

Beyond *g*: The Impact of *Gf-Gc* Specific Cognitive Abilities Research on the Future Use and Interpretation of Intelligence Tests in the Schools

Kevin S. McGrew
St. Cloud State University

Dawn P. Flanagan
St. John's University

Timothy Z. Keith
Alfred University

Mike Vanderwood
University of Minnesota

Abstract: This article describes recent developments in intelligence theory, applied measurement, and research methodology that suggest that new research is needed to reexamine the relative importance of both general and specific cognitive abilities in explaining school achievement. A summary of the results of a set of studies which examined the relationship between *g* and seven *Gf-Gc* specific abilities and general and specific reading and mathematics skills, is presented. These studies were designed to reexamine the *g* versus specific abilities issue in a manner that reflects progress in theory, measurement, and methodology. Analyses were conducted in separate model calibration and cross-validation samples ($n = 222$ to 255) at each of five grade levels (i.e., grades 1-2, 3-4, 5-6, 7-9, and 10-12). These studies used structural equation modeling procedures and tests from a validated *Gf-Gc* organized intelligence battery to operationalize a hierarchical *g + Gf-Gc* model consistent with Carroll's (1993) three stratum model of intelligence. Across all analyses, the relationship of *g* to general reading and math was as expected — significant and strong across all developmental levels. However, a number of significant and strong cross-validated direct effects for specific *Gf-Gc* abilities on specific reading and math skills also were found. The results of these studies suggest that (a) some specific abilities, including auditory processing, and fluid and crystallized intelligence, may be important for understanding the development of specific reading and math skills, above and beyond the understanding gained from general cognitive (*g*) and achievement constructs; (b) the predominant "Just say no" to subtest analysis position presented in much of the school psychology literature may need to be reevaluated; and (c) practitioners need to broaden their assessments beyond traditional instruments (e.g., Wechsler, 1991), in order to measure specific cognitive abilities that may be important in the development of specific reading and math skills.

Commonly used individually administered intelligence test batteries provide reliable and valid estimates (e.g., Full Scale IQ) of an underlying general ability construct (psychometric *g*) that is typically interpreted as an individual's global level of cognitive functioning. Considerable empirical evidence indicates that this general ability is "among the most dominant and enduring factors, both causal and corollary, associated with scholastic and occupational success; environmental adaptation; physical propensity and morbidity; and scientific, cultural, and political acumen" (McDer-

mott, Fantuzzo, & Glutting, 1990, p. 291). However, many practicing school psychologists place less emphasis on general ability in favor of specific intellectual constructs due to the belief that subtest scores yield useful diagnostic and treatment information (e.g., Kaufman, 1979, 1994). The practice of interpreting specific cognitive abilities has been encouraged in school psychology training programs through the use of general assessment texts (e.g., Kamphaus, 1993; Sattler, 1992) and specific intelligence test texts (Kaufman, 1979, 1990, 1994; McGrew, 1994) that present procedures for in-

Address correspondence concerning this article to Kevin S. McGrew, Department of Applied Psychology, Education Building A-253, St. Cloud State University, 720 4th Ave. So., St. Cloud, MN 56301.
mcgrew@tigger.stcloud.msus.edu

Copyright 1997 by the National Association of School Psychologists, ISSN 0279-6015.

interpreting subtest scores or patterns of subtests. Furthermore, recently revised and new intelligence test batteries emphasize both global and specific abilities (e.g., Differential Abilities Scale [DAS], Elliott, 1990; Kaufman Adolescent and Adult Intelligence Test [KAIT], Kaufman & Kaufman, 1993; Stanford-Binet Intelligence Scale-Fourth Edition [SB-FE], Thorndike, Hagen, & Sattler, 1986; Woodcock-Johnson Tests of Cognitive Ability-Revised [WJ-R COG], Woodcock & Johnson, 1989).

As a result, prevailing practice supports the belief that a multiple ability model of intelligence has greater predictive and clinical utility than a general factor model — hence, “the one versus the many” debate. Although the “many abilities” position has intuitive appeal and tends to prevail in practice, it is said to lack scientific merit (Bracken, McCallum, & Crain, 1993; McDermott, Glutting, Fantuzzo, Watkins, & Baggaley, 1992). What do the data say?

The Failure of Individual Intelligence Test Interpretation: Stuck on *g*

“Just say no to subtest analysis.” With the use of this catchy phrase, Glutting, McDermott, Watkins, Kush, and Konold (this issue), McDermott and Glutting (this issue) and McDermott et al. (1990) capture the prevailing position of academic school psychology that there is little empirical support for the practice of ipsative analysis and interpretation of intelligence tests. As popularized by Kaufman (1979, 1990) for the Wechsler batteries, ipsative interpretation is the process of generating strength and weakness hypotheses based on the analysis of individual subtest scores or patterns of subtest scores that demonstrate significant positive or negative deviations from the average (mean) of all the subtests. In their original “Just say no” article (McDermott et al., 1990) as well as in articles included in this issue, McDermott and colleagues presented data that they interpreted as providing little evidence to support any form of intelligence test subtest analyses by school psychologists (see also McDermott et al., 1992). Although a variety of sources of evidence were cited by McDermott and colleagues, probably the most powerful nail

they pounded into the coffin of subtest interpretation was the failure of ipsatized Wechsler subtest scores to add anything significant to the prediction of achievement beyond that accounted for by the global Full Scale IQ. McDermott et al. (1990) stated that

to the extent that the ultimate purpose of any psychological measure rests on its ability to improve prediction, the observation that ipsative scores fail to exceed, match, or even approach conventional scores in predictive efficiency effectively vitiates any claim that ipsative assessment has relative merit. (p. 293)

McDermott and colleagues’ “Just say no” position is similar to the conclusion of others. Cronbach and Snow’s (1977) review of the aptitude-treatment interaction (ATT) research found that treatments interacted primarily with general level of intelligence, and that few, if any, meaningful specific-level ability-treatment interactions had been established. Reviewing the research on the effectiveness of multiscore aptitude batteries (viz., Differential Aptitude Test Battery; Army Classification Battery; General Aptitude Test Battery) in the prediction of different outcomes, McNemar (1964) concluded that “the worth of the multitest batteries as differential predictors of achievement in school has not been demonstrated” (p. 875). In general, the extant literature on the use of differentiated measures of specific abilities has found that “*g* accounts for all of the significantly predicted variance; other testable factors, independently of *g*, add practically nothing to the predictive validity” (Jensen, 1984, p. 101). Across studies that have investigated the effectiveness of how measures of *g* and specific abilities predict educational and non-educational outcomes, the results suggest that we are “stuck on *g*.” The inability to move beyond *g* has provided little hope for the development of educational interventions designed according to an individual’s profile of specific ability strengths and weaknesses.

The failure to demonstrate that measures of specific abilities add anything useful to the prediction of school achievement, plus the current zeitgeist in academic

scho
intel
little
tant
cogn
In fa
psyc
polit
the
eral
ticol
genc
far in
genc
som
are
(anti
exar
of re
tellig
sure
metl
ating
and
and

opm
sure
desc
new
relat
cific
mer
mos
sche
mea
con
Wec
sear
emp
to ea
ing
spec
achi
mar
sear
wer
wea
sugg
cific
lum
then
tion

school psychology that may be largely anti-intelligence testing in nature, has instilled little optimism for ever identifying important links between measures of specific cognitive abilities and school achievement. In fact, we believe that some of our school psychology colleagues might argue that it is politically incorrect to conduct research on the usefulness of intelligence tests in general and subtest level interpretation in particular. We believe that before the intelligence testing pro/con pendulum swings too far in the direction of "Just say no to *intelligence tests*," there is a need to reexamine some of the pivotal research findings that are driving the pendulum in its current (anti-intelligence testing) direction. This re-examination is particularly critical in light of recent developments in (a) theories of intelligence, (b) theoretically organized measures of intelligence, and (c) research methodology that is better suited to evaluating the relative influence of both general and specific abilities within a theoretical and explanatory analytic framework.

With the aid of Figure 1, recent developments in intelligence theory, applied measurement, and research methodology are described. These developments suggest that new research is needed to reexamine the relative importance of both general and specific abilities in explaining school achievement. This section highlights the fact that most of the anti-specific ability research in school psychology has been conducted with measures that are based on an outdated conceptualization of intelligence (viz., the Wechsler batteries) and have employed research methods that have placed primary emphasis on *prediction* with little attention to *explanation* and *theoretical understanding* of the relations between general and specific cognitive abilities and school achievement. That is followed by a summary of select results from a series of research studies in reading and math that were designed to address these research weaknesses. The results of these studies suggest that the momentum of the anti-specific ability and anti-intelligence test pendulum should be slowed or even reversed. We then present a discussion of the implications of this research for the future of indi-

vidual intelligence testing in the schools. We conclude that there yet may be hope for moving "beyond *g*," a conclusion that is consistent with Carroll's (1993) position that "there is no reason to cease efforts to search for special abilities that may be relevant for predicting learning" (p. 66).

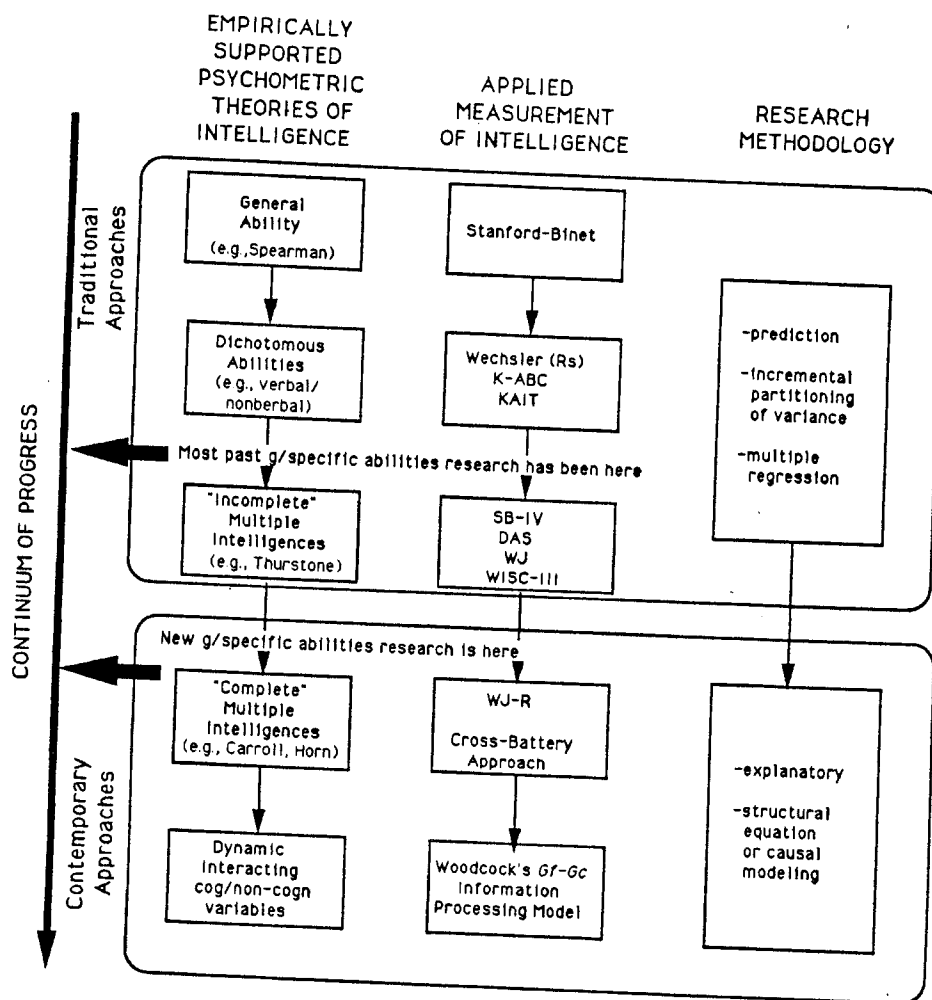
Moving Beyond *g*: The Use of Better Theory, Measurement, and Methodology

Advances in Intelligence Theory

Several theories of intelligence have received increased attention in psychology and education in recent years. Among the most prominent are Carroll's three stratum theory of cognitive abilities, Gardner's theory of multiple intelligences, the Horn-Cattell fluid-crystallized (*Gf-Gc*) theory of intelligence, the Luria-Das model of information processing, and Sternberg's triarchic theory of intelligence (see Flanagan, Genshaft, & Harrison, 1997, for a review). Each of these theories provides a framework from which to understand the multidifferentiated structure of cognitive abilities and the interrelation among them. Based on our review of the empirical evidence, it is apparent that the Horn-Cattell *Gf-Gc* Theory (Horn, 1991, 1994; Horn & Noll, 1997) and the three-stratum theory of cognitive abilities (Carroll, 1993, 1997) represent the most comprehensive and well researched frameworks of the structure of intelligence to date (also see Flanagan & McGrew, 1996, 1997; McGrew & Flanagan, 1996a, 1996b; Messick, 1992).

Both the Horn-Cattell and Carroll theories represent the culmination of a lengthy and ongoing effort of many theorists and researchers, most working from the dominant *psychometric* perspective (Neisser et al., 1996), who have attempted to specify a "complete" taxonomy of human cognitive abilities. The first column in Figure 1 represents the progress that has occurred in the development of the empirically supported psychometric theories of intelligence. Since Spearman's (1904, 1927) presentation of the general factor or *g* theory of intelligence, the psychometric research tradition has pushed the "many abilities" position from the specification of relatively simple di-

Figure 1. Theoretical, measurement, and methodological advances that indicate a need to reexamine the relations between *g*-specific abilities and school achievement.



chotomous models (e.g., verbal/nonverbal) to multiple but "incomplete" multiple intelligence models (e.g., Thurstone's primary mental abilities), that have since evolved into the "complete" (in a relative sense; no theoretical model is ever complete) multiple ability or intelligence models of Horn and Carroll (Woodcock, in press). While research continues to focus on the identification of the major abilities in the multiple intelligences taxonomy (Carroll, 1993), a number of researchers are attempting to move farther along the intelligence theory continuum through the specification of models that describe and explain cognitive

performance as a dynamic integration of both cognitive and noncognitive variables (e.g., see the work of Snow, 1986, and Woodcock, in press).

Reviews of the extant factor analytic research conducted during the past 50 to 60 years have converged on the *Gf-Gc* multiple intelligences taxonomy that serves as the organizational framework for both the Carroll and Horn-Cattell models (Carroll, 1983, 1989, 1993; Gustafson, 1984, 1988; Horn, 1988, 1991, 1994; Lohman, 1989; Snow, 1986). *Gf-Gc* theory has evolved through the analyses of several hundred data sets that were not restricted or limited to the cogni-

Table 1
Gf-Gc Broad Cognitive Ability Definitions

Gf-Gc Ability	Gf-Gc Symbol	Definition
Fluid Reasoning	Gf	Ability to reason, form concepts, and problem solve using novel information and/or procedures
Crystallized Intelligence	Gc	Breadth and depth of general knowledge and knowledge of a culture including verbal communication and reasoning with previously learned procedures
Visual Processing	Gv	Ability to analyze and synthesize visual information
Auditory Processing	Ga	Ability to analyze and synthesize auditory information
Processing Speed	Gs	Ability to perform quickly automatic cognitive tasks, particularly when under pressure to maintain focused concentration
Short-Term Memory	Gsm	Ability to hold information temporarily in immediate awareness and then use it within a few seconds
Long-Term Retrieval	Glr	Ability to store information and retrieve it later through association
Quantitative Knowledge	Gq	Ability to comprehend quantitative concepts and relationships and to manipulate numerical symbols
Correct Decision Speed	CDS	Quickness in providing correct answers to a variety of moderately difficult problems in comprehension, reasoning, and problem solving

tive ability tests of any specific intelligence test battery. Following from the work of Raymond Cattell (1941), Horn conducted a program of Gf-Gc research that, to date, has resulted in the identification of nine broad cognitive abilities (see Horn, 1991, 1994; Horn & Noll, 1997): Fluid Intelligence (Gf), Crystallized Intelligence (Gc), Short-Term Acquisition and Retrieval (Gsm), Visual Processing (Gv), Auditory Processing (Ga), Long-Term Storage and Retrieval (Glr), Cognitive Processing Speed (Gs), Correct Decision Speed (CDS), and Quantitative Knowledge (Gq). A brief description of these Gf-Gc abilities is provided in Table 1. Following an extensive review and reanalysis of most of the theoretical and empirical research on human cognitive abilities and their measurement since the early 1900s, Carroll (1993) concluded that "[t]he Cattell-Horn model ... is a true hierarchical model covering all major domains of intellectual functioning ... among available models it appears to offer the most well-founded and reasonable approach to an acceptable the-

ory of the structure of cognitive abilities" (p. 62).

Carroll's (1993) seminal work culminated in a three-stratum theory of cognitive abilities that, according to Snow (1993), "defines the taxonomy of cognitive differential psychology for many years to come" (cf. Carroll, 1993, p. back cover). The three strata in Carroll's framework differ in degree of generality. *General* cognitive ability or *g* is located at *stratum III*. This general ability subsumes eight *broad* cognitive abilities (located at *stratum II*), which, in turn, subsume 69 *narrow* abilities (located at *stratum I*). The broad (stratum II) cognitive abilities identified by Carroll include Fluid Intelligence (Gf), Crystallized Intelligence (Gc), General Memory and Learning (Gy), Broad Visual Perception (Gv), Broad Auditory Perception (Gu), Broad Retrieval Ability (Gr), Broad Cognitive Speediness (Gs), and Processing Speed/Reaction Time Decision Speed (Gt). The similarities between the broad (stratum II) cognitive abilities proposed by Carroll and the Gf-Gc abilities

proposed by Horn are worthy of notice (see McGrew & Flanagan, 1996b). It is beyond the scope of this article to discuss either the Horn-Cattell *Gf-Gc* theory or the three stratum theory of cognitive abilities in more detail. Therefore, the reader is referred to Horn (1991, 1994), Horn and Noll (1997), and Carroll (1993, 1997) for a comprehensive description of these theoretical frameworks, respectively.

Support for the *Gf-Gc* constructs underlying the Horn and Carroll models has been documented extensively through many forms of validity evidence, including (a) structural — individual differences in factor analytic investigations, (b) developmental — changes in cognitive abilities across age, (c) neurocognitive — relations to indicators of physiological and neurological functioning, (d) achievement — predictions of academic capability and occupational success, and (e) heritability — relations among individuals who are related biologically to differing degrees (Horn, 1994; Horn & Noll, 1997). In general, *Gf-Gc* theoretical models are based on a more comprehensive network of validity evidence than other contemporary multidimensional models of intelligence (e.g., Gardner's theory of multiple intelligences, Sternberg's triarchic theory) (Messick, 1992).

Although there is considerable support for the *Gf-Gc* theory, it is not without notable limitations. According to Horn and Noll (1997), some of the major limitations of *Gf-Gc* theory are that (a) the theory is largely a descriptive empirical generalization of research findings and much less a deductive explanation of these findings, (b) the structure implied for the *Gf-Gc* factors in rotated factor solutions is most likely not a good indication of the organization of actual human abilities, (c) the theory is largely a product of linear equations (viz., factor analysis), while natural phenomena most likely are nonlinear in nature, and (d) the theory provides little information on how the *Gf-Gc* abilities develop or how the cognitive processes work together.

Notwithstanding these limitations, given the breadth of empirical support for the *Gf-Gc* structure of intelligence, models based on *Gf-Gc* theory (i.e., the Horn-Cattell

and Carroll models) seem to provide the most useful frameworks for understanding cognitive functioning (Carroll, 1997; Flanagan & McGrew, 1995a, 1995b, 1997; McGrew, 1997; McGrew & Flanagan, 1996a, 1996b; Messick, 1992; Woodcock, 1990), and for investigating the relations between general and specific abilities and important outcomes (e.g., school achievement). Therefore, when considering empirically supported theories of intelligence along a continuum (see Figure 1), it is clear that psychology has progressed considerably since intelligence was conceptualized as a single general ability or dichotomous abilities. Indeed, the contributions that both Horn and Carroll have made to the scientific literature are significant and provide a *map* of the currently known cognitive abilities that can be used to guide cognitive assessment, interpretation, test development, and research. Until cognitive science advances further, the work of Horn and Carroll provide the most complete empirically supported multiple intelligences frameworks to date. However, as described in the next section, most of our cognitive assessment batteries do not adequately measure the breadth of abilities included in the *Gf-Gc* model of intelligence.

Advances in Applied Measurement

The Wechsler scales have dominated the practice of applied intelligence testing for nearly 6 decades (Harrison, Kaufman, Hickman, & Kaufman, 1988; Wilson & Reschly, 1996). As a result, many psychologists conceptualize intelligence as either general ability (the Full Scale IQ), dichotomous abilities (verbal/nonverbal or verbal/performance) or a set of multiple constructs (verbal comprehension, perceptual organization, freedom from distractibility, processing speed), such as those underlying the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991). When these interpretations are considered within the context of current advances in theory and research on the structure of intelligence (see Figure 1), it is clear that they are "ill-fitted and no longer viable" (Flanagan & McGrew, 1996).

Alth
Wechsle
practice
expansi
and inte
been ref
sler bat
1993; K
1993). A
that son
sure a w
the Wecl
telligenc
ize the c
dated Gf

Usir
recent j
McGrew
1992; W
broad (e
derlie th
Additior
and Fla
abilities
major in

I) al
steps a
KAIT w
General
tive reas
based on
ability d
Grew, i
two qu
tors of
stratur

Mc
Flanaga
with the
current
broad r
cluded i
roll's (1
ligence.
tery inc
Gf-Gc r
(i.e., sul
hypothe
when tr
tively di
rion to
from jo
nitiy

Although the contributions that the Wechsler scales have made to research and practice in psychology are significant and expansive, advances in cognitive science and intelligence test technology have not been reflected in the revisions of the Wechsler batteries (Carroll, 1993; Kamphaus, 1993; Keith & Witta, in press; Sternberg, 1993). More importantly, despite the fact that some intelligence test batteries measure a wider range of cognitive abilities than the Wechsler scales, none of our existing intelligence batteries adequately operationalize the comprehensive and empirically validated Gf-Gc model of intelligence.

Using the Gf-Gc framework as a guide, recent joint factor analyses (Flanagan & McGrew, in press; McGhee, 1993; Stone, 1992; Woodcock, 1990) have identified the broad (stratum II) Gf-Gc abilities that underlie the major intelligence test batteries. Additionally, McGrew (1997) and McGrew and Flanagan (1996b) have classified the abilities measured by the subtests of the major intelligence tests at the narrow (stratum I) ability level. For example, the Logical Steps and Mystery Codes subtests of the KAIT were classified as measuring mainly *General Sequential Reasoning* (i.e., deductive reasoning) and *Induction*, respectively, based on Carroll's (1993) narrow, stratum I ability definitions (see also Flanagan & McGrew, in press). Thus, the KAIT includes two *qualitatively different* narrow indicators of Fluid Intelligence (Gf) — a broad, stratum II ability.

McGrew's (1997) and McGrew and Flanagan's (1996b) analyses indicate that, with the exception of the WJ-R, none of our current intelligence batteries assess the broad range of Gf-Gc abilities that are included in either Horn's (1991, 1994) or Carroll's (1993) models of the structure of intelligence. Furthermore, no intelligence battery includes enough qualitatively different Gf-Gc narrow stratum I ability indicators (i.e., subtests) to warrant the generation of hypotheses about all abilities. For instance, when the presence of two or three qualitatively different measures is used as a criterion to identify broad factors, it is evident from joint factor analysis studies that most cognitive batteries include only two or

three composite scores that may be interpreted as representing two or three broad Gf-Gc abilities.

Current intelligence batteries generally measure a limited number of broad Gf-Gc abilities adequately, including mainly Crystallized Intelligence (Gc), Visual Processing (Gv), and Short-Term Memory (Gsm). Interestingly, Fluid Intelligence (Gf), often referred to as the "hallmark" of intelligent behavior, is not represented (or is underrepresented) in most intelligence batteries (McGrew & Flanagan, 1996b, 1996c). Although the WJ-R was developed deliberately to operationalize contemporary Gf-Gc theory, and as such, measures validly eight Gf-Gc abilities (Keith, 1997; McGrew, Werder, & Woodcock, 1991; Woodcock, 1990; Yseldyke, 1990), some of these abilities (or constructs) are underrepresented. For example, *Glr* is represented by two measures of the *same* narrow ability — associative memory — rather than two qualitatively different abilities (see McGrew & Flanagan, 1996b and Messick, 1995 for a discussion). However, because of the *breadth* of coverage of Gf-Gc abilities provided by the WJ-R it is depicted in Figure 1 as a relatively "complete" measure of multiple intelligences.

Interpretation of the findings of joint factor analyses within the context of contemporary Gf-Gc theory has resulted in one unequivocal conclusion — the use of a single battery to measure intelligence will result in an incomplete assessment of important cognitive abilities (see Flanagan & McGrew, 1995a, 1996, in press; McGrew & Flanagan, 1996b). For example, Auditory processing (*Ga*), Long-Term Retrieval (*Glr*) and Fluid Intelligence (*Gf*) are not assessed when using the Wechsler scales or the Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983). The practice of conducting an incomplete assessment of intelligence via the use of a single battery may have significant implications for making diagnostic, classification, and placement decisions and for conducting theoretically sound research on intelligence.

More importantly, with regard to the present article, is the conclusion that most

of the research that has investigated the relation between *g* and specific abilities and school achievement has been conducted with intelligence test batteries that have not provided a comprehensive measurement of most *Gf-Gc* abilities. In particular, the "Just say no to subtest analysis" admonition is based largely on research with an intelligence battery (i.e., the Wechsler) that does *not adequately measure several Gf-Gc abilities*, such as Fluid Intelligence (*Gf*), Long-Term Retrieval (*Glr*), and Auditory Processing (*Ga*). Given the importance attached to fluid ability in the definition of intelligence and the growing body of literature that has demonstrated the critical importance of phonological awareness (*Ga*) in reading (Felton & Pepper, 1995; McBride-Chang, 1995; Torgesen, Wagner, & Rashotte, 1994; Wagner & Torgesen, 1987; Wagner, Torgesen, Laughon, Simmons, & Rashotte, 1993; Wagner, Torgesen, & Rashotte, 1994), this is a potentially significant gap in the Wechsler-based *g*-specific ability research.

As a result of this gap, we argue that the failure to find that specific abilities add anything to the prediction of achievement beyond that already provided by a *g* score from the Wechsler's may be a *correct* interpretation for the set of constructs measured by the Wechsler batteries, but may be a *premature generalization* to apply to all intelligence batteries, particularly those developed to measure the *Gf-Gc* model or those interpreted from a *Gf-Gc* "cross-battery" perspective (McGrew & Flanagan, 1996b). We believe that before the interpretation of specific abilities on intelligence tests is declared legally dead, the *g*-specific ability research needs to be reexamined in light of current theory and with instruments or combinations of instruments that provide a more accurate and complete measurement of most *Gf-Gc* abilities. Progress has been made in both theories and measurement of intelligence. This progress requires a second look at "the one versus the many" debate.

Advances in Research Methodology

As suggested in Figure 1, advances in research methodology also enable more valid and theoretically coherent analyses of

the effects of specific abilities on learning and other criteria. Previous research has relied primarily on ordinary multiple regression analysis (MR) to determine whether the use of specific abilities improves the prediction of achievement beyond their prediction by general ability. Such methods, although sophisticated at the time, also have notable limitations — limitations which are particularly telling when analyzing the effects of specific abilities.

MR requires a two-step process for the joint prediction of achievement from specific and general ability. Briefly, MR will not allow the prediction of a criterion from both a composite score (e.g., Full Scale IQ) and the components (e.g., the subtests) that comprise that composite score (i.e., the correlation matrix will be singular and its inverse cannot be found). Therefore, various "tricks" are required to conduct such analyses. So, for example, in one analysis McDermott et al. (1990) subtracted the average subtest score from each individual subtest (in effect removing *g* from each subtest), used these "ipsatized" subtest scores to predict achievement, and compared the variance explained by the ipsatized subtest scores to the variance explained by the unadjusted subtest scores in a separate regression. Thorndike (1986) compared the variance explained by all subtests to that explained, in a separate regression, by a general, or *g* factor. These procedures, although imaginative, do not allow a direct comparison of the effects of specific versus general abilities; a more suitable approach would be to analyze the effects of general and specific abilities *simultaneously*.

Previous research has generally attempted, directly or indirectly, to partition variance into that accounted for by a general factor or score versus that accounted for separately by the variations in subtest scores. (Interestingly, many such analyses have shown that specific abilities do indeed aid in prediction, but the variance explained is often discounted as trivial compared to that explained by general abilities.) Yet methodologists warn that variance partitioning is an inadequate method for determining the relative importance of the effects of different variables on a criterion

(Kenn,
7). Vari
stingy
tween
see also
pending
their o
the effe
hazur,
tained
varianc
as guid
(Pedha
Fin
marily
termin
ties in
rion b
ity. Th
fects
have l
howev
transl
cific a
to dev
dents
t &
e. ag
dicts
resear
know
ing co
R
struct
many
searc
deter
one c
sever
p. 2).
firms
(For
meth
(
the r
hypc
for e
of ir
with
first-
hier
eral
ing r

(Kenny, 1979, chap. 4; Pedhazur, 1982, chap. 7). Variance partitioning provides a "highly stingy ... measure of the relationship between two variables" (Abelson, 1995, p. 7; see also Rosenthal & Rubin, 1979), and, depending on the variables in the analysis and their ordering, can severely underestimate the effects of one variable on another (Pedhazur, 1982). Furthermore, "results obtained from an incremental partitioning of variance are of dubious value, if not useless, as guides for policy and decision making" (Pedhazur, 1982, p. 218).

Finally, previous research has been primarily *predictive* in nature, seeking to determine whether the focus on specific abilities improves the *prediction* of some criterion beyond that predicted by general ability. Theory and efforts at determining the *effects* of specific abilities on achievement have been noticeably absent. We believe, however, that predictive findings have little translation into practice. To translate specific ability research into practice, to use it to develop meaningful interventions for students with learning problems, an *explanatory* approach is needed. That is, it is not enough to know simply that ability "x" *predicts* reading comprehension; to translate research into practice it is necessary to know whether or not ability "x" *affects* reading comprehension.

Recent developments in latent variable structural equation modeling (SEM) obviate many of the weaknesses of previous research. SEM "is a multivariate method for determining the magnitude of influence of one or several presumed causes on one or several presumed effects" (Keith, in press, p. 2). SEM subsumes such methods as confirmatory factor analysis and path analysis. (For more information concerning the method, see Hoyle, 1995, or Keith, in press.)

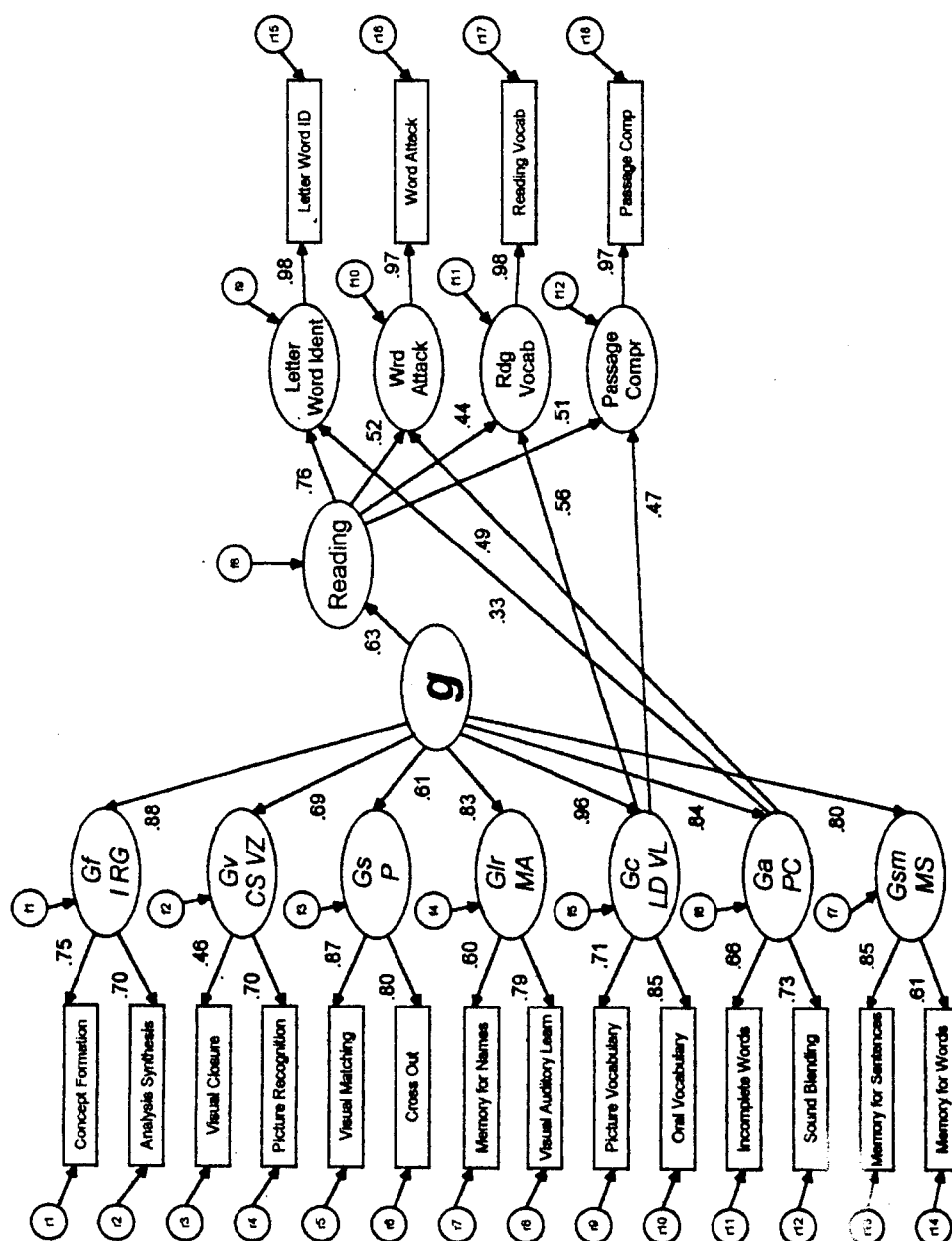
One advantage of SEM is that it allows the modeling of quite complex theories and hypotheses. The model shown in Figure 2, for example, proposes a hierarchical model of intelligence that generally is consistent with Carroll's model, with Gf-Gc abilities as first-order factors and *g* as a second-order, hierarchical factor. The factor *g* affects general reading ability — itself a factor affecting more specific reading skills — in a man-

ner consistent with claims that it is general ability that affects students' achievement. In addition, however, several of the specific abilities are hypothesized to affect specific reading achievements. Crystallized Intelligence (Gc), for example, is hypothesized to affect Passage Comprehension above and beyond the effects of *g* on Reading. Thus, unlike previous (MR) research, SEM allows for a more direct test of complex models, rather than the indirect, incomplete testing of those models. SEM also allows the testing of specific hypotheses of the influence of the effects of specific abilities rather than the blanket, blind prediction of achievements.

Also, SEM allows the specification of models (such as that represented in Figure 2) which examine the effects of specific abilities that are represented by the reliable *common Gf-Gc* ability variance shared between subtests that measure aspects of the same construct and that have been shown to be valid indicators of that construct. This is a significant difference from the use of ipsatized subtests in the MR approach which results in specific abilities being represented by the *residual* variance of a single subtest. That is, in the MR approach specific abilities are defined by a combination of error variance (or unreliability) and unique variance (or invalidity) — variance that remain after the reliable common (valid) variance among subtests is removed. SEM reduces the confounding effects of error by estimating the variation in measured variables due to error and by removing it from consideration of the effects of one variable on another. A related advantage is that SEM gets closer to the *constructs* of primary interest in research. All researchers are interested in the constructs underlying their imperfect measures, but must generally be satisfied with those imperfect measures as approximations of those constructs. SEM, by removing error (unreliability) and unique variance (invalidity) from the structural model, more accurately estimates the effects of one construct to another. Finally, SEM focuses on the analyses of effects rather than the partitioning of variance.

We argue, then, that the method used in the research summarized in this article

Figure 2. Structural Equation Model of the influence of g and specific abilities on reading achievement. The rectangles are measured variables and the ovals represent latent or unmeasured variables/factors. The arrows from the large ovals to the rectangles define the *measurement model*, a confirmatory factor analysis of the latent and measured variables. The small circles labeled r_1 through r_{18} represent the residual variance (unique and error) of the measured variables. The arrows from one oval to another define the *structural model* that represents the standardized effects of the latent variables on one another. For example, the arrow from g to reading has a *structural coefficient* of .63, which is interpreted to indicate that a standard deviation change in g is associated with .63 standard deviation change in general reading. All other structural coefficients are interpreted in a similar manner. Finally, the small circles labeled f_1 through f_{12} represent all other influences on the latent variables besides those shown in the model.



avoids t net
may have conf
ous research. S
pare directly, i
effects of gene
method focus
able on anothe
theoretically d
predictive app
moves the ubi
more closely
fects of the co

Spec
are

Recently,
ducted to ex
both g and sev
general and sp
(Flanagan, Ke
1996, McGre
Keith, 1996).
to reexamine
issue in a ma
in theory, me
summarized

Briefly, th
across the
1-2, 3-4, 5-6,
ples were dra
representative WJ
al., 1991). For
jects who had
1 thru 14) ar
lected. At ea
samples, sep
idation samp
The design o
the illustrati
Figure 2. T
used SEM pr
validated Gf
tery (i.e., W
chical $g + G$
roll's three
(Bickley, Ke
Reading and
were opera
indicators
tests), and
cific reading
lowing the
and Balke

avoids the methodological difficulties that may have confounded the results of previous research. SEM is better suited to compare directly, in the same model, the relative effects of general and specific abilities. The method focuses on the effects of one variable on another, and SEM is an explanatory, theoretically driven approach rather than a predictive approach. Furthermore, SEM removes the ubiquitous effects of error and more closely approaches a test of the effects of the constructs on one another.

Specific Abilities and *g* are Both Important

Recently, a set of studies were conducted to examine the relations between both *g* and seven Gf-Gc specific abilities and general and specific reading and math skills (Flanagan, Keith, McGrew, & Vanderwood, 1996; McGrew, Vanderwood, Flanagan, & Keith, 1996). These studies were designed to reexamine the *g* versus specific abilities issue in a manner that reflects the progress in theory, measurement, and methodology summarized above.

Briefly, these analyses were conducted across five grade categories (i.e., grades 1-2, 3-4, 5-6, 7-9, and 10-12). The five samples were drawn from the nationally representative WJ-R norm sample (McGrew et al., 1991). For each analyses, only those subjects who had complete cognitive (viz., tests 1 thru 14) and achievement tests were selected. At each of these five grade-based samples, separate calibration and cross-validation samples were used ($n = 222$ to 255). The design of these analyses is reflected in the illustrative causal model presented in Figure 2. The causal modeling analyses used SEM procedures and measures from a validated Gf-Gc organized intelligence battery (i.e., WJ-R) to operationalize a hierarchical *g* + Gf-Gc model consistent with Carroll's three stratum model of intelligence (Bickley, Keith, & Wolfe, 1995; Keith, 1997). Reading and math achievement constructs were operationalized by reliable and valid indicators (i.e., WJ-R reading and math tests), and organized into general and specific reading and math skill hierarchies following the recommendations of Gustafson and Balke (1993). Based on a review of

prior theoretical and research literature, an initial model that included a path from *g* to general reading or mathematics and paths from select specific Gf-Gc abilities to specific reading and math subskills was specified in each calibration model. Through an iterative model readjustment or "trimming" procedure, final models were selected that showed good model fit and that contained only significant structural paths. These final calibration models (which fit the data as well as or better than the single *g* models when various fit statistics were compared) were then cross-validated in an independent sample at each grade level. For example, the results presented in Figure 2 indicate that in the grade 1-2 sample, *g* had a significant and cross-validated effect on reading (.63). In addition, however, *Ga* had significant, cross-validated effects on Letter-Word Identification (.33) and Word Attack (.49) above and beyond the effects of *g* on reading. *Gc* had significant, cross-validated effects on Reading Vocabulary (.56) and Passage Comprehension (.47) above and beyond the effects of *g* on reading.

Although *g* obviously plays an important role in reading (.63) in this model, specific Gf-Gc abilities also were found to be significantly associated with specific forms of reading achievement. The results suggest that reading subskills could improve by $\frac{1}{3}$ (.33) to more than $\frac{1}{2}$ (.56) of a standard deviation for each standard deviation increase in the respective *Ga* or *Gc* ability, after the effect of *g* is already accounted for. Also, these effects are very consistent with the extant reading literature that has reported significant relations between general language (*Gc*) abilities and phonological awareness/processing (*Ga*) and reading (Felton & Pepper, 1995; McBride-Chang, 1995; McGrew, 1994; Wagner & Torgesen, 1987; Wagner et al., 1993; Wagner et al., 1994).

Across all the reading and mathematics concurrent analyses, the relation of *g* to general reading and mathematics was as expected — significant and strong across all developmental levels. In reading, the *g* direct effects on general reading were strong across developmental levels (ranging from .57 to .88) (McGrew, Vanderwood, et al.,

1996). The *g* effect on general mathematics was slightly lower and consistently ranged from .43 to .56 (with the exception of a lower .23 effect at grades 5-6) (Flanagan et al., 1996). Of greater interest was the finding of a number of significant and strong cross-validated direct effects for specific *Gf-Gc* abilities for specific reading and mathematics skills in these studies. Given the space limitation of this article, we present graphic summaries of select cross-validated results of the specific *Gf-Gc* effects. These results are presented in Figures 3 through 5.

The results summarized in Figure 3 indicate that *Ga* abilities were significantly related to Word Attack skills (pronunciation or decoding of unfamiliar words). The *g* factor displayed a significant indirect effect on Word Attack skills (as mediated through its strong direct effect on Reading), which in turn had a consistently strong direct effect on Word Attack at all grade levels. Even with the *general* direct and indirect effects present, *Ga* abilities contributed to the explanation of Word Attack skills as reflected by significant direct effects ranging from approximately .20 to .50 in grades 1 through 9. Thus, both general and specific abilities (viz., *Ga*) were found to be important for understanding Word Attack skills.

The results presented in Figure 4 for Passage Comprehension (a person's skill in identifying a key word missing from a reading passage) reflect the importance of a different specific ability (i.e., *Gc*), and reveal potentially important developmental trends. Although the effects of *g*, general reading, and *Gc* were all significant during grades 1-6, the general effects (i.e., *g* and general reading) decreased in importance and became nonsignificant after grade 6, whereas the specific *Gc* ability effect gradually increased with age. Interestingly, *Gc* was the strongest of all effects on Passage Comprehension at all grades. These results again highlight the importance of a specific *Gf-Gc* ability and indicate that in both research and practice it may be important to recognize that specific and general ability effects may vary as a function of developmental level.

Finally, the mathematics results presented for Applied Problems (Figure 5) sug-

gest that a person's skill in analyzing and solving practical mathematics problems may be a function of a complex combination of multiple general and specific abilities, and these effects may change with age. The indirect and direct effects for the general constructs (i.e., *g* and general mathematics) increased in importance with age. *Gs* abilities, although less consistent in significance, also demonstrated a trend that suggested increased importance with age. In contrast, the specific abilities of *Gf* and *Gc* decreased in relative importance with increasing age. An interesting developmental observation was the finding that *Gf* abilities were much more strongly associated with solving applied mathematics problems at grades 1-2 than any other general or specific ability, with its importance decreasing and becoming nonsignificant after grade 9.

The results summarized in Figures 3 through 5 suggest a number of important conclusions. First, some specific abilities (e.g., *Ga*, *Gf*, and *Gc*) may be important for understanding the development of specific reading and mathematics skills, above and beyond the understanding gained from general cognitive (*g*) and achievement constructs. That is, *both general and specific abilities are important* in understanding reading and mathematics achievement. And, the specific effect sizes (many from approximately .30 to .50) are not trivial. These effects are much larger than those that have been judged to have enormous practical and scientific significance (i.e., effect sizes of .10 to .20) in the medical and behavioral sciences (Gage, 1996).

Second, the relative importance of both general and specific abilities change developmentally, and these changes may in turn vary (either increasing or decreasing in importance) as a function of the specific academic skill being investigated. Third, for some academic skills (e.g., Word Attack and Passage Comprehension) there may exist only one specific cognitive ability (e.g., *Ga* or *Gc*) that makes a consistent contribution to the understanding of that academic skill beyond the understanding gained from the knowledge of a person's general abilities, while for other academic skills (e.g., applied mathematics problems) a complete under-

Figure 3. Significant *g*-specific abilities effects on Word Attack (Reading) across five school-age samples. The x-axis presents the three latent factors (i.e., *g*, Reading, and *Ga*) that had significant cross-validated effects on the latent Word Attack factor. The size of the effects by sample are represented by the vertical bars. The *g* effects are indirect because these effects are mediated through their strong direct effects on general Reading, which in turn had consistently strong direct effects on Word Attack (the *Reading* effect section of the figure).

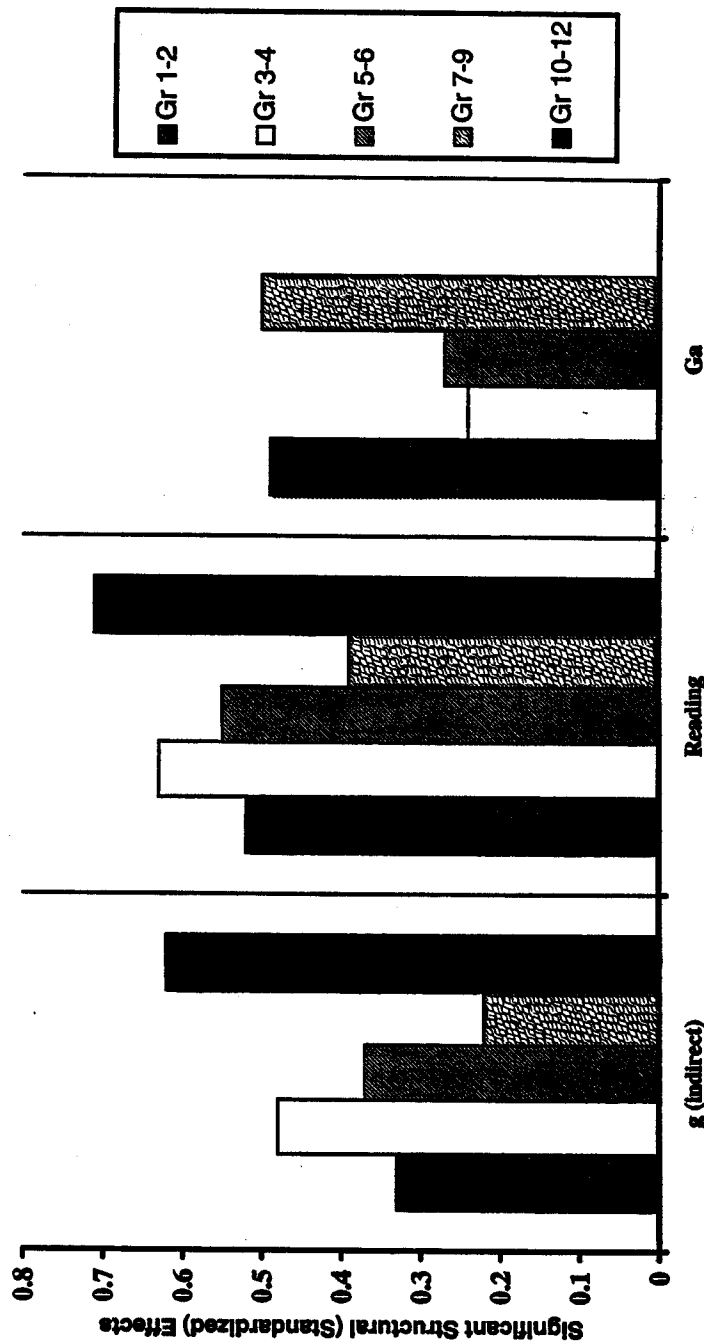


Figure 4. Significant *g*-specific abilities effects on Passage Comprehension (Reading) across five school-age samples. The x-axis presents the latent factors (i.e., *g*, Reading, *Gc*, and *Gs*) that had significant cross-validated effects on the latent Passage Comprehension factor. The size of the effects by sample are represented by the vertical bars. The *g* effects are indirect because these effects are mediated through their strong direct effects on general Reading, which in turn had consistently strong direct effects on Passage Comprehension (the *Reading* effect section of the figure).

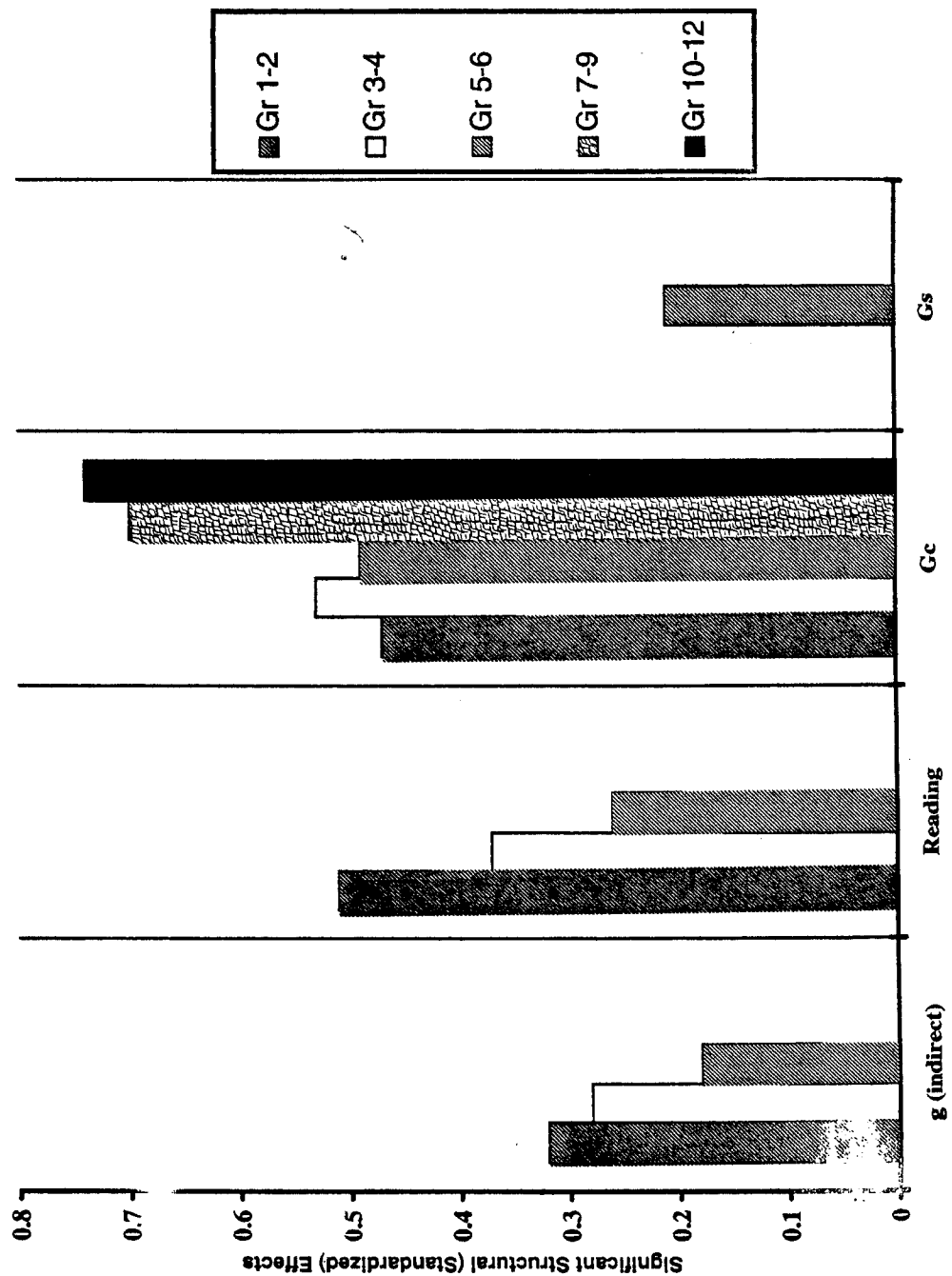
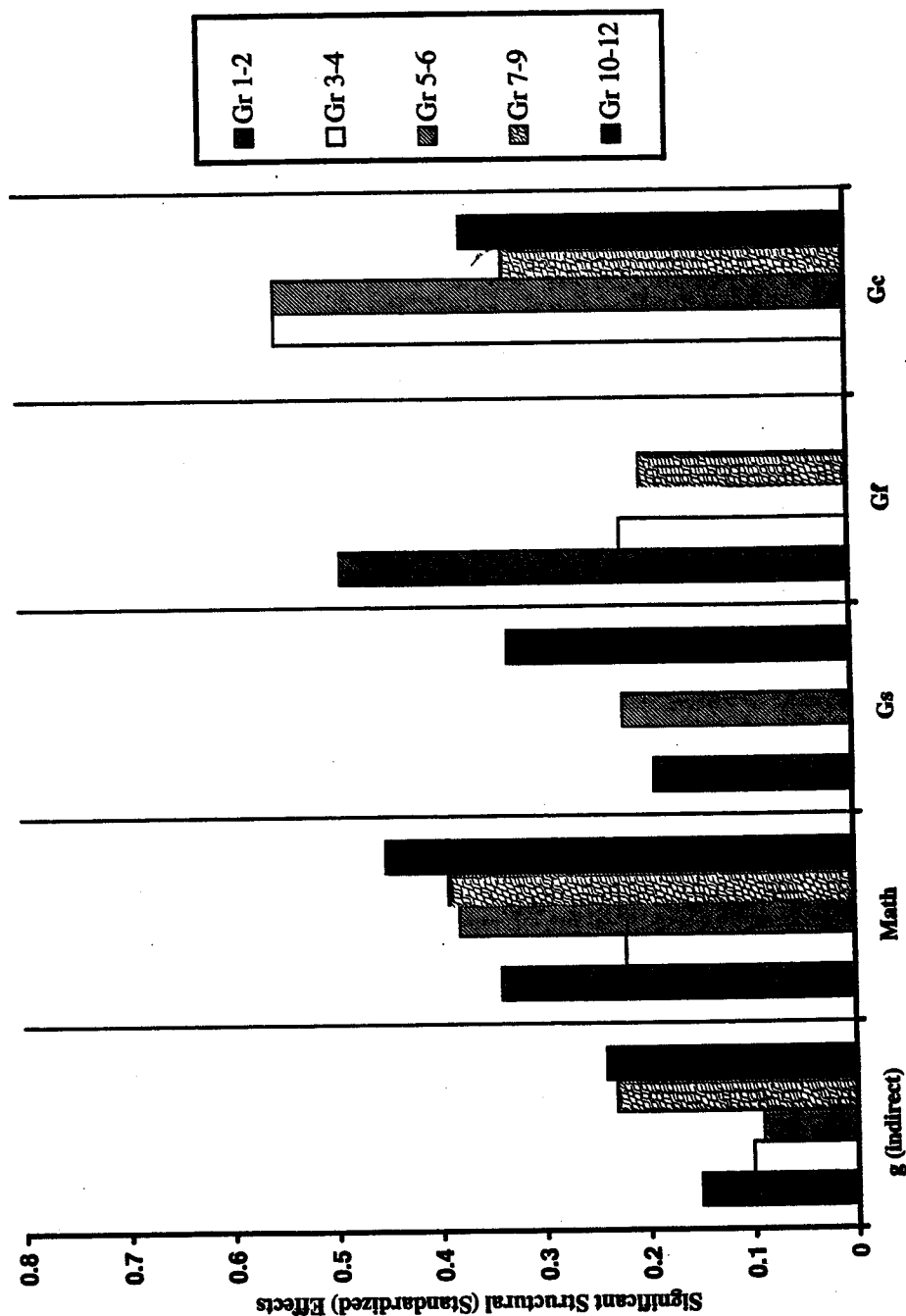


Figure 5
across fi
ics, *Gs*,
lems fac
effects i
fects on
plied Pr

Figure 5. Significant *g*-specific abilities effects on Applied Problems (Mathematics) across five school-age samples. The x-axis presents the latent factors (i.e., *g*, Mathematics, *Gs*, *Gf*, *Gc*) that had significant cross-validated effects on the latent Applied Problems factor. The size of the effects by sample are represented by the vertical bars. The *g* effects are "indirect" because these effects are mediated through their strong direct effects on general Mathematics, which in turn had consistently strong direct effects on Applied Problems (the *Math* effects section of the figure).



standing may rest on the complex combination of a number of general and specific abilities. Finally, past research that was based on intelligence test batteries (viz., Wechsler scales) that did not include indicators of important abilities (e.g., *Ga* for Word Attack; *Gf* for applied problems), which focused on samples with too broad or too limited an age range of subjects, and which did not use methodology that allowed for the examination of the simultaneous effects of both general and specific cognitive abilities, may have produced an incomplete conclusion regarding the *g*-specific abilities issue.

The above results, which, of course, need additional replication, suggest that the "Just say no" to specific abilities admonition in school psychology may need to be revised to "Just say maybe." Recent advances in theory, applied measurement, and research methodology suggest that the *g*-specific abilities issue is complex, but it may deserve a reexamination if research and practice are guided by an empirically supported framework (e.g., contemporary *Gf-Gc* theory) within a developmental context, and if sound applied measurement and research methodologies are employed.

Implications

Implications for Practice

The research summarized in this article suggests a number of implications for practice. First, practitioners should recognize that most of the negative findings reported in the school psychology literature regarding the importance of specific abilities may be due, in part, to the use of an intelligence battery that does not assess certain *Gf-Gc* abilities that appear to be important for school achievement (viz., Wechsler scales). If specific ability interpretation is to survive as an important and empirically supported practice, the research presented here suggests that practitioners may need to broaden their assessments beyond the confines of most of their favorite test kits. For example, serious consideration may be given to assessment batteries that are designed specifically to measure the widest breadth of *Gf-Gc* abilities (e.g., the WJ-R

cognitive battery of tests). Alternatively, practitioners may broaden their assessments by supplementing their batteries with measures of important (yet missing) abilities through a *Gf-Gc* organized "cross-battery" assessment approach (Flanagan & McGrew, 1997; McGrew & Flanagan, 1996a, 1996b) — an approach that presents procedures and guidelines to maximize the breadth and depth of abilities measured while minimizing the technical difficulties (e.g., different norm groups) present when combining tests across different intelligence batteries.

Second, practitioners might consider conducting more focused and selective intellectual assessments. For example, in the reading research reviewed in this article, *Gv* abilities showed no effect on reading achievement. This raises questions about the value of spending significant time assessing *Gv* abilities for reading-related referrals. The research studies of Flanagan et al., (1996), and McGrew, Vanderwood, et al. (1996), together with the *Gf-Gc* achievement related literature reviewed by McGrew and Flanagan (1996b), highlight those potential specific *Gf-Gc* abilities that may or may not be included (or may receive more or less time and attention) in intellectual assessments as a function of the nature of the referral concerns. Selectively focused *Gf-Gc* organized intellectual assessments may help reduce the time practitioners spend in assessment activities and, more importantly, make this time potentially more valuable.

Third, if specific ability interpretation is to be a valuable practice in school and clinical psychology, current approaches need be modified. In addition to the use of *Gf-Gc* validated intelligence batteries or a *Gf-Gc* organized cross-battery approach, the focus on interpretation should not be on individual subtests. The positive specific ability findings reviewed here suggest that a more promising practice may be the interpretation of specific abilities that are operationally defined by the reliable common variance shared by at least two strong indicators of each specific ability. That is, individual subtest interpretation, which focuses on the *smallest* portion of a test's reliable

variance, may be replaced with *cluster interpretation*, which focuses on the *larger* reliable common Gf-Gc variance shared by two or more strong indicators of each broad Gf-Gc ability (Flanagan, Andrews, & Genshaft, 1997). This approach *increases the validity* of the constructs measured through greater *construct representation* and the reduction of *construct irrelevant variance* (McGrew & Flanagan, 1996b; Mesick, 1995).

Fourth, it is important for practitioners to recognize that the valid interpretation of specific abilities is a complex enterprise that needs to account for the finding that the effects of specific abilities may vary as a function of curriculum area and development. A specific Gf-Gc ability that is featured prominently in assessments and interpretations at younger ages may have little relevance, either diagnostically or instructionally, at a later age. For example, a review of Figure 3 suggests that school psychologists could pay particular attention to *Ga* abilities as they relate to basic reading skills during the early school years.

Finally, we would be remiss if we did not add a serious note of caution. The primary message from this article is that there may be hope for specific ability interpretation. However, practitioners should not misinterpret this research to support any form of individual subtest interpretation. The current results only suggest that *some* specific Gf-Gc abilities may be important for understanding *some* academic skills at *some* developmental levels. Also, these findings only generalize to interpretations grounded in contemporary Gf-Gc theory and to an approach that operationally defines specific abilities as a combination of at least two factorially strong measures of each Gf-Gc ability. Of critical importance is the fact that no empirical evidence to date substantiates the use of specific Gf-Gc ability scores in making diagnostic or treatment recommendations. That is, diagnostic and treatment validity is not yet available and needs to be determined. The current results, therefore, are suggestive, not definitive. Additional research is needed.

Implications for Multicultural Populations

Although the need for understanding the relation between intelligence theory, tests, school achievement, and multicultural variables has been rapidly increasing in importance, the research necessary to address these issues in the context of contemporary Gf-Gc theory is limited. Although the number of research studies is small, Carroll's review of the factor analytic research on cognitive abilities suggests that "with reference to the major types of cognitive ability, there is little evidence that factorial structure differs in any systematic way across male and female groups, different cultures, racial groups, and the like" (1993, p. 687). Support for the Horn-Cattell Gf-Gc model in particular, as operationalized by tests from the WJ-R and KAIT, in a non-White sample of elementary school students was recently reported by Flanagan and McGrew (in press). Thus, the available research suggests that, in general, the Gf-Gc model of intelligence reported in this article is most likely invariant across different cultural and racial groups.

However, currently there is no *g*-specific ability research similar to that summarized in this article that has been conducted with other cultural or racial groups. It is unknown whether the relative effects of *g* and specific abilities on reading and mathematics achievement is similar in different cultural or racial groups. There is a critical need to investigate whether the same specific Gf-Gc abilities affect the same specific academic skills across cultures. These findings would have significant implications with regard to diagnostic and instructional decision making across cultural groups. In addition, there is a need to extend the present research beyond the realm of Gf-Gc abilities by exploring whether noncognitive variables (e.g., motivation, cultural-attitudinal variables, exposure to mainstream American environmental influences, etc.) moderate (or change) the relations between specific cognitive abilities and achievement. Until that research is completed, implications for the relative influence of general and specific cognitive abilities on important outcome variables (e.g., academic achieve-

ment) for different cultural groups is unknown.

Implications for Research

Additional research is needed to confirm and extend the results outlined here. First, the findings presented here should be replicated with additional samples. One advantage of the SEM approach is that it is possible to constrain the models from one sample to be similar or even identical to those from another analysis, thus allowing a strong test of the replicability of findings. An advantage of the reading and math studies reviewed here is that replicability was built into the research design; the samples were split in half and the validation subsamples were used to replicate the findings of the calibration samples. Nevertheless, independent cross-validation with distinct samples would provide even firmer evidence of the importance of specific abilities. Second, researchers should use the methodology illustrated here to investigate the effects of specific abilities with other instruments. It will be particularly interesting to see if the influences demonstrated to be important in this research generalize to other instruments or combinations of instruments (e.g., the cross-battery approach).

Third, the influence of specific abilities on other criteria, especially other measures of learning, should be investigated. For example, the grades that students earn in school provide another, important measure of learning related to, but different from, achievement test scores. It will be important to determine the effects of specific *Gf-Gc* abilities on students' grades. Likewise, it will be important to examine the effects of specific abilities on other measures of learning more closely tied to the curriculum, such as teacher-made unit tests and curriculum-based assessments. Generalization of the effects shown here to other measures of learning would strengthen the conclusions of the importance of particular, specific abilities and will help in the translation of this research into practice. It will be interesting to see if the influences demonstrated in this research have stronger or

weaker effects on other measures of learning.

Finally, it is critical that the positive research summarized here be extended to well-designed studies of diagnostic and treatment validity. The research results presented here suggest that use of the Wechsler intelligence batteries (which do not measure a number of important *Gf-Gc* abilities found in this research) in most of the diagnostic and treatment validity research in school psychology and special education may have resulted in the specific ability position not having "its day in court." This point echoes the words of Ysseldyke (1990) who, when commenting on the WJ-R and *Gf-Gc* theory, said that "we may now be able to investigate the extent to which knowledge of pupil performance on the various factors is prescriptively predictive of relative success in school. That is, we may now begin to address treatment relevance" (p. 273). Appropriately designed aptitude treatment interaction (ATI) studies that reflect the *Gf-Gc* ability by curriculum area by developmental level interactions reported here are sorely needed.

Summary Comments

"Every tradition grows ever more venerable ... the reverence due to it increases from generation to generation. The tradition finally becomes holy and inspires awe" (Nietzsche, cited in Bartlett, 1992). Although the "Just say no" to the specific abilities position, a position that some have generalized to most all forms of intelligence test use in the schools, may not be a "tradition" that "inspires awe," we believe that this position is on the verge of becoming the politically correct position in academic school psychology. As a result, it may become a fundamental belief that is being imparted to new generations of school psychologists. It is important for the field of school psychology, a profession that prides itself on being grounded in the scientist-practitioner model, to adhere to the rules of the "scientist" half of the model. Just as the long-entrenched practice of interpreting specific abilities on intelligence tests was correctly challenged by systematic programs of research that resulted in the "Just say no" ad-

monition, so
"Just say no"
scrutinized.

Many im-
science are t
issues being
better theor
nology, and
lieve that th
in this artic
more empir
structure of
porary *Gf-G*
ment of the
ties, and m
(viz., struct
focused m
mental par
that the m
maybe" or
eral and sp
important for
of reading
children. I
specific ab
— effects
of *Gf-Gc* a
fac any
most rese
results su
to Ysseld
temporar
may fina
means to
specific
educatio
yet may

Abelson,
ment.

Bartlett,
Boston

Bickley,
three
the st
tellig

Bracker
WIS
ficti
Jou

Carroll
cog
ana

monition, so must the foundation of the "Just say no" position be systematically scrutinized.

Many improvements and discoveries of science are the result of old problems and issues being reexamined in light of new and better theories, better measurement technology, and improved methodology. We believe that the research studies summarized in this article, studies that are based on more empirically supported models of the structure of cognitive abilities (i.e., contemporary Gf-Gc theory), improved measurement of the complete range of Gf-Gc abilities, and more powerful research methods (viz., structural equation modeling methods focused more on explanation than incremental partitioning of variance), suggest that the more correct position is "Just say maybe" or "Wait just a minute." Both general and specific Gf-Gc abilities appear important for understanding different aspects of reading and mathematics in school-aged children. More importantly, many of the specific ability effects we have summarized — effects that are present *after the effects of g are accounted for* — are not trivial. In fact, many are quite large (e.g., .20 to .50) by most research standards (Gage, 1996). The results summarized in this article resonate to Ysseldyke's (1990) comment that contemporary Gf-Gc theory and measurement may finally provide psychologists with a means to address the relationship between specific Gf-Gc abilities and diagnosis and educationally relevant intervention. There yet may be hope for moving "beyond g."

References

- Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ: Lawrence Erlbaum.
- Bartlett, J. (1992). *Familiar quotations* (16th ed.). Boston: Little, Brown.
- Bickley, P. A., Keith, T. Z., & Wolfe, L. M. (1995). The three-stratum theory of cognitive abilities: Test of the structure of intelligence across the life span. *Intelligence*, 20, 309-328.
- Bracken, B. A., McCallum, S. R., & Crain, R. M. (1993). WISC-III subtest composite reliabilities and specificities: Interpretive aids. [WISC-III Monograph] *Journal of Psychoeducational Assessment*, 20-34.
- Carroll, J. B. (1983). Studying individual differences in cognitive abilities: Through and beyond factor analysis. In R. F. Dillon & R. R. Schneek (Eds.), *Individual differences in cognition* (Vol. 1, pp. 1-33). New York: Academic Press.
- Carroll, J. B. (1989). Factor analysis since Spearman: Where do we stand? What do we know? In R. Kanfer, P. L. Ackerman, & R. Cudeck (Eds.), *Abilities, motivation, and methodology* (pp. 43-67). Hillsdale, NJ: Lawrence Erlbaum.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, England: Cambridge University Press.
- Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 122-130). New York: Guilford.
- Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin*, 38, 592.
- Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods*. New York: Irvington.
- Elliott, C. D. (1990). *Differential Ability Scales*. San Antonio, TX: The Psychological Corporation.
- Felton, R. H., & Pepper, P. P. (1995). Early identification and intervention of phonological deficits in kindergarten and early elementary children at risk for reading disability. *School Psychology Review*, 24, 405-414.
- Flanagan, D. P., Andrews, J. T., & Genshaft, J. L. (1997). The functional utility of intelligence tests with special education populations. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 457-483). New York: Guilford.
- Flanagan, D. P., Genshaft, J. L., & Harrison, P. L. (Eds.) (1997). *Contemporary intellectual assessment: Theories, tests and issues*. New York: Guilford.
- Flanagan, D. P., Keith, T. Z., McGrew, K. S., & Vanderwood, M. L. (1996). *Is g all there is? An investigation of the relationship between Gf-Gc specific cognitive abilities and mathematics achievement in individuals from grades 1 through 12*. Manuscript in preparation.
- Flanagan, D. P., & McGrew, K. S. (1995a). The field of intellectual assessment: A current perspective. *The School Psychologist*, 49, 7, 14.
- Flanagan, D. P., & McGrew, K. S. (1995b, March). *Will you evolve or become extinct? Interpreting intelligence tests from modern Gf-Gc theory*. Paper presented at the 27th annual convention of the National Association of School Psychologists, Chicago, IL.
- Flanagan, D. P., & McGrew, K. S. (1996). A "cross-battery" approach to assessing and interpreting cognitive abilities: An alternative to the Wechsler tradition. Manuscript submitted for publication.
- Flanagan, D. P., & McGrew, K. S. (1997). A cross-battery approach to assessing and interpreting cognitive

- abilities: Narrowing the gap between practice and cognitive science. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 314-325). New York: Guilford.
- Flanagan, D. P., & McGrew, K. S. (in press). Interpreting intelligence tests from contemporary *Gf-Gc* theory: Joint confirmatory factor analyses of the WJ-R and KAIT in a non-white sample. *Journal of School Psychology*.
- Gage, N. L. (1996). Confronting counsels of despair for the behavioral sciences. *Educational Researcher*, 25(3), 5-22.
- Gustafson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, 8, 179-203.
- Gustafson, J. E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 35-71). Hillsdale, NJ: Lawrence Erlbaum.
- Gustafson, J. E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, 28(4), 407-434.
- Harrison, P. L., Kaufman, A. S., Hickman, J. A., & Kaufman, N. L. (1988). A survey of tests used for adult assessment. *Journal of Psychoeducational Assessment*, 6, 188-198.
- Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselrode & R. B. Cattell (Eds.), *Handbook of multivariate psychology* (rev. ed., pp. 645-685). New York: Academic Press.
- Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock (Eds.), *Woodcock-Johnson technical manual* (pp. 197-232). Chicago: Riverside.
- Horn, J. L. (1994). Theory of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 443-451). New York: Macmillan.
- Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: *Gf-Gc* theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 53-91). New York: Guilford.
- Hoyle, R. H. (Ed.). (1995). *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: Sage.
- Jensen, R. A. (1984). Test validity: *g* versus the specificity doctrine. *Journal of Social and Biological Structures*, 7, 93-118.
- Kamphaus, R. W. (1993). *Clinical assessment of children's intelligence*. Boston: Allyn & Bacon.
- Kaufman, A. S. (1979). *Intelligent testing with the WISC-R*. New York: Wiley & Sons.
- Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence*. Boston: Allyn and Bacon.
- Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. New York: Wiley & Sons.
- Kaufman, A. S., & Kaufman, N. L. (1983). *K-ABC: Kaufman Assessment Battery for Children*. Circle Pines, MN: American Guidance Service.
- Kaufman, A. S., & Kaufman, N. L. (1993). *The Kaufman Adolescent and Adult Intelligence Test manual*. Circle Pines, MN: American Guidance Service.
- Keith, T. Z. (1997). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 373-402). New York: Guilford.
- Keith, T. Z. (in press). Structural equation modeling in school psychology. In C. R. Reynolds & T. B. Gutkin (Eds.), *Handbook of school psychology* (3rd ed.). New York: Wiley & Sons.
- Keith, T. Z., & Witta, E. L. (in press). Hierarchical and cross-age confirmatory factor analysis of the WISC-III: What does it measure? *School Psychology Quarterly*.
- Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley & Sons.
- Lohman, D. F. (1989). Human intelligence: An introduction to advances in theory and research. *Review of Educational Research*, 59, 333-373.
- McBride-Chang, C. (1995). What is phonological awareness? *Journal of Educational Psychology*, 87(2), 179-192.
- McDermott, P. A., Fantuzzo, J. W., & Glutting, J. J. (1990). Just say no to subtest analysis: A critique on Wechsler theory and practice. *Journal of Psychoeducational Assessment*, 8, 290-302.
- McDermott, P. A., Glutting, J. J., Fantuzzo, J. W., Watkins, W. M., & Baggaley, A. R. (1992). Illusions of meaning in the ipsative assessment of children's ability. *The Journal of Special Education*, 25(4), 504-526.
- McGhee, R. (1993). Fluid and crystallized intelligence: Confirmatory factor analysis of the Differential Abilities Scale, Detroit Tests of Learning Aptitude-3, and Woodcock-Johnson Psycho-Educational Battery-Revised. [WJ-R Monograph]. *Journal of Psychoeducational Assessment*, 20-38.
- McGrew, K. S. (1994). *Clinical interpretation of the Woodcock-Johnson Tests of Cognitive Ability-Revised*. Needham Heights, MA: Allyn and Bacon.
- McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive *Gf-Gc* framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 151-179). New York: Guilford.

- McGrew, K. S., & Flanagan, D. P. (1996a, March). A cross-battery approach to intelligence test interpretation. Workshop presented at the 28th annual convention of the National Association of School Psychologists, Atlanta, Georgia.
- McGrew, K. S., & Flanagan, D. P. (1996b). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment*. Boston: Allyn and Bacon.
- McGrew, K. S., & Flanagan, D. P. (1996c). The Wechsler Performance Scale debate: Fluid intelligence (Gf) or visual processing (Gv)? *Communiqué*, 24(6), 14, 16.
- McGrew, K. S., Vanderwood, M. L., Flanagan, D. P., & Keith, T. Z. (1996). *Stuck on g? Gf-Gc specific cognitive abilities and their relationship to reading achievement from grades 1 through 12*. Manuscript in preparation.
- McGrew, K. S., Werder, J. K., & Woodcock, R. W. (1991). *WJ-R technical manual*. Chicago: Riverside.
- McNemar, Q. (1964). Lost: Our intelligence? Why? *American Psychologist*, 19, 871-872.
- Messick, S. (1992). Multiple intelligences or multilevel intelligence? Selective emphasis on distinctive properties of hierarchy: On Gardner's *Frames of Mind* and Sternberg's *Beyond IQ* in the context of theory and research on the structure of human abilities. *Psychological Inquiry*, 3, 365-384.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
- Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51(2), 77-101.
- Pedhazur, E. J. (1982). *Multiple regression in behavioral research: Prediction and explanation* (2nd ed.). New York: Holt, Rinehart, & Winston.
- Rosenthal, R., & Rubin, D. B. (1979). A note on percent variance explained as a measure of the importance of effects. *Journal of Applied Social Psychology*, 9, 395-396.
- Sattler, J. M. (1992). *Assessment of children*. San Diego, CA: Author.
- Snow, R. E. (1986). Individual differences and the design of educational programs. *American Psychologist*, 41, 1029-1039.
- Spearman, C. E. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Spearman, C. E. (1927). *The abilities of man: Their nature and measurement*. London: Macmillan.
- Sternberg, R. J. (1993). Rocky's back again: A review of the WISC-III [WISC-III Monograph]. *Journal of Psychoeducational Assessment*, 161-164.
- Stone, B. J. (1992). Joint confirmatory factor analysis of the DAS and WISC-R. *Journal of School Psychology*, 30, 185-195.
- Thorndike, R. L. (1986). Historical and theoretical perspectives. In R. L. Thorndike, E. P. Hagen, & J. M. Sattler (Eds.), *Technical manual: Stanford-Binet Intelligence Scale* (4th ed.) (pp. 1-7). Chicago: Riverside.
- Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *The Stanford-Binet Intelligence Scale: Fourth Edition technical manual*. Chicago, IL: Riverside.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities*, 27(5), 276-286.
- Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, 102(2), 192-212.
- Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology*, 85(1), 83-103.
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bi-directional causality from a latent variable longitudinal study. *Developmental psychology*, 30(1), 73-87.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-Third Edition*. San Antonio, TX: The Psychological Corporation.
- Wilson, M. S., & Reschly, D. J. (1996). Assessment in school psychology training and practice. *School Psychology Review*, 25, 9-23.
- Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment*, 8, 231-258.
- Woodcock, R. W. (in press). Extending Gf-Gc theory into practice. In J. J. McArdle, & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice*. Chicago: Riverside.
- Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psychoeducational Battery-Revised*. Chicago: Riverside.
- Ysseldyke, J. E. (1990). Goodness of fit of the Woodcock-Johnson Psychoeducational Battery-Revised to the Horn-Cattell Gf-Gc theory. *Journal of Psychological Assessment*, 8, 268-275.

Kevin S. McGrew is Professor of Applied Psychology at St. Cloud State University. His research interests include the measurement of intelligence and personal competence, applied psychometrics, and the development validation of models and measurements of intelligence, personal competence, and community adjustment.

Dawn P. Flanagan is Associate Professor of School Psychology, Department of Psychology, St. John's University, New York. Her research focuses on the structure of intelligence, psychoeducational and preschool assessment, and professional issues in school psychology.

Timothy Z. Keith is Professor in the Division of School Psychology at Alfred University. His research interests include the influences on school learning and the nature and measurement of intelligence. He also has a continuing interest in the methodologies of structural equation modeling and confirmatory factor analysis.

Mike Vanderwood is a doctoral candidate in the School Psychology Doctoral program at the University of Minnesota. He is currently working on his dissertation and as a school psychologist in Iowa. His areas of professional interest include alternative service delivery models, traditional and alternative assessment strategies, research methodology and statistics and outcome-driven educational evaluation.

Ass
broad r
plied to
to pr
unders
consid
tests d
dersta
develo
ments
stanti
these
threa

A
demo
educ
scho
ceiv
give
199
test
spe
par
dis
sp
dr
ab
co

Ac