.062 (.046–.077)	486.85
14a. EF & *Gc* to Sorting: FSCC	348.21	265	2.49	1	.115				.91	.90	.056 (.039–.072)	468.21
14b. EF to Sorting: FSCC	359.33	266				11.12	1	.001	.90	.89	.060 (.043–.075)	477.33
15a. EF & *Gc* to Sorting: SRD	348.66	265	2.04	1	.153				.91	.90	.056 (.039–.072)	468.66
15b. EF to Sorting: SRD	350.49	266				1.83	1	.176	.91	.90	.057 (.039–.072)	468.49
16a. EF & *Gc* to Twenty Questions: Total Achievement	350.17	265	0.53	1	.467				.91	.90	.057 (.039–.072)	470.17
16b. EF to Twenty Questions: Total Achievement	352.67	266				2.50	1	.114	.91	.90	.057 (.040–.073)	470.67
17a. *Glr* & EF to Verbal Fluency: CSA	347.76	265	2.94	1	.086				.91	.90	.056 (.038–.072)	467.76
17b. *Glr* to Verbal Fluency: CSA	360.21	266				12.45	1	>.001	.90	.89	.060 (.043–.075)	478.21

Note. EF: Executive Function; *Gc:* Comprehension–Knowledge; *Gv:* Visual–Spatial Thinking; *Gsm:* Short–Term Memory; *Gs:* Processing Speed; *Glr:* Long–Term Storage and Retrieval; LNS: Letter Number Switching; FSCC: Free Sorting Confirmed Correct; SRD: Sort Recognition Description; CSA: Correct Switching Accuracy; CWI: Color–Word Interference; CFI: comparative fit index; TLI: Tucker Lewis Index; RMSEA: Root Mean Square Error of Approximation; AIC: Akaike Information Criterion.
[a]Compared to final EFA Model.
[b]Compared to most parsimonious nested model.

Finally, because the final EFA model indicates that both Verbal Fluency: Letter Fluency and Verbal Fluency: Category Fluency are measures of Long-Term Storage and Retrieval, whereas Verbal Fluency: Category Switching Accuracy is a measure of the broad Executive Function factor—despite relatively similar response requirements across the three conditions—two alternative models were tested. They included (a) a model in which Verbal Fluency: Category Switching Accuracy loaded on both Executive Function and Long-Term Storage and Retrieval (Model 17a) and (b) a model in which it loaded on only Long-Term Storage and Retrieval (Model 17b). As evident in Table 4, only Model 17a produced a slightly better fitting model than the more parsimonious final EFA model, $\Delta\chi^2 = 2.94$, $\Delta df = 1$, $p = .086$, but it was rejected due to lack of significantly better fit with additional model paths.

Discussion

The results of this study suggest that measures of executive functions and measures of CHC cognitive abilities are not easily distinguished. Instead, results reveal that every DKEFS test or condition measures the general factor as well as broad ability factors outlined in CHC theory. As indicated by Eslinger (1996) and others, if there are measures of abilities associated with executive functions, they are contaminated by the general factor and more specific ability factors, so that there is probably little unique about them.

The General Factor

All correlations between WJ III and DKEFS tests were positive, revealing the positive manifold associated with the general factor. Furthermore, EFA revealed that all tests across batteries demonstrated general factor loadings of at least .29 and that all DKEFS tests demonstrated general factor loadings of at least .42. Although the majority of measures that loaded most strongly on the second-order general factor were from the WJ III, the average loading of the DKEFS measures (at least those we selected) was somewhat higher than that from the WJ III tests. Furthermore, all of the first-order broad ability factors we specified were highly affected by the general factor. Based on the best-fitting CFA model tested, the general factor to first-order factor loadings were highest for the Comprehension–Knowledge factor (.95)—followed by broad Executive Function (.85), Short-Term Memory (.74), Visual Processing (.71), Long-Term Retrieval (.55), and Processing Speed (.52) factors.

Broad Abilities and DKEFS Tests and Conditions

*S*cores from every DKEFS test or condition loaded on at least one of the CHC broad ability factors we specified in the CFA modeling, and we offer these tentative conclusions regarding the abilities that they measure. Of all the tests and conditions included in the DKEFS and employed in this study, tasks like Trail Making: Letter Number Switching appear to have received the most study via factor-analytic research. Although some studies indicated that this task measures abilities associated with sound discrimination, tactile awareness, and visuospatial abilities (Chittooran et al. 1993, D'Amato, Gray, & Dean 1988; Lamar, Zonderman, & Resnick, 2002; Larrabee & Curtiss, 1995; Leonberger et al., 1991; Sherman et al., 1995), the results of this study are consistent with some prior research indicating that it measures Processing Speed (Larabee, 2000; Salthouse, 2005). It also appears, however, to measure Short-Term Memory, which is facilitated by interference control and verbal rehearsal.

It is also no surprise that both DKEFS Category Fluency and DKEFS Letter Fluency loaded on the Long-Term Storage and Retrieval factor. Scores from tests like these have been shown to load on factors associated with verbal or knowledge-related abilities and to correlate with related measures

(Larabee, 2000; Salthouse, 2005; Salthouse & Davis, 2006), but most prior research has not included marker tests of Long-Term Storage and Retrieval. It was surprising, however, that a variant of DKEFS Category Fluency and Letter Fluency—DKEFS Category Switching Accuracy—loaded instead on the broad Executive Function factor, which suggests that this factor measures some processes other than memory retrieval. Some research indicated that tests like Color-Word Interference: Inhibition measure some of the same abilities as tests measuring Perceptual Organization (Sherman et al., 1995), but they have more commonly been revealed to be measures of Processing Speed (Burns et al., 2009; Salthouse, 2005; Salthouse & Davis, 2006). The current research supports this consensus. Similarly, Design Fluency appears to be a measure of Processing Speed, not Visual Processing (cf. Salthouse, 2005).

It may be a surprise to some that all four of the remaining DKEFS tests and conditions—Sorting: Free Sorting Confirmed Correct, Sorting: Sort Recognition Description, Twenty Questions: Total Achievement, and Word Context: Total Consecutively Correct—appear to be measures of the CHC broad ability Comprehension–Knowledge. It may have been hypothesized that they (perhaps barring Word Context) would measure the CHC broad ability Fluid Intelligence because they seem to require the processes associated with the narrow abilities subsumed by Fluid Intelligence, General Sequential Reasoning and Induction, as do the WJ III tests Analysis–Synthesis and Concept Formation, respectively. (It is notable that Analysis–Synthesis also loaded on the Comprehension–Knowledge factor and that Concept Formation did not. These results with the WJ III tests are anomalies and cannot be easily explained.) It may be that these four DKEFS tests and conditions required reasoning that was dependent on prior verbal knowledge—in contrast to "novel reasoning." It is apparent that Word Context requires such verbal knowledge, and this assertion is supported by its factor loading on the Comprehension–Knowledge factor in the best-fitting model (.74) that was almost as high as the WJ III test General Information (.76), which is clearly a measure of breadth of knowledge.

How Do Executive Functions Fit with the CHC Model?

We envision three ways in which executive functions fit with the CHC model—as components of social competence, as measures of cognitive abilities, and as measures of cognitive processes. For one, it may be that executive functions are components of an extremely general construct that is similar to *personal competence* offered by Greenspan and Driscoll (1997). Personal competence subsumes specific competencies represented well by executive functions, such as *affective competence*, *everyday competence*, and *academic competence*, whereas the abilities described by CHC theory represent only narrow subcomponents of academic competence (i.e., *conceptual intelligence* and *language*). It may be that measures from both individually administered tests of executive functions and tests of cognitive abilities reside at this subcomponent level of personal competence—whereas other methods of measuring executive functions, such those from behavior rating scales (e.g., Gioia et al., 2000), will better represent other subcomponents, such as affective competence.

In contrast, the results of this study appear to indicate that measures from both individually administered tests of executive functions and tests of cognitive abilities measure the same types of abilities (e.g., the general factor or conceptual intelligence). Certainly, results from two of our models (Models 2 and 3) indicate that these two sets of measures are affected by the same latent ability, so perhaps the true executive function factor is the CHC general factor, as postulated by several scholars (e.g., Crinella & Yu, 2000; Detterman, 1982; Sternberg & Gardner, 1983). It is also possible that executive functions are represented by more specific abilities in CHC theory. Our broad Executive Function factor was shown to influence performance on WJ III Concept Formation and, in part, WJ III Rapid Picture Naming and DKEFS Verbal Fluency: Category Switching Accuracy and

DKEFS Design Fluency. Perhaps all of the tests and conditions loading on this factor require mental flexibility—the ability to "switch set"—in a manner that the other tests included in the batteries do not. It is possible that "switching ability" should be represented as a CHC ability, but we believe that it would be premature to do so. Nonetheless, we are confident that our results and prior research suggest confluence in the measurement of executive functions and cognitive abilities.

Finally, school psychologists and other assessment professionals should refer to executive functions in the same manner that they refer to cognitive processes (a) targeted by specific items or task designs and (b) inferred from observing individuals completing such tasks (Crinella & Yu, 2000). According to Jensen (1998), cognitive processes are "hypothetical constructs used by cognitive theorists to describe how persons apprehend, discriminate, select, and attend to certain aspects of the vast welter of stimuli that impinge on the sensorium to form internal representations that can be mentally manipulated, transformed, related to previous internal representations, stored in memory . . . and later retrieved from storage to govern the person's decision and behavior in a particular situation" (pp. 205–206). Because all voluntary behaviors are the result of some sequence of cognitive processes, consequentially, all measures of cognitive abilities can be said to be the result of cognitive processes—with some abilities reflecting a subset of processes called executive functions. In light of our results, those from previous research, and prominent cognitive models (see Floyd, 2005), we encourage restriction of the use of the term *executive functions* to either (a) cognitive processes inferred based on the performance of individuals (Jensen, 1998) or (b) components of social competence and reference to norm-based scores from both tests of executive functions and tests of cognitive abilities as measures of individual differences in cognitive abilities.

Limitations and Suggestions for Future Research

Although the results of this study offer insight into the relations among executive functions and cognitive abilities, this study has a number of weaknesses that limit the applicability its findings. First, participants were drawn from two school districts in Tennessee, and the ratio of sample size to model size was small (see Herzog & Boomsma, 2009; Nevitt & Hancock, 2004). As the result of our sampling, the *M* values for the vast majority of the measures were at least slightly greater than the population mean, and many measures demonstrated restriction or expansion of range. The ratio of sample size to model size may also have limited our ability to detect observed differences between CFA models. Following procedures outlined by MacCallum, Browne, and Cai (2006), it appears, however, in every case that we had adequate power ($>.80$) to detect such observed differences between models. Nonetheless, well-informed caution should be used in interpreting our results; it is possible that some results are due to sampling error and that they would not be replicated using other samples of children, homogeneous clinical samples (Delis, Jacobson, Bondi, Hamilton, & Salmon, 2003), and samples of adults. Future research should include larger and more thoughtfully selected samples.

Second, although the DKEFS includes the only conormed executive function tests that yield norm-based scores derived from a nationally representative norm sample and independent reviews of the battery have been complimentary, many DKEFS measures have demonstrated low levels of reliability. Future research should draw on measures of executive functions with better evidence of reliability. Finally, one DKEFS test, Tower—for which we had data—was excluded from the analysis based on its weak relations with other measures included in the factor analysis. The conclusions drawn about the other DKEFS based on this research do not necessarily apply to the Tower test or other laboratory- or computer-based measures of executive functions.

Conclusion

The results from this study are similar to those of previous factor-analytic studies using measures from both intelligence and neuropsychological test batteries. As such, these results suggest that there are similarities in terms of constructs measured across intelligence and neuropsychological test batteries. We suggest that measures of executive functions obtained from individually administered tests are so remarkably similar to measures of CHC cognitive abilities and consistent with CHC theory that distinctions between them should be minimized and that the term *executive functions* should be used to describe either some components of social competence or cognitive processes inferred based on the performance of an individual during test taking.

References

Arbuckle, J. L., & Wothke, W. (2004). Amos 5.0 user's guide. Chicago: Smallwaters.

Burns, N., Nettelbeck, T., & McPherson, J. (2009). Attention and intelligence: A factor analytic study. Journal of Individual Differences, 30, 44–57.

Carroll, J. B. (1993). Human cognitive abilities: A survey of factor analytic studies. New York: Cambridge University.

Chittooran, M. M., D'Amato, R. C., Lassiter, K. S., & Dean, R. S. (1993). Factor structure of psychoeducational and neuropsychological measures of learning-disabled children. Psychology in the Schools, 30, 109–118.

Cizek, G. J. (2003). Review of the Woodcock-Johnson III. In B. S. Plake & J. C. Impara (Eds.), The fifteenth mental measurements yearbook (pp. 1020–1024). Lincoln, NE: Buros Institute of Mental Measurements.

Crinella, F. M., & Yu, J. (2000). Brain mechanisms and intelligence: Psychometric g and executive function. Intelligence, 27, 299–327.

D'Amato, R. C., Gray, J. W., & Dean, R. S. (1988). A comparison between intelligence and neuropsychological functioning. Journal of School Psychology, 26, 282–292.

Delis, D. C., Jacobson, M., Bondi, M. W., Hamilton, J. M., & Salmon, D. P. (2003). The myth of testing construct validity using factor analysis or correlations with normal or mixed clinical populations: Lessons from memory research. Journal of International Neuropsychological Society, 9, 936–946.

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

Detterman, D. K. (1982). Does "g" exist? Intelligence, 6, 99–108.

Eslinger, P. J. (1996). Conceptualizing, describing, and measuring components of executive function. In G. R. Lyon & N. A. Krasnegor (Eds.), Attention, memory, and executive function (pp. 367–395). Baltimore, MD: Brooks.

Floyd, R. G. (2005). Information-processing approaches to interpretation of contemporary intellectual assessment instruments. In D. P. Flanagan & P. Harrison (Eds.), Contemporary intellectual assessment (2nd ed., pp. 203–233). New York: Guilford Press.

Floyd, R. G., Keith, T. Z., Taub, G. E., & McGrew, K. S. (2007). Cattell–Horn–Carroll cognitive abilities and their effects on reading decoding skills: g has indirect effects, more specific abilities have direct effects. School Psychology Quarterly, 22, 200–233.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). Behavior Rating Inventory of Executive Function. Lutz, FL: Psychological Assessment Resources.

Gorsuch, R. L. (1983). Factor analysis (2nd ed.). Hillsdale, NJ: Erlbaum.

Greenspan, S., & Driscoll, J. (1997). The role of intelligence in a broad model of personal competence. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), Contemporary intellectual assessment (pp. 131–150). New York: Guilford Press.

Herzog, W., & Boomsma, A. (2009). Small-sample robust estimators of noncentraility-based and incremental model fit. Structural Equation Modeling, 16, 1–27.

Homack, S., Lee, D., & Riccio, C. A. (2005). Test review: Delis–Kaplan Executive Function System. Journal of Clinical and Experimental Neuropsychology, 27, 599–609.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. Psychometrika, 30, 179–185.

Jensen, A. R. (1998). The g factor: The science of mental ability. Westport, CT: Preager.

Keith, T. Z., Kranzler, J. H., & Flanagan, D. P. (2001). What does the Cognitive Assessment System (CAS) measure? Joint confirmatory factor analysis of the CAS and the Woodcock–Johnson Tests of Cognitive Ability (3rd ed.). School Psychology Review, 30, 89–119.

Kelly, M. D., Arceneaux, J. M., Dean, R. S., & Anderson, J. L. (1994). Neuropsychological significance of IQ summary scores. International Journal of Neuroscience, 75, 175–179.

Lamar, M., Zonderman, A. B., & Resnick, S. (2002). Contributions of specific cognitive processes to executive functioning in an aging population. Neuropsychology, 16, 156–162.

Larrabee, G. J. (2000). Association between IQ and neuropsychological test performance: Commentary on Tremont, Hoffman, Scott, and Adams (1998). The Clinical Neuropsychologist, 14, 139–145.

Larrabee, G. J., & Curtiss, G. (1995). Construct validity of various verbal and visual memory tests. Journal of Clinical and Experimental Neuropsychology, 17, 536–547.

Leonberger, F. T., Nicks, S. D., Goldfader, P. R., & Munz, D. C. (1991). Factor analysis of the Wechsler Memory Scale–Revised and the Halstead–Reitan Neuropsychological Battery. The Clinical Neuropsychologist, 5, 83–88.

Leonberger, F. T., Nicks, S. D., Larrabee, G. J., & Goldfader, P. R. (1992). Factor structure of the Wechsler Memory Scale–Revised within a comprehensive neuropsychological battery. Neuropsychology, 6, 239–249.

Loehlin, J. C. (2004). Latent variable models: An introduction to factor, path, and structural equation analysis (4th ed.). Mahwah, NJ: Erlbaum.

MacCallum, R. C., Browne, M. W., & Cai, L. (2006). Testing differences between nested covariance structure models: Power analysis and null hypotheses. Psychological Methods, 11, 19–35.

McClosky, G., Perkins, L. A., & Van Diviner, B. (2008). Assessment and intervention for executive function difficulties. New York: Routledge.

McGrew, K. S., & Woodcock, R. W. (2001). Technical manual. Woodcock–Johnson III. Itasca, IL: Riverside Publishing.

Nevitt, J., & Hancock, G. R. (2004). Evaluating small sample approaches for model test statistics in structural equation modeling. Multivariate Behavioral Research, 39, 439–478.

Pennington, B. F., & Ozanoff, S. (1996). Executive functions and developmental psychopathology. Journal of Child Psychology and Psychiatry, 37, 51–87.

Phelps, L., McGrew, K. S., Knopik, S. N., & Ford, L. A. (2005). The general (g), broad, and narrow CHC stratum characteristics of the WJ III and WISC-III tests: A confirmatory cross-battery investigation. School Psychology Quarterly, 20, 66–88.

Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. Neuropsychology, 4, 532–545.

Salthouse, T. A., & Davis, H. P. (2006). Organization of cognitive abilities and neuropsychological variables across the lifespan. Developmental Review, 26, 31–54.

Schafer, J. L. (1997). Analysis of incomplete multivariate data. London: Chapman & Hall.

Sherman, E., Strauss, E., Spellacy, F., & Hunter, M. (1995). Construct validity of WAIS–R factors: Correlates in adults referred for evaluation of possible head injury. Psychological Assessment, 7, 440–444.

Sternberg, R. J., & Gardner, M. K. (1983). Unities in inductive reasoning. Journal of Experimental Psychology: General, 112, 80–116.

Tabachnick, B. G., & Fidell, L. S. (2007). Using multivariate statistics (5th ed.). Needham Heights, MA: Allyn & Bacon.

Taub, G. E., & McGrew, K. S. (2004). A confirmatory factor analysis of CHC theory and cross-age invariance of the Woodcock–Johnson Tests of Cognitive Abilities III. School Psychology Quarterly, 19, 72–87.

Thompson, B. (2004). Exploratory and confirmatory factor analysis: Understanding concepts and applications. Washington, DC: American Psychological Association.

Vernon, P. (1964). The structure of human abilities. London: Muthen and Co. Ltd.

Wechsler, D. (1974). Wechsler Intelligence Scale for Children–Revised. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1981). Wechsler Adult Intelligence Scale–Revised. New York: Psychological Corporation.

Willcutt, E. G., Pennington, B. F., Boada, R., Ogline, J. S., Tunick, R. A., Chhabildas, N. A., et al. (2001). A comparison of the cognitive deficits in reading disability and attention-deficit/hyperactivity disorder. Journal of Abnormal Psychology, 110, 157–172.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock–Johnson III Tests of Cognitive Abilities. Itasca, IL: Riverside Publishing.

Zelazo, P. D., & Mueller, U. (2002). Executive function in typical and atypical development. In U. Goswami (Ed.), Blackwell handbook of childhood cognitive development (pp. 445–469). Malden, MA: Blackwell.