The Wechsler Intelligence Scales and *Gf-Gc* Theory

A Contemporary Approach to Interpretation

2000

Dawn P. Flanagan
St. John's University, New York

Kevin S. McGrew
St. Cloud State University

Samuel O. Ortiz
San Diego State University

(Theory material is update of McGrew & Flanagan, 1998)

Foreword by Alan S. Kaufman

Allyn and Bacon
Boston ▪ London ▪ Toronto ▪ Sydney ▪ Tokyo ▪ Singapore

Series editor: Becky Pascal
Series editorial assistant: Susan Hutchinson
Marketing manager: Caroline Croley

A Pearson Education Company
Needham Heights, MA 02494

Internet: www.abacon.com

Portions of this book first appeared in *The Intelligence Test Desk Reference (ITDR):* Gf-Gc *Cross-Battery Assessment,* by Kevin S. McGrew and Dawn P. Flanagan, copyright © 1998 by Allyn & Bacon.

Library of Congress Cataloging-in-Publication Data

Flanagan, Dawn P.
The Wechsler intelligence scales and Gf-Gc theory : a contemporary approach to interpretation / Dawn P. Flanagan, Kevin S. McGrew, and Samuel O. Ortiz.
p. cm.
Includes bibliographical references and index.
ISBN 0-205-29271-2
1. Wechsler Adult Intelligence Scale. 2. Intelligence tests. 3. Intellect. I. McGrew, Kevin S. II. Ortiz, Samuel O. III. Title.
BF432.5.W4F53 1999
153.9'32--dc21
99-35058
CIP

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1 03 02 01 00 99

Psychology Archives University of Akron

In Memory of David Wechsler
(1896–1981)

Every step of progress the world has made has been
from scaffold to scaffold, and from stake to stake
—Wendell Phillips, 1851

This book is dedicated to those scholars whose work has provided us with a scaffold on which to build

David Wechsler
For developing a rich clinical instrument that has provided important insights into the theoretical construct of intelligence and the cognitive capabilities of individuals

Raymond B. Cattell and John L. Horn
For developing the Gf-Gc *theory of intelligence, a theory that demonstrates that an individual's intelligence is not a single ability, but rather a mosaic of many distinct cognitive abilities*

John B. Carroll
For articulating a comprehensive, empirically supported cognitive ability taxonomic foundation that can be used to bridge intelligence test theory and practice

Richard W. Woodcock
For modeling how to bridge the Gf-Gc *theory/measurement gap with* Gf-Gc *designed test batteries and for first recognizing that "crossing" different batteries may be necessary to measure a great breadth of a person's* Gf-Gc *cognitive abilities*

Samuel Messick
For reminding psychologists that all psychological measures are not created equal and that all psychological tests must be based on strong theory-based construct validity evidence

Alan S. Kaufman
For providing the "intelligent testing" interpretive framework—a framework that captures the delicate balance between the art and science of intelligence test interpretation

CONTENTS

FOREWORD

ALAN S. KAUFMAN
Clinical Professor of Psychology
Yale University School of Medicine

The history of intelligence testing has often been chronicled—from Galton to Binet to Wechsler—and that type of historical perspective can be useful for understanding contemporary test practices and controversies. Sometimes, however, it is even more prudent to stand back and ponder the history of the *interpretation* of intelligence tests. That history is less often discussed, but has its own place in making sense out of the vitriolic bile that seems to follow IQ tests like a shadow. This book by Flanagan, McGrew, and Ortiz has stirred my own memory bank regarding the history of IQ test interpretation—in part because of their summary of Kamphaus, Petoskey, and Morgan's (1997) division of the history of IQ test interpretation into four *waves,* and in part because the comprehensive theory-based approach to Wechsler interpretation that forms the crux of the Flanagan, McGrew, Ortiz text reflects the crest of the fourth wave, "applying theory to intelligence test interpretation."

My own work, primarily *Intelligent Testing with the WISC-R* (Kaufman, 1979), was credited with forming an aspect of the third wave (psychometric profile analysis), in conjunction with the pioneering factor-analytic studies published by the recently deceased Jacob Cohen (i.e., Cohen, 1959), and paving the way for the fourth wave of theoretical applications. Yet, I know that other people were far more influential than I was in forming the bridge between the third and fourth waves, and that a serendipitous set of circumstances tilled the soil for the interpretive method that would be attributed to me. Let me explain.

First, there was the burgeoning learning disabilities movement that began to gain steam during the decade of the 1960s. At the end of the decade of the 1950s, the old Stanford-Binet was still the king of assessment, with the Wechsler scales merely the pretenders to the throne. But spokespersons for the learning disabilities movement such as Sam Kirk, more likely to reside in Departments of Special Education than Educational Psychology, made it clear that multiscore IQ tests were essential for proper diagnosis and remediation and that the one-score Binet was as outmoded as the Model T Ford. Waiting in the wings was the 1949 WISC, until then a bridesmaid, ready to take over the field. In a way, the Wechsler takeover of the Binet was similar to Binet's ultimate triumph over Galton's psychophysical measure of intelligence. Binet had been developing tasks in his laboratory and gathering data on children's abilities since the mid-1880s—Galton's time as reigning King. When the French government knocked on Binet's door in the early 1900s to weed out the slow thinkers from the crème-de-la-crème in Paris's school system, Binet delivered his 1905 test in record time; he had it ready and waiting, and just needed to dust off a few parts. Similarly, when the Binet scale was found wanting by the learning disabilities leaders in the United States more than a half-century later, the three-IQ and multiple-scaled-score WISC, waiting a decade for its opportunity, stepped to the forefront of the testing scene.

Arm in arm with Cohen's factor loadings (fine-tuned with "g" loadings and estimates of subtest specificity), and given renewed life by new interpretive heroes like Alex Bannatyne, the WISC began its reign.

During the early 1970s, two fortuitous occurrences happened to me. In 1970, I began working with master clinician and genius David Wechsler, the same year I earned my Ph.D. under master psychometrician and genius Robert L. Thorndike, at Columbia University. In my clinical courses on IQ interpretation, I was taught the clinical profile analysis that the second wave featured, as promoted by Rapaport and others, that somehow was only able to interpret low scores on Information in terms of the Oedipal conflict, and the inability to reverse digits in terms of hostility and a weak super-ego. In my psychometric theory courses, Thorndike's bent dominated, and the Cohen statistics assumed primacy. The latter influence was the strongest in my own development, constantly reinforced by my superiors at The Psychological Corporation—most of whom, I found out, were also trained by Robert L. Thorndike (although one went back to his father, E. L. Thorndike). Despite Dr. Wechsler's frequent attempts to get his dull-witted student (me) to see things through an astute clinician's eye (his), I always fell back on the psychometric approach. When I wrote an article on the factor analysis of the WISC-R (Kaufman, 1975), one that continued the work of Cohen and was to be frequently cited, I was a practicing and devout coward; I didn't even have the courage to change the name of Cohen's Freedom from Distractibility factor, even though my instincts told me that it measured some cognitive ability or other, and I could barely spell the damn thing.

All the while that Wechsler was trying to hammer some sense into my head, so was my wife and colleague, Nadeen, who was also in a doctoral program at Columbia, but one that didn't rely on psychologists to interpret IQ tests and whose clinical and diagnostic coursework barely overlapped with the coursework in Psychology programs. Nadeen was in the Learning Disabilities and Neuroscience division of the Department of Special Education, and she learned IQ tests not by begging her neighbor to volunteer their children for practice testing, but by testing individuals of all ages in the Learning Disabilities clinic under the watchful eye of some superb clinicians. She was being taught to focus on intraindividual differences, to apply theory to the test-score profile, and to integrate behavioral observations with the profile of scores. Her mentors, Dr. Margaret Jo Shepherd and the late Dr. Jeannette Fleischner, were using different words than my mentor, Dr. Wechsler, but their messages had a similar ring: Focus on the client, not the profile of test scores; understand this profile in terms of theory and clinical observations of behavior, background information, and reasons for referral; and use the IQ test to understand the person's strengths and weaknesses so that the test results can be used productively. However, I listened with half a brain, convinced that the real answers could be found by sticking to psychometric formulas.

Then I left my ivory tower and took a position in the School Psychology program at the University of Georgia, taught the IQ course, and had to grade case reports. Help! I conferred with Nadeen. I called Dr. Wechsler. This time I listened with both hemispheres, integrated what they taught me with my own strait-laced psychometric background, and developed the method that ultimately was seen as transitioning from wave three to wave four in the Kamphaus and colleagues (1997) scheme of interpretive history.

And now Flanagan, McGrew, and Ortiz have taken my pleas for an integrated research-based and theoretical approach to IQ test interpretation to a new level. In my writings, I asked for research results to be applied to profile interpretation. This book, *The Wechsler Intelligence Scales and* Gf-Gc *Theory: A Contemporary Approach to Interpretation,* is based on an impressive compilation and integration of research investigations. Every chapter has research at its foundation. I asked for theory to be applied to profile interpretation. Flanagan, McGrew, and Ortiz have achieved more than anyone else in operationalizing my plea into action. They have accomplished their ambitious goal of applying Carroll's (1993a) research-based, comprehensive theory to the complex task of interpreting Wechsler's intelligence and memory tests, and have done it in a way that translates to a veritable guide for examiners to follow. One of the basic tenets of my approach to IQ test interpretation is to supplement Wechsler's scales with pertinent tasks to round out the assessment and to follow-up hunches and hypotheses. This psychoeducational approach to assessment (courtesy of patient, hands-on teaching by Nadeen, and vicarious mentorship from Dr. Shepherd and Dr. Fleischner) has been implemented to near perfection by Flanagan and colleagues via their numerous valuable tables that systematically categorize tasks from the diversity of other IQ tests that are available to clinicians. The authors of this text have also systematically applied research concerning the inclusion of basic concepts in the directions to children, the need to understand item difficulty gradients, and other subtle aspects of Wechsler's tasks, in their development of helpful interpretive sheets for each component subtest.

The 1990s have witnessed two major sophisticated, high-quality psychometric approaches to intelligence test interpretation: The research conducted by Glutting, McDermott, and their colleagues on profile interpretation and the research by Flanagan, McGrew, and their colleagues on the cross-battery technique. Both sets of research programs have been based, either directly or indirectly, on the now controversial approach to profile interpretation espoused by me, Randy Kamphaus, Jack Naglieri, Alex Bannatyne, and others, and tracing its roots both to Wechsler's clinical use of test scores and Kirk's psycholinguistic use of test scores. Both groups of researchers have contributed significantly to the test-interpretation scene by advancing the application of psychometrics to profile analysis.

Yet Glutting and McDermott have used the results of their research as an obstacle for clinicians, as purveyors of gloom-and-doom for anyone foolish enough to engage in profile interpretation. In contrast, Flanagan and McGrew have applied their research findings to elevate profile interpretation to a higher level, to add theory to psychometrics and thereby to improve the quality of the psychometric assessment of intelligence.

In a footnote to a reference to a Glutting and McDermott study, Anastasi and Urbina (1997) state: "One problem with several of the negative reviews of Kaufman's approach is that they seem to assume that clinicians will use it to make decisions based solely on the magnitude of scores and score differences. While it is true that the mechanical application of profile analysis techniques can be very misleading, this assumption is quite contrary to what Kaufman recommends, as well as to the principles of sound assessment practice" (p. 513).

One thing is obvious to me. Flanagan, McGrew, and Ortiz have internalized sound assessment principles. And they might even understand my method of profile interpretation better than I do.

A. S. K.

PREFACE

This book has one overarching goal—to modernize the interpretation of the Wechsler Intelligence Scales by applying *Gf-Gc* theory and the cross-battery approach to intellectual assessment and interpretation. This book represents a focused extension of the *Intelligence Test Desk Reference (ITDR):* Gf-Gc *Cross-Battery Assessment* (McGrew & Flanagan, 1998), in which the cross-battery approach was first introduced. The *Gf-Gc cross-battery approach* is a time-efficient method of intellectual assessment that allows practitioners to measure validly a wider range (or a more in-depth but selective range) of cognitive abilities than that represented by any one intelligence battery in a manner consistent with contemporary psychometric theory and research. Whereas the ITDR briefly described how to use the cross-battery approach to supplement the major intelligence batteries (including the Wechsler scales), this book provides an *in-depth treatment* of how to use cross-battery principles and techniques to augment the Wechsler Intelligence Scales (WPPSI-R, WISC-III, WAIS-III) in a psychometrically defensible manner and interpret the results of Wechsler-based cross-battery assessments within the context of current theory and research.

In the process of writing this book and applying the *Gf-Gc* cross-battery approach to the Wechsler Intelligence Scales, we have gained a greater appreciation of the foundational sources of information on which our assessment approach is based. In particular, this book builds on the seminal work of David Wechsler, Raymond Cattell, John Horn, John Carroll, Richard Woodcock, Samuel Messick, and Alan Kaufman. Through their extensive research, ideas, and writings, these scholars have contributed significantly to our development of the Wechsler-based *Gf-Gc* cross-battery approach in several important ways.

First, few psychologists would argue the fact that David Wechsler's Intelligence Scales currently dominate the practice of intelligence testing. Although we have been critical of certain aspects of the Wechsler Scales in our writings, there is little doubt in the pages that follow that our work has been influenced directly and indirectly by the writings of David Wechsler.

According to one who knew him well, Wechsler succeeded largely because he was able to anticipate the needs of practitioners and had the courage to challenge the prevailing Stanford-Binet monopoly (Kaufman, 1990a). Similarly, it is our hope that the Wechsler *Gf-Gc* cross-battery approach presented in this book will meet the emerging needs of assessment professionals who have recognized the gap between intelligence theory and practice. However, because our ideas and procedures necessitate a change in thinking and practice with regard to the Wechsler Intelligence Scales, our recommended use and interpretation of these batteries may be met with resistance by some, this time due to the prevailing Wechsler monopoly, which carries with it a limited test-kit focus. Our approach suggests a shift in focus from a circumscribed set of measures (as represented by a single intelligence battery) to a theory-driven method of organizing assessments and making interpretations. Thus, our Wechsler-based *Gf-Gc* cross-battery approach is not intended to denigrate the Wechsler Scales, but rather, to modernize these measures thereby extending Wechsler's legacy.

Second, the seminal theoretical work of Raymond Cattell, John Horn, and John Carroll played a prominent role in the development of our assessment approach. Cattell, Horn, and Carroll's contributions have, in our judgement, provided a convincing argument that a hierarchical multiple-ability theory, such as those represented by the Horn-Cattell *Gf-Gc* and Carroll *Three-Stratum* models, best describes the structure of human intelligence. Furthermore, the network of validity evidence (e.g., substantive, structural, and external) that supports the *Gf-Gc* structure of intelligence argues strongly for the use of this framework as a guide to the selection and interpretation of all intelligence batteries.

Third, readers familiar with the *ITDR* will recognize the "three pillars of cross-battery assessment" presented in Chapter 6 of this book. Briefly, the *Gf-Gc* cross-battery approach is predicated on three major sources of information: (1) the *Gf-Gc* theory of intelligence, (2) cross-battery factor-analysis-based classifications of the individual tests in all major intelligence batteries at the *broad* (stratum II) level of the *Gf-Gc* model, and (3) expert consensus-based classifications of individual tests at the *narrow* (stratum I) level of the *Gf-Gc* model. All three pillars provide evidence from which valid inferences can be drawn from cross-battery organized test scores. The second and third pillars focus on increasing validity through the reduction of *construct irrelevant variance* and *construct underrepresentation,* respectively—two ubiquitous sources of invalidity in traditional assessment and interpretation approaches. The three cross-battery pillars (the latter two in particular) are best conceptualized as being part of a larger overarching theory-based construct validity framework that rests solidly on the work of Samuel Messick. The writings of Messick have allowed us to more accurately place the Wechsler-based *Gf-Gc* cross-battery approach within a "big picture" construct validity structure.

Fourth, the application of the cross-battery approach to the Wechsler Intelligence Scales is consistent with the influential writings of Alan Kaufman. Kaufman's prominent "intelligent" approach to Wechsler intelligence test interpretation is at the core of our teaching, writing, research, and practice. To be sure, Kaufman's approach to intelligence test interpretation—an approach that recognizes that "clinical assessment is part art, part science" (Kaufman, 1994, p. 27)—permeates much of the Wechsler-based *Gf-Gc* cross-battery approach presented in this text.

In summary, the process of extending the *Gf-Gc* cross-battery approach to the interpretation of the Wechsler Intelligence Scales has made us more cognizant of the shoulders on which we stand—Wechsler, Cattell, Horn, Carroll, Woodcock, Messick, and Kaufman. We hope our humble efforts to integrate their contributions and extend them to the use and interpretation of the Wechsler Intelligence Scales do justice to their work.

Organization

This book is organized in three sections. Part I (Linking Contemporary Intelligence Theory and Practice: An Overview) consists of three chapters. In Chapter 1, the Wechsler Intelligence Scales are placed in historical and contemporary perspective. This chapter is arranged around Kamphaus's conceptualization of intelligence test interpretation as representing four waves, beginning in the late 1900s through present day. The Wechsler Scales have been an integral part of each wave of test interpretation. This book is an attempt to ground the Wechsler Scales firmly in the fourth, *theory-based wave* of intelligence test interpretation. Chapter 2 provides an overview of the current state-of-the-art of intellectual assessment, and describes the progress that has been made in both psychometric theory development and intelligence test development. An integrated Cattell-Horn-Carroll *Gf-Gc* model is presented and defined in this chapter. Support for the *Gf-Gc* framework as the most well validated and researched theoretical model of multiple cognitive abilities within the psychometric tradition is presented in this chapter. Also, the Wechsler Scales are described according to the extent to which they operationalize prominent abilities specified in the *Gf-Gc* structure of intelligence. In Chapter 3, the extant literature on the validity of the Wechsler Intelligence Scales is evaluated according to substantive, structural, and external validation criteria. This chapter shows how *Gf-Gc* theory can be linked to the applied measurement of cognitive abilities using the Wechsler Scales. Specifically, we impose a strong substantive framework to the interpretation of the Wechsler Intelligence Scales via McGrew and Flanagan's (1998) *Gf-Gc* cross-battery approach. The end result is the derivation of more valid inferences from Wechsler test scores. Together, Chapters 1 through 3 of this book provide the foundational knowledge on which our approach to using and interpreting the Wechsler Intelligence Scales was based.

Part II of this text (Descriptions and Evaluations of the Wechsler Intelligence and Wechsler-Linked Memory Scales) provides a comprehensive review of the psychometric, theoretical, and qualitative characteristics of the individual subtests of the WPPSI-R, WISC-III, WAIS-III, WMS-III, and CMS. Chapter 4 describes these test characteristics in detail and relates each characteristic to the test interpretation process. In addition, the psychometric (e.g., reliability, *g* loading, floors/ceilings), theoretical (e.g., *Gf-Gc* broad and narrow ability classifications), and qualitative (e.g., individual/situational factors that influence performance, degree of cultural loading, degree of linguistic demand) characteristics for each individual test of the Wechsler Scales are presented on summary pages (one page per test) in an easy-to-read, visual-graphic format at the end of Chapter 4. Chapter 5 provides a brief description of the importance of *supplemental* cognitive ability tests in the assessment and interpretation process. All cognitive ability tests included in this chapter are described according to the *Gf-Gc* theoretical model and are used in subsequent chapters to supplement the Wechsler Intelligence Scales. Finally, like Chapter 4, Chapter 5 provides test characteristic summary pages for the Wechsler-linked memory batteries (i.e., WMS-III and CMS).

Part III describes the product of grounding cognitive ability assessment and interpretation with the Wechsler Scales in strong theory and research—the *Wechsler-based* Gf-Gc *cross-battery approach*. The foundation, rationale, and application of this approach are presented in Chapter 6. In this chapter, we argue strongly for a theory- and research-based approach and highlight the utility of this approach in uncovering intracognitive strengths and weaknesses particularly as it applies to the identification and diagnosis of learning disabilities. Chapter 7 provides a comprehensive approach to interpreting *Gf-Gc* cross-battery data using a case example. Finally, Chapter 8 extends the Wechsler-based *Gf-Gc* cross-battery approach to multicultural and multilingual populations. Numerous tables, figures, flowcharts, and *Gf-Gc* cross-battery worksheets are provided throughout this book to assist the reader in the process of infusing this material in their current practice. In addition, the appendices provide valuable information, such as a *Gf-Gc* cross-battery interpretive report, a "user-friendly" guide to understanding *Gf-Gc* abilities, a percentile and standard score conversion table, and information about ability-achievement discrepancy analyses, that is also intended to assist Wechsler users in the application of the psychometrically and theoretically defensible cross-battery approach.

Intended Audience

This book is intended for practitioners, researchers, and scholars who seek to infuse current theory and research in their use and interpretation of the Wechsler Intelligence Scales. Practitioners, university trainers, students, researchers and other professionals in school, clinical, counseling, and educational psychology as well as neuropsychology, who use the Wechsler Intelligence and Memory Scales in applied settings would find this book valuable. This book would be appropriate for a graduate course in beginning or advanced intelligence testing, measurement, and psychoeducational assessment. This book is also particularly valuable for those who seek an organized, systematic, and theory-based method for evaluating cognitive functioning in children, adolescents, and adults, including those from culturally and linguistically diverse backgrounds.

Acknowledgments

We are grateful to a number of individuals who facilitated the preparation of this book. Our deepest appreciation is extended to Jennifer Mascolo who spent endless hours preparing tables, organizing materials, conducting library research, editing, and performing countless other tasks related to the preparation of this book. Jennifer Mascolo's dedication, commitment, insightfulness, energy, and attention to detail greatly enhanced the quality of this work. We also wish to thank Matthew Broudy, Bhupin Butaney, Theresa Huettl, Jennifer Jablonski, Dennis Russell, and Joe Wehrman for their assistance in locating articles, entering data, and participating in an extensive review of the literature on the validity of the Wechsler Intelligence and Memory Scales. We are particularly grateful for the diligence of Jennifer Jablonski who carefully edited earlier drafts of this manuscript.

Our appreciation also goes to the following reviewers for their valuable thoughts and comments on an initial draft of this manuscript: Vincent C. Alfonso, Fordham University; Michael E. Gerner, President, Consulting Psychologists, Flagstaff, AZ; Steven K. Kaplan, Lesley College; Cynthia A. Riccio, Texas A&M University; Alexander I. Troster, The University of Kansas Medical Center; Christopher Tolsdorf, University of Connecticut; Glenna B. Rubin, St. John's University.

Finally, the contributions of Sean Wakely, Susan Hutchinson, and the rest of the staff at Allyn and Bacon are gratefully acknowledged. Their expertise and pleasant and cooperative working style made this book an enjoyable and productive endeavor.

D. P. F.
K. S. M.
S. O. O.

CHAPTER

1 The Wechsler Scales in Perspective

Historical and Contemporary Views

No theory is good unless it permits, not rest, but the greatest work. No theory is good except on condition that one use it to go on beyond.

—André Gide (1913)

Few things in life achieve preeminent stature without some merit. Substance is, after all, the fundamental criteria against which we assess greatness in nearly every case. Within the field of applied psychological assessment, the substantive elements underlying the Wechsler Intelligence Scales have served to propel these instruments to positions of dominance and popularity unrivaled in the history of intellectual assessment. The concepts, methods, and procedures embodied in the design of the Wechsler Scales have been so influential, that they have guided the majority of development and research in the field over the last half century. Virtually every reviewer of these scales, including those who voice significant concerns about the test, acknowledge the monumental impact and position of central importance that the scales have occupied in scientific endeavors aimed at understanding the nature of human intelligence and cognitive abilities. For example, despite the critical content and tone of their review, McDermott, Fantuzzo, and Glutting (1990) assert their "deep respect for most of the Wechsler heritage" by stating that "were we to say everything we might about the Wechsler Scales and their contributions to research and practice, by far our comments would be quite positive" (p. 291).

Kamphaus (1993) has also observed that praise flows from the pages of the majority of those who have written about the Wechsler Scales. The titles of many articles in the professional literature continue to illustrate the heights to which the Wechsler Scales have been elevated; for example, "King WISC the third assumes the throne" (Kaufman, 1994b). Although such praise of the Wechsler Scales has always exceeded their criticisms, they have not been without their detractors. In fact, critics of the Wechsler Scales offer compelling arguments that outline one or more significant deficiencies in these instruments (e.g., Braden, 1995; Little, 1992; McGrew, 1994; Shaw, Swerdlik, & Laurent, 1993; Sternberg, 1993; Witt & Gresham, 1985). Nonetheless, it remains clear that when viewed from a historical perspective, the importance, influence, and contribution of David Wechsler's instruments to the science of intellectual assessment can neither be disputed nor diminished.

The purpose of this chapter is neither to pay another tribute to the Wechsler Scales nor present a thesis regarding its failings. Rather, the purpose of this chapter is to provide factual and historical information regarding the Wechsler Scales and to trace developments that have occurred in attempts to interpret and derive meaning from the Wechsler scores.

The Wechsler Scales: History and Approaches to Interpretation

Kamphaus, Petoskey, and Morgan (1997) offered an extended treatment of the historical precedents and contemporary developments regarding interpretive approaches with the Wechsler Scales. These authors describe the history of intelligence test interpretation in terms of four waves: (1) quantification of a general level; (2) clinical profile analysis; (3) psychometric profile analysis; and (4) application of theory to intelligence test interpretation. Kamphaus and colleagues' four-wave framework will be used to organize the current treatment of the development of the Wechsler Scales and, more importantly, the evolution of approaches to interpreting the Wechsler Scales.

The First Wave: Quantification of General Level

To a large extent, the widespread acceptance of the early intelligence tests (the Stanford-Binet, in particular) was grounded in the conclusion that intelligence tests offered an objective method for creating distinct groups of people differentiated on the basis of their general intelligence. According to Kamphaus and colleagues (1997), this represented the first wave of intelligence test interpretation and was driven by practical considerations related to classification of individuals into separate groups.

During this period, the focus in interpretation for most all individually administered intelligence tests was on the omnibus IQ. The dominant influence of Spearman's *g* theory of intelligence and the age-based Stanford-Binet Scale, combined with the fact that factor-analytic and psychometric methods were not available for the identification of multiple cognitive abilities, contributed to an almost exclusive focus on using a global IQ to classify individuals. In turn, a number of classification systems were proposed for organizing individuals according to their global IQ.

Some of these early classification systems used labels that corresponded to medical and legal terminology (e.g., *idiot, imbecile,* and *moron*). Although the Wechsler Scales did not contribute to the early classification efforts during most of this interpretive wave, Wechsler eventually made a significant contribution. He proposed a classification scheme that relied less on evaluative terminology (albeit, it still contained the terms *defective* and *borderline*) and more on meaningful deviations from the mean that reflected the "prevalence of certain intelligence levels in the country at that time" (Kamphaus et al., 1997, p. 35). With some refinements, interpretation of intelligence tests in the present day continue to be based on this type of classification system, as distinctions are still made between individuals who are mentally retarded, learning disabled, and gifted, for example.

It appears that Wechsler accepted the prevailing ideas regarding *g* and the definition of intelligence as a global entity along the lines already postulated by Terman, Binet, Spearman, and others (Reynolds & Kaufman, 1990) when he offered his own definition of intelligence as being "the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment." Wechsler specified that this definition "avoids singling out any ability, however esteemed (e.g., abstract reasoning), as crucial or overwhelmingly important" (Wechsler, 1939, p. 3) and implies that any one intelligence subtest is readily interchangeable with another.

The Second Wave: Clinical Profile Analysis

Kamphaus and colleagues (1997) identified the second wave in interpretation as *clinical profile analysis* and suggested that the publication of the Wechsler-Bellevue (W-B; Wechsler, 1939) was pivotal in spawning the profile approach to interpretation, an approach that sought to understand individuals beyond identification of their global intellectual ability. The relationship between the development of the Wechsler Scales and the second wave of interpretation, as well as subsequent historical and conceptual developments of the Wechsler Scales and approaches to interpretation, is summarized in Figure 1.1.

The Wechsler-Bellevue Intelligence Scale, Form I, published in 1939 (a slightly updated version, Form II, was published in 1946), represented an approach to intellectual assessment in adults that was differentiated clearly from other instruments available at that time (e.g., the Binet scales). The W-B was comprised of 11 separate subtests, including Information, Comprehension, Arithmetic, Digit Span, Similarities, Vocabulary, Picture Completion, Picture Arrangement, Block Design, Digit Symbol, and Coding. Perhaps the most notable feature introduced with the W-B that contributed to an emphasis in interpretation on more than a global IQ was the grouping of subtests into the now familiar Verbal and Performance dichotomy, an organizational structure that was based on the notion that intelligence could be expressed and measured through both verbal and nonverbal modes of communication. In attempting to clarify his use of and the distinction between the verbal and nonverbal methods for assessing intelligence, Wechsler asserted that this dichotomy:

> [D]oes not imply that these are the only abilities involved in the tests. Nor does it presume that there are different kinds of intelligence, e.g., verbal, manipulative, etc. It merely implies that these are different ways in which intelligence may manifest itself. (Wechsler, 1958, p. 64)

Another important feature pioneered in the W-B revolved around the construction and organization of subtests. At the time, the Binet Scale was ordered and administered sequentially according to developmental age, irrespective of the task. In contrast, Wechsler utilized only 11 subtests, each scored by points rather than age, and each with a sufficient range of item difficulties to encompass the entire age range of the scale.

In his writings, Wechsler often shifted between conceptualizing intelligence as a singular entity (the first wave) and conceptualizing it as a collection of primary mental abili-

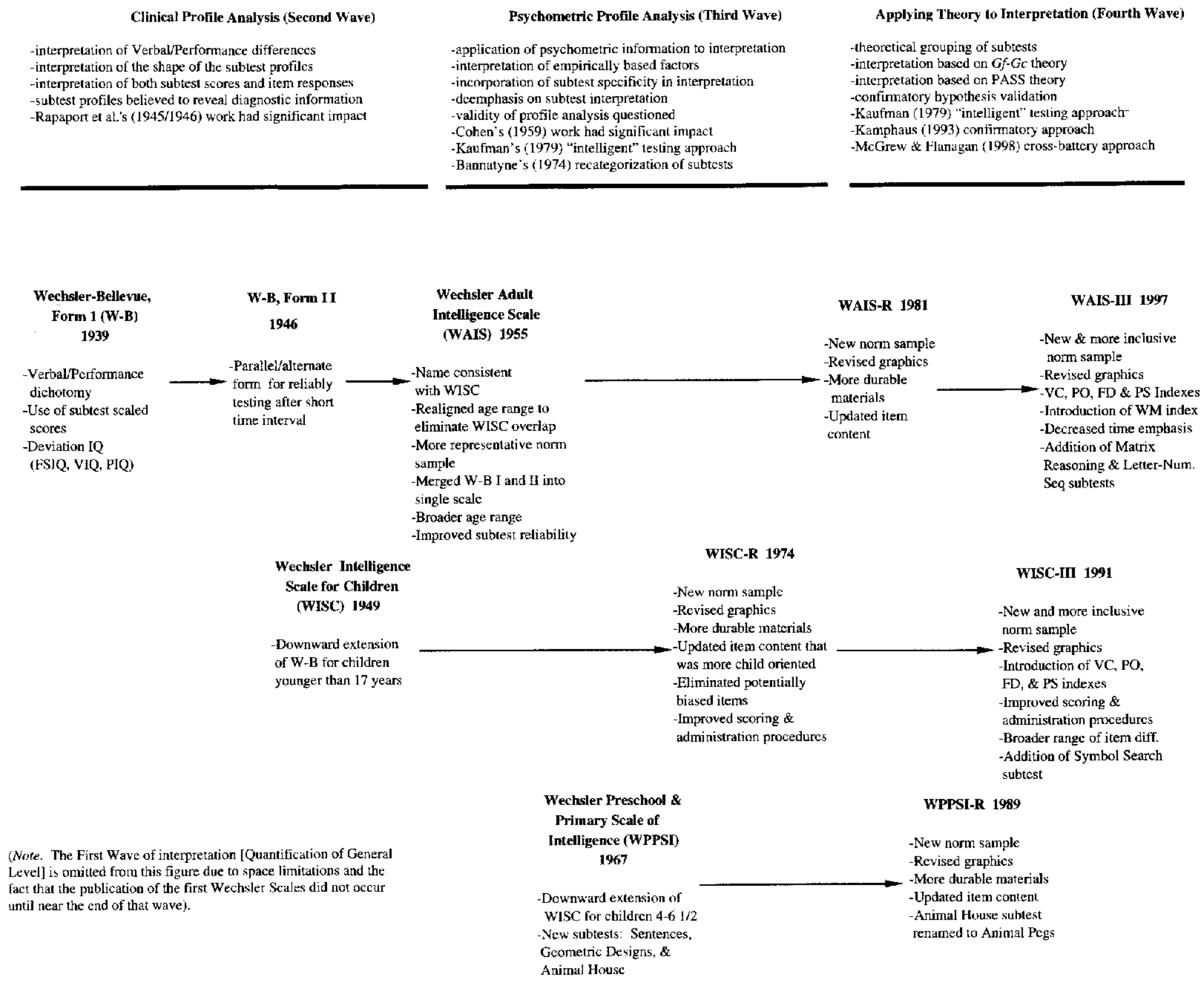

FIGURE 1.1 Timeline of Wechsler Intelligence Scale Revisions and Corresponding Interpretive Approaches

ties, a notion more consistent with the emphasis on profile interpretation during the second wave. At times he appeared to encourage the practice of individual interpretation of subtests, and suggested that each one represented a relatively distinct and different measure of intellectual ability (McDermott, Fantuzzo, & Glutting, 1990). To many, this position seems to represent a theoretical contradiction to his prior meticulous attempts not to equate general intelligence with the sum of separate intellectual abilities. This shift in viewpoint may have been responsible, in part, for the development of methods for interpreting the constructs underlying individual subtests that established the trend toward profile analysis.

Unquestionably, the structure, organization, and innovations found in the original W-B were impressive, practical, and in many respects, superior to any other instruments available in 1939. More importantly, the structure and organization of the W-B scale stimulated the pioneering efforts of Rapaport, Gill, and Schafer (1945–46) to invent approaches to test interpretation that focused on understanding the meaning behind the shape of a person's profile of subtest scores. According to Kamphaus and colleagues (1997), a new method of test interpretation developed under the assumption that "patterns of high and low subtest scores could presumably reveal diagnostic and psychotherapeutic considerations" (p. 36). Thus, during the second wave of intelligence test interpretation, the W-B (Wechsler, 1939) provided the major impetus for developing a variety of procedures for deriving diagnostic and prescriptive meaning from not only the shape of Wechsler subtest profiles, but also Verbal and Performance discrepancies and, in some cases, individual item responses.

In addition to the enormous scope of Rapaport and colleagues' (1945–46) diagnostic suggestions, their approach to understanding profile shape triggered a furious rush of investigations that sought to establish the psychological functions underlying the infinite variety of profile patterns and the nature of their relationships to each other. Perhaps as a consequence of the enormous clinical appeal of the approach espoused by Rapaport and colleagues, Wechsler (1944) helped relegate general level assessment to the back burner while increasing the heat on the analysis of profile shape.

The search for meaning in discrepancies and profiles was carried over to interpretation of the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949), a downward extension of the W-B. The WISC was comprised of the same 11 subtests used in the W-B, but was modified to assess intellectual functioning in children within the age range of 6 to 16 years. The subtests were grouped into the verbal and performance categories as before, with Information, Comprehension, Arithmetic, Digit Span, Similarities, and Vocabulary making up the verbal subtests, and Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Coding comprising the performance subtests. The WISC also provided scaled scores for each subtest and yielded the ubiquitous Wechsler composite scores: Full Scale IQ (FSIQ), Verbal IQ (VIQ), and Performance IQ (PIQ).

Although the search for diagnostic meaning in differences between Wechsler scores represented a more sophisticated approach to intelligence test interpretation, it also created additional methodological problems. With enough practice, just about any astute clinician could provide a rational interpretation of an obtained profile to fit the known functional or dysfunctional patterns of any individual. Notwithstanding, simple analysis of profile shape or scatter did not create diagnostic or treatment utility automatically. Although the next wave in intelligence test interpretation sought to address such methodological flaws with the clinical profile analysis method, this dominant interpretive approach remains in practice today (Kamphaus et al., 1997).

The Third Wave: Psychometric Profile Analysis

As presented in Figure 1.1, the original W-B scales were revised and updated into a single instrument in 1955. The name was aligned with the existing juvenile version (i.e., WISC) and became known as the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955). Major changes and revisions included incorporating Forms I and II of the W-B into a single scale with a broader range of item difficulties, realigning the target age range to include ages 16 years and older (which eliminated overlap with the WISC, creating a larger and more representative norm sample) and refining the subtests to improve reliability.

Within this general time period, technological developments in the form of computers and readily accessible statistical software packages to assist in intelligence test interpretation, provided the impetus for what Kamphaus and colleagues (1997) called the third wave of interpretation—*psychometric profile analysis*. The work of Cohen (1959), which was based extensively on the then new WAIS (Wechsler, 1955), sharply criticized the clinical profile analysis tradition that defined the second wave. For example, Cohen's factor-analytic procedures revealed a viable three-factor solution which rivaled the dichotomous Verbal-Performance model and remained the de facto standard for the Wechsler Scales for decades. Also, by examining and removing the variance shared between subtests, Cohen demonstrated that the majority of Wechsler subtests had very poor specificity (i.e., reliable, specific variance). Thus, the frequent clinical practice of interpreting subtests as reliable measures of a *presumed* construct was not supported. Kamphaus and colleagues (1997) summarize Cohen's significant contributions that largely defined the third wave of test interpretation as threefold: (1) empirical support for the Full Scale IQ based on analysis of shared variance between subtests; (2) development of the three-factor solution for interpretation of the Wechsler Scales; and (3) revelation of limited subtest specificity questioning individual subtest interpretation.

The most vigorous and elegant application of psychometric profile analysis to intelligence test interpretation occurred with the revision of the venerable WISC (Wechsler Intelligence Scale for Children—Revised; Wechsler, 1974). Briefly, as summarized in Figure 1.1, the WISC-R utilized a larger, more representative norm sample than its predecessor, included more contemporary-looking graphics and updated items, eliminated content that was differentially familiar to specific groups, and included improved scoring and administration procedures. Armed with the WISC-R, Kaufman (1979) articulated the essence of the psychometric profile approach to intelligence test interpretation in his seminal book, *Intelligent Testing with the WISC-R* (now superceded by *Intelligent Testing with the WISC-III;* Kaufman, 1994).

Kaufman emphasized flexibility in interpretation and provided a logical and systematic approach that utilized principles grounded in measurement theory. Reflective of the underlying philosophy of the psychometric profile analysis wave, Kaufman's approach required the examiner to have a greater level of psychometric expertise than might ordinarily be possessed by the average clinician. Anastasi (1988) lauded and recognized that "the basic approach described by Kaufman undoubtedly represents a major contribution to the clinical use of intelligence tests. Nevertheless, it should be recognized that its implementation requires a sophisticated clinician who is well informed in several fields of psychology" (p. 484). In some respects, publication of Kaufman's work can be viewed as an indictment against the poorly reasoned and unsubstantiated interpretation of the Wechsler Scales that had sprung up in the second wave (clinical profile analysis). Kaufman's focal message was the notion that interpretation of Wechsler intelligence test performance must be conducted with a higher than usual degree of psychometric precision and must be based on credible and dependable evidence, rather than merely the clinical lore which surrounded earlier interpretive methods. The psychometric profile analysis approach was also applied readily to Wechsler's downward extension of the WISC (i.e., the Wechsler Preschool and Primary Scale of Intelligence; WPPSI; Wechsler, 1967).

Despite the enormous body of literature that has mounted over the years regarding profile analysis of the Wechsler Scales, this form of interpretation, even when upgraded with the rigor of psychometrics, must be regarded as a perilous endeavor, because it is largely without empirical support, and it is not grounded in a well-validated theory of intelligence. With over 75 different profile types discussed in a variety of areas, including neuropsychology, personality, learning disabilities, and juvenile delinquency (McDermott, Fantuzzo, & Glutting, 1990), there is considerable temptation to believe that such analysis is reliable. It must be remembered, however, that many studies (e.g., Hale, 1979; Hale & Landino, 1981; Hale & Saxe, 1983) have demonstrated consistently that "profile and scatter analysis is not defensible" (Glutting, McDermott, Watkins, Kush, & Konold, 1997; Kavale & Forness, 1984, p. 136). In a meta-analysis of 119 studies of the WISC-R subtest data, Mueller, Dennis, and Short (1986) concluded that using profile analysis with the WISC-R in an attempt to differentiate various diagnostic groups is clearly not supported. Recent evaluations regarding the merits of profile analysis have produced similar results (e.g., Glutting, McDermott, & Konold, 1997; Glutting et al., 1997b; Kamphaus, 1993; McDermott, Fantuzzo, Glutting, Watkins, & Baggaley, 1992; Watkins & Kush, 1994). The nature of the controversy surrounding clinical profile analysis, with or without the application of psychometric theory, was brought to the forefront by McDermott and colleagues (1990) in their substantive discussion of the subject. After extensively reviewing the subtest analysis literature and investigating the diagnostic utility of Wechsler subtest scatter based on their own sound analyses of nationally representative datasets, McDermott and colleagues concluded, "until preponderant and convincing evidence shows otherwise, we are compelled to advise that psychologists 'just say no' to subtest analysis" (p. 299).

The Fourth Wave: Application of Theory

The third wave's less-than-impressive results at improving intelligence test interpretation set the stage for the fourth and current wave, described by Kamphaus and colleagues (1997) as *application of theory.* The need to integrate theory and research in the intelligence test interpretation process was articulated best by Kaufman (1979). Specifically, Kaufman commented that problems with intelligence test interpretation can be attributed largely to the lack of a specific theoretical base to guide the practice. He suggested that it was possible to enhance interpretation significantly by reorganizing subtests into clusters specified by a particular theory. In essence, the end of the third wave of intelligence test interpretation and beginning of the fourth wave was marked by Kaufman's pleas for practitioners to ground their interpretations in theory, as well as his efforts to demonstrate the importance of linking intellectual measurement tools to empirically supported and well-established conceptualizations of human cognitive abilities.

In contrast to the Wechsler Scales' central role in the development of new approaches to test interpretation (e.g., clinical and psychometric profile analysis), recent revisions of the Wechsler trilogy (i.e., WPPSI-R, WISC-III, WAIS-R/WAIS-III) have (unfortunately) failed to ride the next wave of test interpretation (i.e., the fourth "theoretical" wave). As seen in Figure 1.1, since the early 1980s the WPPSI and WISC-R have undergone one revision each (WPPSI-R; WISC-III) and the WAIS has undergone two revisions (WAIS-R; WAIS-III). However, neither instrument changed substantially from its predecessor. Changes to the basic structure, item content, and organization of the WPPSI-R and WISC-III were relatively minimal, with the most obvious changes being cosmetic. However, the WISC-III introduced four new composite score indexes, Verbal Comprehension (VC), Perceptual Organization (PO), Freedom from Distractibility (FD), and Processing Speed (PS), to supplement the subtest scaled scores and the FSIQ, VIQ, and PIQ. This version of the WISC also contains one new subtest, Symbol Search.

In terms of structure, organization, and content, the WAIS-R did not represent a significant departure from the WAIS. As summarized in Figure 1.1, the more salient changes reflected in the WAIS-R included a new norm sample, revised item graphics, more durable materials, and updated item content. The WAIS-III, however, reflected more substantive revisions, including more careful attention to the factor structure and statistical linkage to other measures of cognitive functioning and achievement. Of course, the WAIS-III also included the more typical changes, such as the use of updated color graphics with improved item content, and the adoption of the composite indexes first introduced with the WISC-III. The WAIS-III, like the WISC-III, yields a VC, PO, and PS index; however, the FD index was renamed the *Working Memory (WM) Index* in the WAIS-III. This change most likely reflected the considerable controversy over whether FD was a viable construct, as well as the increasing recognition of the importance of working memory in understanding (predicting) specific academic skills. Adding a working memory construct to the underlying factor structure of the WAIS-III led to the development of a working memory test (i.e., Letter-Number Sequencing) and a slight reorganization of subtests, based on the results of factor analyses (Wechsler, 1997). The organization of subtests according to the respective underlying factor structures of the WPPSI-R, WISC-III, and WAIS-III is presented in Table 1.1.

Although the latest versions of the WISC-III and WAIS-III provide more factor-based composite scores for interpretation than their predecessors (i.e., VC, PO, FD, PS, and WM), the fact remains that nearly all current options for interpreting Wechsler test performance are not grounded in any contemporary theoretical model of intelligence. This failure to ground the latest revisions of the Wechsler Scales in a contemporary theoretical model is at variance with Kaufman's (1979) admonition for intelligence tests and intelligence test interpretation to become more theory-based.

The fact that the Wechsler Intelligence Scales lack theoretical substance cannot, in his absence, be attributed to David Wechsler. Admittedly, David Wechsler was not often considered a theoretician in his own right and the name "*Wechsler*" is rarely followed by the word *theory* in the professional literature. Discussions of his work invariably refer to

TABLE 1.1 Factor Indexes and Organization of Subtests in the Current Wechsler Intelligence Scales

TEST	FACTOR	SUBTEST
WPPSI-R	Verbal Comprehension (VC)	Information Similarities Vocabulary Comprehension Arithmetic Sentences
	Perceptual Organization (PO)	Picture Completion Block Design Mazes Animal Pegs Geometric Designs
WISC-III	Verbal Comprehension (VC)	Information Similarities Vocabulary Comprehension
	Perceptual Organization (PO)	Picture Completion Picture Arrangement Block Design Object Assembly
	Freedom from Distractibility (FD)	Arithmetic Digit Span
	Processing Speed (PS)	Coding Symbol Search
WAIS-III	Verbal Comprehension (VC)	Information Similarities Vocabulary Comprehension
	Perceptual Organization (PO)	Picture Completion Block Design Matrix Reasoning
	Working Memory (WM)	Arithmetic Digit Span Letter-Number Sequencing
	Processing Speed (PS)	Digit Symbol-Coding Symbol Search

Wechsler's *views*, or Wechsler's *definition*, and not Wechsler's *theory*. However, if he were alive today, the possibility exists that the Wechsler Scales would have kept abreast with the theoretical focus of contemporary intelligence test development and, may have even led these efforts.

Although Wechsler's name is still listed as the author of all revisions since his death in 1981, this most likely reflects contractual obligations. The publisher of the Wechsler Scales, as well as the members of the work groups organized by the publisher, are more accurately responsible for the slow, incremental changes in the Wechsler Scales. Historical continuity and tradition apparently has played a stronger role than theoretical considerations in the revisions of the Wechsler Scales.

In addition to a greater focus on the development and revision of intelligence tests based on a theoretical framework during the fourth wave, Kamphaus (1993, 1998) extended the portion of Kaufman's intelligent testing approach that espouses a method for integrating theory and hypothesis validation in the test interpretation process. Briefly, Kamphaus warned against the practice of using results from intelligence tests in isolation, without the benefit of other supporting data. In addition, he emphasized the need to base interpretations on research evidence and theory. Central to this approach is the *a priori* specification of hypotheses relevant to the referral questions. The development of such hypotheses changes the nature of the assessment and interpretive process from exploratory to confirmatory. Typical clinically based profile analyses generally require the clinician to gather a wide variety of evidence and, subsequently, engage in attempts to make sense of it *a posteriori*. Conversely, Kamphaus suggested that a hypothesis should arise from already existing evidence and then be tested specifically (see case study presentation in Chapter 7). A crude analogy of this distinction may be expressed as the difference between a leisurely fishing trip and a hunting expedition—catching anything that bites versus setting out for specific game.

The primary purpose of this book is to provide a method for guiding interpretation of the Wechsler Scales from an underlying modern theoretical foundation. The theory-based interpretive approach outlined in the subsequent chapters represents an attempt to move the Wechsler Scales into the currents of the fourth wave of intelligence test interpretation (i.e., application of theory). Figure 1.2 illustrates the various theoretical, empirical, and interpretive components that are combined and integrated to create the basic principles of the proposed assessment and interpretive approach.

The approach described herein is an outgrowth of the publication of the *Intelligence Test Desk Reference (ITDR):* Gf-Gc *Cross-Battery Assessment* (McGrew & Flanagan, 1998), which introduced a comprehensive application of theory-driven assessment and interpretation known as the *cross-battery approach.* One of the core components of the cross-battery approach rests on the adoption and application of an empirically supported, modern theory of intelligence. The two upper boxes on the left side of Figure 1.2 illustrate the development of an integrated Carroll (1989; 1993a) and Horn-Cattell (Horn, 1985; 1988; 1989; 1991) *Gf-Gc* theoretical model that is supported with considerable empirical research that establishes construct validity. The arrows leading from these boxes indicate their integration as the foundation for the *Gf-Gc* based classification of intelligence batteries and supplemental tests according to the broad (Stratum II) and narrow (Stratum I) abilities they measure. In other words, theory specifies the structure and empiricism supports

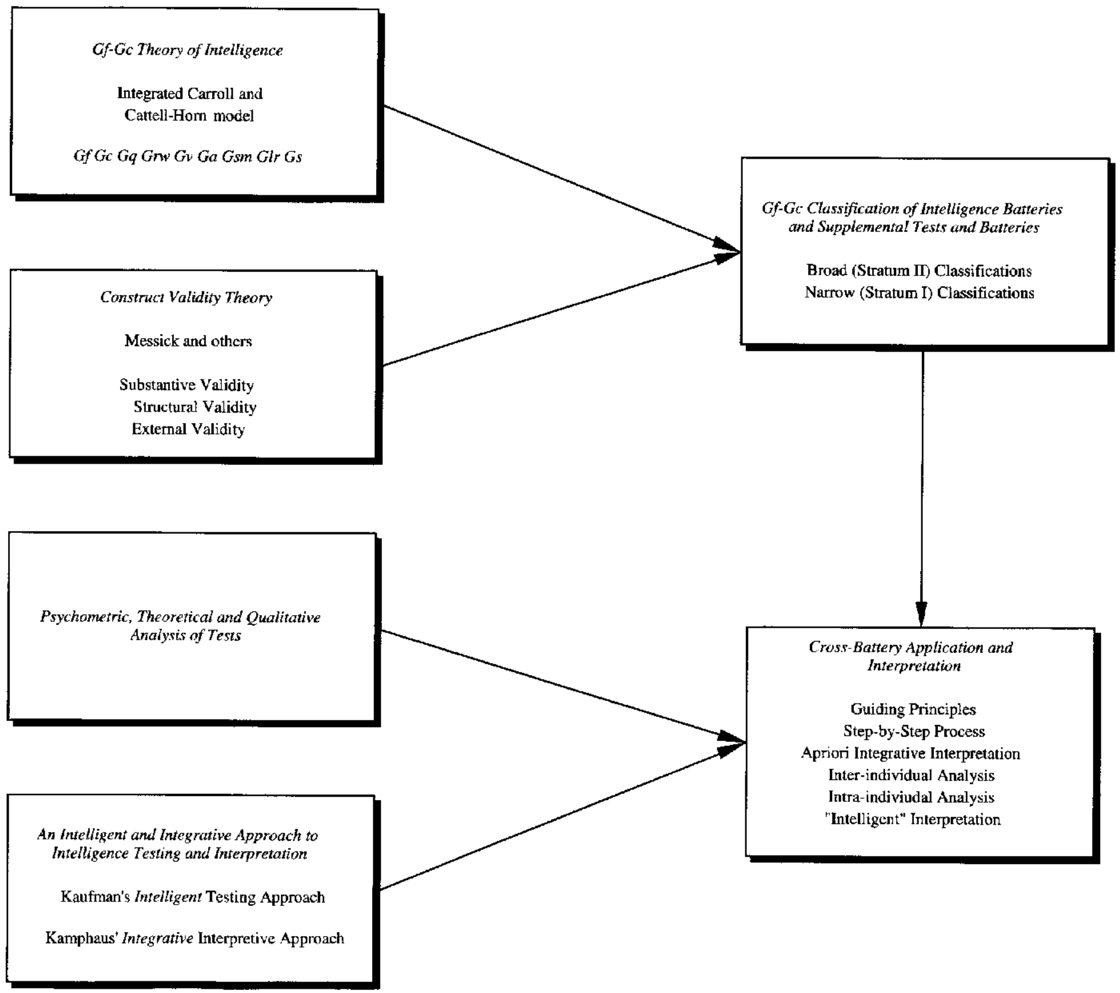

FIGURE 1.2 Major Developmental and Structural Components of *Gf-Gc* Cross-Battery Application and Interpretation

it. The two lower boxes on the left side of Figure 1.2 illustrate the integration of knowledge related to the psychometric, theoretical, and qualitative properties of tests with a rational and sound approach to assessment and interpretation (e.g., Kaufman's intelligent testing approach). The arrows leading from these boxes indicate their integration into a comprehensive, systematic, hypothesis-driven approach to cross-battery based assessment and interpretation, which also rests on the knowledge base of the broad and narrow *Gf-Gc* classifications described in the previous text.

Each of these major developmental and structural components are presented and discussed in detail in the chapters that follow, including the manner in which they may be applied in the assessment of culturally and linguistically diverse populations (see Chapter 8). As such, the remaining chapters of this book offer a formal Wechsler-based *Gf-Gc* cross-battery approach that merges intelligent testing with intelligent interpretation. When fully integrated, we believe that the resulting cross-battery approach represents a pioneering effort in line with the fourth wave of intelligence test interpretation.

Conclusion

The contributions to the science of intellectual assessment made by David Wechsler through his intelligence scales are many and substantial, if not landmark. Although he is not recognized as an important theoretician in the strictest sense, this neither detracts from his accomplishments nor diminishes his innovations in applied psychometrics. Wechsler was a well-known clinician and, as such, he intentionally placed significant importance in developing tasks that had practical, clinical value, and not merely theoretical value. Thus, the driving force behind the development of the Wechsler Scales was no doubt based equally, if not more, on practical considerations rather than theoretical ones. Zachary (1990) stated, "when David Wechsler published the original Wechsler-Bellevue Scales in 1939, he said relatively little about the theoretical underpinnings of his new instrument; rather, he followed a pragmatic approach. He selected a set of tasks that were easy to administer and score" (p. 276). Detterman (1985) also attributed much of the popularity of the Wechsler family of tests to their "ease of administration fostered by an organization of subtests that are brief. . . and have long clinical histories" (p. 1715). For better or worse, Wechsler's primary motivation for constructing his tests was to create an efficient, easy-to-use tool for clinical purposes; operationalizing them on a specific theory of intelligence was not of paramount importance.

It can be argued reasonably that the popularity and longevity of the Wechsler Scales is more strongly attributed to atheoretical features. Wechsler introduced numerous innovations into the arena of applied intelligence testing that had an immediate appeal to assessment professionals and aided in creating the Wechsler Scales as a viable challenger and alternative to the Binet scales (Reynolds & Kaufman, 1990). Some of the more notable features introduced with the various Wechsler Intelligence Scales included: (1) separate norms for children and adults (with the introduction of the WISC; Wechsler, 1949); (2) the provision for the calculation of subtest standard scores, which made the test open to profile interpretation; (3) multiple-channel assessment that allowed for a fairer method of determining performance through verbal and nonverbal means; and (4) calculation of a new type of standard score, the Deviation IQ, which greatly reduced the theoretical and statistical inadequacies of the Ratio IQ (Zimmerman & Woo-Sam, 1985). In addition, there is some evidence that Wechsler's instruments might have reached their preeminent status precisely because of his reputation as a clinician (Zimmerman & Woo-Sam, 1985). Irrespective of the true reasons for the immediate acceptance and lasting popularity of the Wechsler Scales, their influence on the research and practice of psychology is unparalleled.

In spite of these accomplishments and accolades, under the critical eye of subsequent advancements in the field, the failure of the Wechsler Scales to keep abreast of contemporary intelligence research cannot be ignored. As will be described in Chapters 2 and 3, the extant literature reveals that the Wechsler Scales lack a modern theoretical base, prefer historical continuity over scientific innovation, and are dominated by practical rather than theoretical considerations—all of which raise doubts about the validity of inferences drawn from some of the Wechsler subtest and composite scores. It is clear that meaningful use and interpretation of the Wechsler Scales require the adoption of an alternative (fourth wave) approach in which contemporary theory, research, measurement principles, and hypothesis validation are integrated. Such alternative approaches to measuring intelligence with the Wechsler Scales are necessary in order to reliably and validly assess a broader and more in-depth range of abilities than that which can be accomplished through traditional methods.

Previous attempts have been made to adapt the Wechsler Scales in ways that improve diagnostic precision and interpretive accuracy. However, the research into clinical and psychometric profile analyses is not convincing, and the reliability and validity of such practices are questionable. We believe that clinical judgment and experience are insufficient stanchions on which defensible interpretations can be built. In contrast, the application of theory to intelligence test use and interpretation is beginning to form a solid empirical base. The *Gf-Gc* cross-battery approach offered in this book has considerable promise as an efficient, theoretically and statistically defensible method for assessing and interpreting the broad array of cognitive abilities specified in contemporary psychometric theory. The subsequent chapters of this book demonstrate how the principles and procedures of this approach can be applied to the Wechsler Scales in order to advance the science of measuring multiple cognitive abilities when using these instruments as the core battery in assessment, including the assessment of individuals from diverse cultural and linguistic backgrounds.

CHAPTER

2 Theories and Measures of Intelligence

A Continuum of Progress within the Psychometric Tradition

Classification is arguably one of the most central and generic of all our conceptual exercises...without classification, there could be no advanced conceptualization, reasoning, language, data analysis, or for that matter, social science research.

—K. D. Bailey (1994)

Although the approaches varied across the four waves of intelligence test interpretation described in Chapter 1, they all shared a common goal—the classification of individuals according to their cognitive abilities. The process of analyzing and classifying human cognitive abilities "has intrigued scientists for centuries" (Kamphaus et al., 1997, p. 33) and is a manifestation of the longstanding quest, since the beginning of our existence, to understand the world by creating order. Intelligence tests (such as the Wechsler Scales) have served as the taxonomic or classification tools of researchers and practitioners who have sought to understand and create order within the domain of human cognitive abilities.

Intelligence Tests as Taxonomic Tools

Systematic attempts to classify various parts of the world date back to the Greeks, notably Aristotle, who developed an elaborate taxonomic system for classifying the animal kingdom (Dunn & Everitt, 1982; Lorr, 1983). Today classification is an "activity that is essential to all scientific work" (Dunn & Everitt, 1982, p. 9). Indeed, a specialized science of classification of empirical entities known as *taxonomy* (Bailey, 1994; Prentky, 1994) is ubiquitous in all fields of study, because it guides our search for information or truth.

For centuries, we have both observed and sought to classify and understand differences between people. For example, Plato believed in assigning individuals to tasks for which they were best suited, and Aristotle studied gender and racial differences in mental characteristics (Minton & Schneider, 1980). These early observations and attempts to classify individuals illustrate one of the few irrefutable laws in psychology—the law of individual differences.

Individual differences are evident in the considerable variability that exists across all human traits such as weight, height, temperament, intellect, social skills, and facial characteristics. Thurstone (1935) captured the essence of individual difference in cognitive abilities when he stated:

> A large class of human activity is that which differentiates accomplishments. Just as it is convenient to postulate physical forces in describing the movements of physical objects, so it is natural to postulate abilities and their absence as primary causes of the successful completion of a task by some individual and of the failure of other individuals in the same task. (p. 45)

More recently, Neisser and colleagues (1996) conveyed a similar theme when they stated that "individuals differ from one another in their ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought" (p. 77). Thus, for many years scholars have offered the trait or construct of *intelligence* to explain and clarify the complex set of phenomena that account for individual differences in various cognitive capabilities.

Attempts to define the construct of intelligence and to explain and classify individual differences in intellectual functioning have spanned decades and have been characterized by significant variability. The differences between theories of intelligence is exemplified by the various multiple intelligences models that have been offered or revised recently to explain the structure of intelligence (some of which serve as the theoretical foundation of current fourth wave attempts to improve test interpretation). These include Carroll's Three-Stratum Theory of Cognitive Abilities, Gardner's Theory of Multiple Intelligences, the Cattell-Horn Fluid-Crystallized *(Gf-Gc)* theory, Feurestein's theory of Structural Cognitive Modifiability (SCM), the Luria-Das Model of Information Processing, and Sternberg's Triarchic Theory of Intelligence (see Flanagan, Genshaft, & Harrison, 1997, for a comprehensive description of these theories). Each of these theories represents an attempt to comprehend a class of phenomena and, ultimately, fulfill the chief goal of science—to minimize the mental effort needed to understand complex phenomena through classification (Thurstone, 1935, p. 45). To achieve this goal, each theory of intelligence provides a taxonomic framework for classifying and analyzing the nature of the cognitive characteristics that account for the variability in observed intellectual performance among and between individuals.

The remainder of this chapter presents a summary of the state-of-the-art theories of intelligence and provides the context within which the Wechsler Intelligence Scales are evaluated according to contemporary *Gf-Gc* theory, the theory that we believe provides the best framework for moving the interpretation of the Wechsler Scales into credible theory-based (fourth wave) interpretation.

Three General Paradigms for Conceptualizing and Measuring Intelligence

"There is an unlimited number of ways in which nature can be comprehended" (Thurstone, 1935, p. 47). At a general level, the variability in the theories and measures of intelligence can be explained by differences in underlying research traditions in psychological measurement. Taylor (1994) suggests that the psychometric, information processing, and cognitive modifiability theories are the most prominent approaches used to conceptualize the measurement of intelligence.[1]

The *psychometric* or structural approach "attempts to measure performance along dimensions which are purported to constitute the fundamental structure of the psychological domain" (Taylor, 1994, p. 185). In the psychometric approach, psychological tests that yield scores on quantitative scales are used. Correlational and factor analytic methods are employed typically to analyze these scores and identify ability dimensions that form the structure of individual differences in cognitive ability (Gustafsson & Undheim, 1996). The Wechsler Scales and the various approaches to the interpretation of intelligence tests that characterized all four waves of test interpretation (described in Chapter 1) are all products of the psychometric approach to measuring intelligence.

Information-processing theories are more recent in origin (largely since the 1960s) and, in general, have taken a cognitive-rational view of human intellectual functioning using the computer analogy of humans as information processors. In general, information processing theories are "limited capacity theories of cognitive competence" (Taylor, 1994, p. 185) that are concerned with how information is processed efficiently during problem solving and everyday tasks. Information processing approaches view individuals who can process information efficiently through one or more "bottlenecks" as being competent or intelligent. For example, working memory is considered to be a bottleneck because it is a limited-capacity system that can only hold and process a finite amount of information at any one time. As a result, if information is not processed efficiently through working memory, the entire system does not perform at an optimal level. Individuals who have developed skills and strategies for efficient processing of information through working memory can perform at higher levels and are thus considered to be more intelligent (Taylor, 1994).

Information processing research typically uses fine-grained computer-administered chronometric measures of human performance and functioning (e.g., inspection time, average evoked potentials, nerve conduction velocity, reaction time). Currently, practical adaptations of chronometric measures to applied intelligence testing have yet to surface. However, the recent addition of a test of working memory (viz., Letter-Number Sequencing) to the WAIS-III (Wechsler, 1997) and the third edition of the Woodcock-Johnson (WJ-III; Woodcock et al., in press), as well as tests of attention, planning, and rapid automatic naming to the WJ-III and Cognitive Assessment System (CAS: Das & Naglieri, 1997) suggests that information processing concepts are beginning to influence the design and interpretation of intelligence batteries.

Cognitive modifiability theories are based primarily on Vygotsky's (1978) view that cognitive development is a social phenomenon. In particular, they have focused on the "capacity of humans to adapt to circumstantial demands—in other words, to learn to function effectively in their environment" (Taylor, 1994, p. 187). Underlying this conception of cognitive development is the belief that intelligence is dynamic, modifiable, and changeable. Dynamic assessment, which evolved from cognitive modifiability theories, "refers to approaches to the development of decision-specific information that most characteristically involve interaction between the examiner and the examinee, focus on learner metacognitive processes and responsiveness to intervention, and follow a pretest-intervene-posttest administration format (Lidz, 1987; 1991)" (Lidz, 1997, p. 281).

Three general approaches to dynamic assessment have been proposed (Laughon, 1990). In general, dynamic assessment approaches are characterized by an attempt to measure processes through the integration of teaching into the assessment process. With the exception of the Kaufman Adolescent and Adult Intelligence Test (KAIT) and the Woodcock-Johnson series of intelligence batteries (Woodcock & Johnson, 1977; 1989; Woodcock et al., in press), none of the major individually administered intelligence tests include actual learning tasks, let alone tasks and procedures that are consistent with the dynamic assessment model. Kaufman's (1979) criticism of the WISC-R's failure to include actual learning tasks, especially of higher-order cognitive abilities, continues to apply to all revisions of the Wechsler trilogy. The reader is referred to Budoff (1968; 1974; 1987), Feuerstein (1970; 1972; 1979), and Campione and Brown (1987) for a description of the varying dynamic assessment procedures. The interested reader is also referred to Feuerstein, Feuerstein, and Gross's (1997) cogent description of perhaps the most well-known of the dynamic assessment procedures, the Learning Potential Assessment Device.

In summary, the psychometric approach to understanding the structure of intelligence is the oldest and most established of the three approaches, dating back to Galton's attempt, in the late 1800s, to measure intelligence with psychophysical measures (Sternberg & Kaufman, 1998). It is also the approach that is the most research based and that has produced the most economically efficient and practical instruments for measuring intelligence (e.g., Wechsler Scales) (Neisser et al., 1996; Taylor, 1994). Psychometric theories and measures continue to be the dominant force even during the current theory-based wave of interest in test interpretation. On the other hand, to date, the newer information processing and cognitive modifiability theories have produced little in the way of practical measurement tools. The reader is referred to Carroll (1993a), Gustafsson and Undheim (1996), Ittenbach, Esters, and Wainer (1997), Kamphaus (1993), Sattler (1988), and Thorndike and Lohman (1990) for historical information on the development of psychometric theories of intelligence.

Psychometric Theories and Measures of Intelligence: Where Do the Wechsler Intelligence Scales Fit?

The evolution of research on intelligence measures and psychometric theories chronicles the many attempts and progressions toward specifying a "complete" taxonomy of human cognitive abilities. In lieu of a detailed discussion of this research literature, an adaptation and extension of Woodcock's (1994) "continuum of progress in theories of multiple intelligences" is presented. This continuum, shown in the top of Figure 2.1, summarizes the *that was then-this is now* progression of psychometric theories of intelligence. In addition to this continuum, the bottom of Figure 2.1 depicts a parallel continuum of prominent approaches to the applied measurement of intelligence. Through an examination of Figure 2.1, the relation between the respective theoretical benchmarks and intelligence measures is readily apparent. It is important to note that the continua presented in Figure 2.1 do not portray linear *timelines;* rather, they portray the *progress* in understanding and measuring the structure of human intelligence. Note also that the specific theories listed under the theory continuum are illustrative examples and are not intended to represent a complete list of theories.

That Was *Then:* Early or "Incomplete" Taxonomic Theories and Models of Intelligence

Spearman's *g*-Factor Theory

Sir Francis Galton is considered by many to be the father of differential psychology. However, the birth of the psychometric research tradition, by and large, is considered to have begun with Spearman's (1904; 1927) presentation of the general or *g*-factor theory of intelligence and his development and application of factor analytic methods to general mental ability measures (Jensen, 1998). In fact, Spearman's 1904 paper, "General Intelligence Objectively Determined and Measured," is "perhaps the single most important paper in the history of differential psychology and psychometrics" (Jensen, 1998, p. 21). The fundamental premise of Spearman's theory is that a single *g* or general intelligence ability accounts for the performance of individuals on most all types of cognitive tasks. As discussed in Chapter 1, Spearman's general factor theory was largely responsible for the focus on the quantification of a general level of intellectual functioning during the first wave of intelligence test interpretation.

Although Spearman's *g* theory is typically described as a single-factor theory (as displayed in Figure 2.1), this characterization is not completely accurate. In addition to the large *g* factor, Spearman's theory also includes smaller specific *(s)* factors. Spearman eventually became interested in specific cognitive ability factors and, together with Karl Holziner, developed the *bi-factor* model. According to Carroll (1993a), if Spearman had lived beyond 1945, he most likely would have converged on a multiple abilities model similar to the ones proposed by other researchers (e.g., Thurstone's primary mental abilities, which is discussed later in this chapter). As shown in Figure 2.1, the most notable applied measure of intelligence that reflects the *g* model was the omnibus Stanford-Binet Intelligence Scale (Terman, 1916; Terman & Merrill, 1937; Terman & Merrill, 1960; Terman & Merrill, 1972), which provided a single composite intelligence score. The *g*-based Binet was the intelligence test that achieved preeminent status during the first wave of intelligence test interpretation.

Dichotomous Theories and Models

The demise of Spearman's bi-factor model had begun as early as 1909, primarily as a result of the evidence presented by Sir Cyril Burt in favor of *group factors.* According to Jensen (1998), by 1911, Burt's data had convinced most psychologists that it was more reasonable

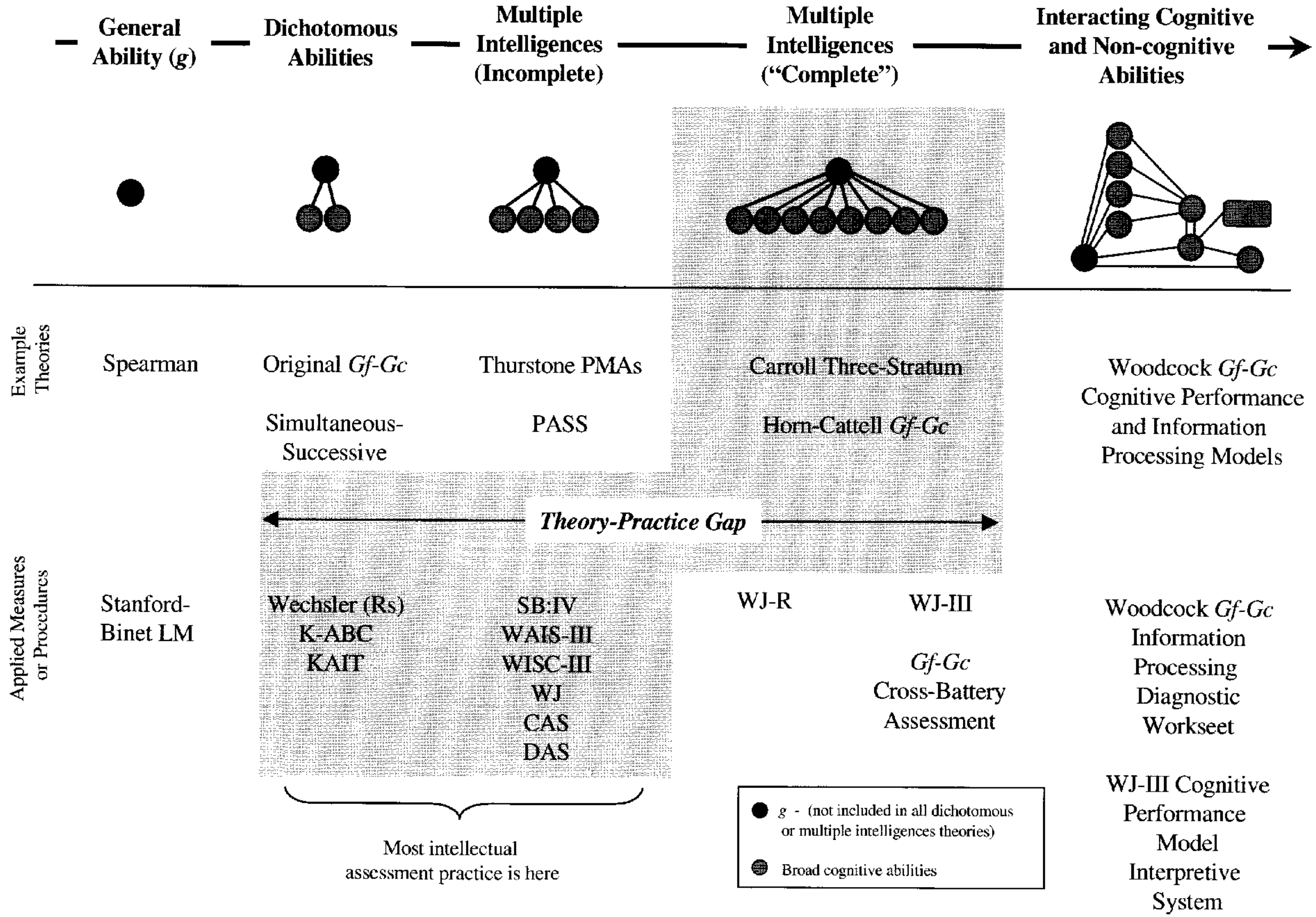

FIGURE 2.1 Progress in Psychometric Theories and Applied Measures of Intelligence

to accept the existence of group factors, in addition to *g* and *s*. The evidence for factors beyond *g* resulted in a variety of attempts to develop both theories and measures of group-factor abilities. The recognition that a complete understanding of intelligence required the measurement and interpretation of abilities beyond *g* set the stage for the eventual development and interpretation of new intelligence tests, or the post-hoc interpretation of existing intelligence tests (viz., the Binet) in a more differentiated manner (e.g., the clinical and psychometric profile analyses that characterized the second and third wave of test interpretation, respectively).

One of the more prominent dichotomous models of intelligence was Cattell's (1941; 1957) original *fluid (Gf)* and *crystallized (Gc)* intelligence theory. Cattell, who earned his doctoral degree under Spearman, suggested that g was not a unitary trait, but rather a composite of two different types of general factors or abilities representing novel problem solving *(Gf)* and consolidated knowledge *(Gc)* (Jensen, 1998).

The dichotomous *Gf-Gc* theory did not result in any widely used practical assessment instrument at the time. However, many attempts were made to interpret intelligence batteries according to the dichotomous *Gf-Gc* model during the second and third waves of test interpretation. In the case of Wechsler's tests, differences between the Verbal and Performance scales have been suggested to be "indicative of differences in fluid and crystallized ability rather than in verbal and nonverbal thinking" (Kaufman, 1994, p. 167). As will be seen in subsequent chapters, only the Verbal/*Gc* Wechsler interpretation is supported by contemporary research.

Although the KAIT (Kaufman & Kaufman, 1993) is a relatively new intelligence battery, its organization around the dichotomous *Gf-Gc* model has warranted its placement near the "older" end of the theory continuum in Figure 2.1. As will be presented later in this section, John Horn's program of research had a significant impact on the evolution of Cattell's model into a multidimensional group-factor theory. This theory has significantly influenced the development of theory-based intelligence tests, as well as the current fourth wave theory-based methods for interpreting intelligence test performance (McGrew & Flanagan, 1998).

Without a doubt, the Wechsler *verbal/nonverbal* (performance) model of intelligence is the most widely recognized dichotomous model of cognitive abilities. It is also the model that has produced the most frequently used intelligence batteries to date (viz., the Wechsler Intelligence Scales). As discussed in Chapter 1, David Wechsler designed his first scale based on a combination of clinical, practical, and empirical considerations (Kaufman, 1990a; Zachary, 1990) and did not regard the Verbal-Performance dichotomy as representing two different types of intelligence. Rather, his intent was to organize the tests to reflect the two different ways (i.e., two different "languages") through which intelligence can be expressed (Kamphaus, 1993; Reynolds & Kamphaus, 1990; Zachary, 1990). Although verbal abilities represent a valid cognitive ability construct (i.e., crystallized intelligence), there is no such thing as *nonverbal ability*—only abilities that are expressed nonverbally. While the above information is presented in greater detail in the "theory" section of this book, it should serve to remind assessment professionals that the various versions of Wechsler's Scales *were not (and are not) based on an empirically supported theory of intelligence*. The relation between the Wechsler model and contemporary empirically based models of intelligence is discussed later in this chapter.

Finally, since the early to mid 1980s there has been interest in developing tests that have their roots in a neuropsychological model of cognitive processing advanced by Soviet neuropsychologist, A. R. Luria (Sternberg, 1997). Luria's work (1966; 1970; 1973; 1980), plus related experimental and cognitive psychological research (Anokhin, 1969; Broadbent, 1958; Das, Kirby & Jarman, 1979; Hunt & Lansman, 1986), has suggested a model of cognitive processing based on two to four mental operations (Kamphaus, 1990; 1993; Kamphaus & Reynolds, 1984; Kaufman, 1984; Kaufman & Kaufman, 1983; Naglieri, 1997; Naglieri & Das, 1990).

As reflected in Figure 2.1, the two-factor *simultaneous/successive* processing model of intelligence spawned the development of the Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983). *Simultaneous processing,* primarily associated with the right cerebral hemisphere, is involved in the integration or synthesis of stimuli into groups when the individual components of the stimuli are interrelated (Kaufman, 1994; Naglieri, 1997). In contrast, *successive processing,* typically associated with left-hemisphere functioning, is involved when the individual stimuli are processed in a serial order and there is no point in time at which the stimuli are interrelated.

The K-ABC was the first norm-referenced cognitive battery that was designed primarily to operationalize the simultaneous and successive processing dichotomy. When evaluated in the context of current psychometric research and theory, the K-ABC has been found to assess only a very limited range of known mental abilities, namely visual processing and short-term memory (Carroll, 1993a; McGrew & Flanagan, 1998; Woodcock, 1990). According to Carroll (1993a), visual processing and short-term memory factors have been included in the cognitive factor analytic research literature well before the publication of the K-ABC; therefore, "there is little if anything that is new in the K-ABC test" (p. 703).

Attempts have also been made to interpret the Wechsler Scales from the simultaneous/successive processing framework (Kaufman, 1994). From this perspective the Wechsler perceptual organization tests (viz., Picture Completion, Block Design, and Object Assembly) are interpreted as measures of simultaneous processing, while the Picture Arrangement, Digit Span (forward) and Coding tests are interpreted as indicators of successive processing. The reader is referred to Kaufman (1994) for details regarding the application of this dichotomous model to the Wechsler Scales.

"Incomplete" Multiple Intelligences Theories and Models

The earliest attempt to identify multiple intelligences—a development that contributed to the proliferation of attempts to understand a person's profile of abilities via clinical and psychometric analysis—was undoubtedly Thurstone's factor-analysis-based efforts to identify *primary mental abilities* (*PMA;* Thurstone, 1938; Thurstone & Thurstone, 1941). The PMA theory suggested that, rather than being a function of *g*, performance on psychometric tests of cognitive ability was due to a number of primary mental abilities or faculties such as Space, Perceptual Speed, Number, Verbal Meaning, Word Fluency, Memory, and Inductive Reasoning (Kamphaus, 1993). Thurstone's PMA model is significant, as most modern test construction tends to be based on it (Taylor, 1994). Although some of the primary mental abilities are reflected in certain Wechsler tests (e.g., Verbal Meaning and the Vocabulary test; Number and the Arithmetic test), these are post-hoc interpretations; the original Wechsler Scales and subsequent revisions were not explicitly constructed to operationalize all or part of the PMA model. Other examples of factor analytically based models are seen in the works of Burt (1949), French, Eckstrom, and Price (1963), and Vernon (1961). Given the benefit of hindsight, the generation of multiple intelligences theories described thus far is now seen as being relatively "incomplete."

As depicted in Figure 2.1, the simultaneous/successive processing model has recently evolved into a four-construct *Planning, Attention, Simultaneous,* and *Successive* model (*PASS;* Das & Naglieri, 1997; Naglieri & Das, 1997) based on the work of A. R. Luria (1966). *Planning,* which has been an important construct in the neuropsychological arena, is one of a number of activity-related executive functions used to identify and organize steps required to achieve a goal or carry out an intention. It is characterized by forward thinking, the generation of alternatives, the weighing and making of choices, and the development of a framework or structure that provides direction in the completion of a plan (Lezak, 1995). According to Naglieri (1997), *attention* involves those processes that allow individuals to focus and respond to a particular stimulus while concurrently ignoring competing stimuli.

The *Cognitive Assessment System* (*CAS;* Das & Naglieri, 1997) is a psychometric intelligence battery specifically designed to operationalize the PASS model. Thus, it is an example of a recent attempt to move test interpretation into the fourth wave. Based on independent research and reviews (Carroll, 1993a; 1995a; Kranzler, Flanagan, & Keith, 1999; Kranzler & Keith, in press; Kranzler & Weng, 1995), the CAS is classified as an incomplete measure of cognitive abilities and its underlying theory. In addition to the above sources, the reader is referred to McGrew and Flanagan (1998), Das, Naglieri, and Kirby (1994), Naglieri (1997), and Chapter 5 of this book for additional information regarding the differing interpretations of the constructs measured by the CAS.

There have been recent attempts to interpret the Wechsler Scales from the perspective of the PASS model. Naglieri (1997) suggested the following: (1) the Wechsler Performance scale is primarily (but not exclusively) a measure of simultaneous processing; (2) Digit Span (forward only) measures successive processing; and (3) the Verbal scale represents verbal/achievement abilities that involve a variety of PASS processes. Although Naglieri believes that planning and attention are not adequately assessed by the Wechsler Scales, Kaufman (1994) suggested that the Wechsler Scales Processing Speed Index measures planning ability and that the Arithmetic, Digit Span, and Digit-Symbol/Coding subtests, which fall under the Freedom from Distractibility factor, measure attention. The reader is referred to Das and colleagues (1994), Kaufman (1994), and Naglieri (1997) for detailed discussions of hypothesized PASS interpretations of the Wechsler Scales. As will be demonstrated in Chapter 3, the extant cognitive abilities factor analytic research does not support either Naglieri's (1997) or Kaufman's (1994) interpretations of the Wechsler-PASS relationship. Rather, this research suggests that the Wechsler subtests are best understood as measures of a narrow range of *Gf-Gc* abilities.

As presented in Figure 2.1, the majority of currently used intelligence batteries are classified as incomplete measures of multiple cognitive abilities (viz., CAS, Differential Abilities Scales [DAS; Elliott, 1990a], Stanford-Binet Intelligence Scale: Fourth Edition

[SB:IV; Thorndike, Hagen, & Sattler, 1986], and the original Woodcock-Johnson Psycho-Educational Battery [WJ; Woodcock & Johnson, 1977]). Of these three batteries, the DAS appears to assess the broadest array of cognitive abilities (McGrew, 1997).

As presented in Figure 2.1, the most recent Wechsler Intelligence Scales (i.e., Wechsler Intelligence Scale for Children—Third Edition [WISC-III; Wechsler, 1991] and WAIS-III) represent progress in the evolution of these batteries. However, the WISC-III and WAIS-III are still classified as "incomplete" since they measure only a subset of the known broad cognitive abilities (viz., three to five *Gf-Gc* abilities). However, as will be discussed in later chapters, when combined with other measures (e.g., the Children's Memory Scale [CMS; Cohen, 1997]; Wechsler Memory Scale—Third Edition [WMS-III; Wechsler, 1997]; WJ-III), the Wechsler system of instruments can aid in narrowing the intelligence theory-practice gap.

This Is *Now:* Contemporary or "Complete" Taxonomic Theories and Models of Intelligence

As portrayed in Figure 2.1, psychometric intelligence theories have converged recently on a more "complete" (in a relative sense, no theory is ever complete) *Gf-Gc* multiple intelligences taxonomy, reflecting a review of the extant factor analytic research conducted over the past 50 to 60 years. This taxonomy serves as the organizational framework for both the Carroll and Cattell-Horn models (Carroll, 1983; 1989; 1993a; 1997; Gustafsson, 1984; 1988; Horn, 1988; 1991; 1994; Horn & Noll, 1997; Lohman, 1989; Snow, 1986), the two most prominent psychometric theories of intelligence proposed to date (McGrew & Flanagan, 1998; Sternberg & Kaufman, 1998).

As depicted in Figure 2.1, only the Woodcock-Johnson Psycho-Educational Battery—Revised (WJ-R; Woodcock & Johnson, 1989) and the Woodcock-Johnson Psycho-Educational Battery—Third Edition (WJ-III; Woodcock et al., in press) come close to measuring the broad abilities specified in the more "complete" psychometrically based *Gf-Gc* multiple intelligences theories. This is not surprising given that the *Gf-Gc* framework drove the design of both the WJ-R and WJ-III. Recent joint or cross-battery factor analyses of the major intelligence batteries with the WJ-R (e.g., Flanagan & McGrew, 1998; McGhee, 1993; McGrew, 1997; Woodcock, 1990) indicated that the majority of these batteries do not adequately assess the complete range of *broad Gf-Gc* abilities included in either Horn's (1991; 1994) or Carroll's (1993a; 1997) model of the structure of intelligence. The one possible exception is the WJ-III which was designed to measure the greatest practically feasible range of *Gf-Gc* abilities. But even the WJ-III may benefit from supplementation via *Gf-Gc* cross-battery procedures (McGrew & Flanagan, 1998), particularly when attempting to assess more thoroughly a person's specific or narrow cognitive abilities.

Finally, while research continues to focus on identifying major abilities in the multiple intelligences taxonomy (Carroll, 1993a), a number of researchers are attempting to push the far end of the intelligence theory continuum by proposing models that describe and explain cognitive performance as a composition of both cognitive and noncognitive variables within an information processing framework (see Figure 2.1). For example, Woodcock (1993; 1997; 1998) has presented a *Gf-Gc Cognitive Performance Model (CPM)* and an *Information Processing Model (IPM)*. These recent theoretical developments, which are described later in this chapter, will continue to stimulate efforts in theory-based interpretation (i.e., the fourth wave) and may eventually contribute to a (yet to be identified) fifth wave of interpretation.

What Is *Gf-Gc* Theory?

Gf-Gc theory is the most comprehensive and empirically supported psychometric theory of intelligence. Therefore, we believe that the *Gf-Gc* theory should serve as a foundation for the development and interpretation of intelligence batteries. In order to implement a *Gf-Gc*-based approach to assessing and interpreting cognitive functioning with the Wechsler Intelligence Scales, it is necessary to understand the major components of the theory.

The Evolution of *Gf-Gc* Theory

Cattell (1941; 1957) first postulated *Gf-Gc* theory as consisting of two major types of cognitive abilities (i.e., *Gf* and *Gc*). Fluid Intelligence *(Gf)* was thought to include inductive and deductive reasoning, abilities thought to be influenced by both biological and neurological factors and incidental learning through interaction with the environment (Taylor, 1994). In contrast, Crystallized Intelligence *(Gc)* was believed to consist primarily of abilities (especially knowledge) that reflected the influences of acculturation (viz., verbal-conceptual knowledge; Gustafsson, 1994; Taylor, 1994). Thus, the original *Gf-Gc* theory was a dichotomous conceptualization of human cognitive ability. Unfortunately (or fortunately, depending on one's belief in maintaining the historical integrity of a theory), the *Gf-Gc* label has been retained as the acronym for this theory, despite the fact that the theory has not been conceived of as a dichotomy since the 1960s (Gustafasson &Undheim, 1986; Horn & Noll, 1997; Woodcock, 1993). As a result, *Gf-Gc* theory is misunderstood often as being a two-factor model of the structure of intelligence.

As early as the mid-1960s, Horn (1965) expanded the *Gf-Gc* model to include four additional cognitive abilities, including visual perception or processing *(Gv)*, short-term memory (Short-Term Acquisition and Retrieval—SAR or *Gsm*), long-term storage and retrieval (Tertiary Storage and Retrieval—TSR or *Glr*), and speed of processing *(Gs)*. By 1968, Horn had refined the definition of *Gv, Gs,* and *Glr,* and added auditory processing ability *(Ga)*. More recently, factors representing a person's quantitative ability or knowledge *(Gq)* and facility with reading and writing *(Grw)* (Horn, 1985; 1988; 1991; Woodcock, 1994) were added to the model, resulting in a ten-factor ability structure.

The Hierarchical Structure of *Gf-Gc* Theory

In his review of the extant factor-analytic research literature, Carroll (1993a) differentiated factors or abilities by three strata that varied according to the "relative variety and diversity of variables" (Carroll, 1997, p. 124) included at each level. The various *"G"* abilities are the most prominent and recognized abilities in the model. They include *Gf, Gc,* and so on. These abilities are classified as broad or stratum II abilities in Carroll's model. The *broad* *abilities* represent "basic constitutional and longstanding characteristics of individuals that can govern or influence a great variety of behaviors in a given domain" and they vary in their emphasis on process, content, and manner of response (Carroll, 1993a, p. 634). What is often not immediately clear when discussing *Gf-Gc* theory is that the broad abilities subsume a large number of narrow or stratum I abilities (currently approximately 70 have been identified; Carroll, 1993a; 1997). *Narrow abilities* "represent greater specializations of abilities, often in quite specific ways that reflect the effects of experience and learning, or the adoption of particular strategies of performance" (Carroll, 1993a, p. 634). The hierarchical structure of *Gf-Gc* theory is demonstrated for the domain of visual processing *(Gv)* in Figure 2.2.

In the *Gf-Gc* taxonomy, *Gv* is classified as a broad stratum II cognitive ability. The 11 narrow or stratum I visual abilities that comprise *Gv* clearly demonstrate the "broadness" or breadth of this factor (see Figure 2.2). Figure 2.2 conveys that 11 different narrow or specialized visual processing abilities have been identified. The broad and narrow *Gv* abilities presented in Figure 2.2, as well as the other *Gf-Gc* broad and narrow abilities, are defined later in this chapter. The significant moderate to high intercorrelations displayed by these narrow abilities suggests the presence of a broader factor or construct that accounts for this shared visual processing variance. The broad *Gv* factor is hypothesized to represent this higher-order explanatory construct and is believed to exert a significant common effect (reflected by the direction of the arrows in Figure 2.2) on the narrow abilities. When extended to the nine other broad cognitive domains, all of which also subsume a number of narrow abilities, it is clear that the contemporary hierarchical *Gf-Gc* theory is extremely comprehensive.

Even without the benefit of the information presented in subsequent chapters of this book, after reflecting on Figure 2.2, the experienced Wechsler user should be able to identify relations between narrow *Gv* abilities and certain Wechsler nonverbal tests (e.g., the narrow ability of Spatial Relations and the Block Design tests). The experienced Wechsler

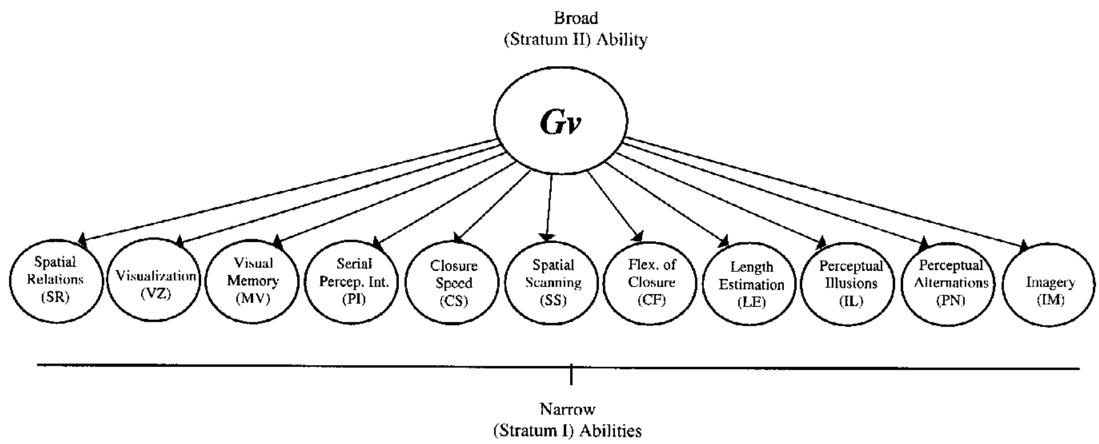

FIGURE 2.2 A Visual Processing *(Gv)* Example Demonstrating the Hierarchical Structure of *Gf-Gc* Theory

user may also observe that the Wechsler Performance scale only measures a subset of the entire *Gv* domain, and that some tests do not seem to fit the empirical taxonomy (e.g., Picture Arrangement). Such observations should set the stage for assessment professionals to recognize the breadth of coverage of certain *Gf-Gc* abilities (or lack thereof) by their favorite intelligence batteries. This will be discussed in detail in subsequent chapters.

The broadest or most general level of ability in the *Gf-Gc* model is represented by stratum III, located at the apex of the Carroll (1993a) hierarchy. This single cognitive ability which subsumes both broad (stratum II) and narrow (stratum I) abilities, is interpreted by Carroll as representing a general factor (i.e., *g*) that is involved in complex higher-order cognitive processes (Gustafsson & Undheim, 1996).

Finally, it is important to recognize that the abilities within each level of the hierarchical *Gf-Gc* model typically display non-zero positive intercorrelations (Carroll, 1993a; Gustafsson & Undheim, 1996). For example, similar to the *Gv* discussion above, the different stratum I (narrow) abilities that define the various *Gf-Gc* domains are correlated positively to varying degrees. These intercorrelations give rise to and allow for the estimation of the stratum II (broad) ability factors. Likewise, the positive non-zero correlations among the stratum II (broad) *Gf-Gc* abilities allows for the estimation of the stratum III (general) *g* factor. The positive factor intercorrelations within each level of the *Gf-Gc* hierarchy indicates that the different *Gf-Gc* abilities do not reflect independent (uncorrelated or orthogonal) traits.

The Carroll and Cattell-Horn *Gf-Gc* Models

The simplified (i.e., narrow abilities omitted) Cattell-Horn and Carroll *Gf-Gc* models are presented together in Figure 2.3. A review of Figure 2.3 reveals a number of notable similarities and differences between the two models. In general, these models are similar in that they both include some form of fluid *(Gf)*, crystallized *(Gc)*, short-term memory and/or learning (*Gsm* or *Gy*), visual *(Gv)*, auditory (*Ga* or *Gu*), retrieval (*Glr* or *Gr*), processing speed *(Gs)*, and decision and/or reaction time speed (*CDS* or *Gt*) abilities. Although there are some differences in broad ability definitions and in the narrow abilities subsumed by the respective broad *Gf-Gc* abilities, the major differences between the two models are primarily four-fold (McGrew, 1997).

First, the Carroll and Cattell-Horn models differ in their inclusion of *g* at stratum III. According to Carroll (1993a; 1997), the general intelligence factor at the apex of his three-stratum theory is analogous to Spearman's *g*. The off-center placement of *g* (to the left side of Figure 2.3) in the Carroll model is intended to reflect the strength of the relations between *g* and the respective broad *Gf-Gc* abilities. As represented in the Carroll model portion of Figure 2.3 (i.e., the top half of the figure), *Gf* has been reported to have the strongest association with *g*, followed next by *Gc*, and continuing on through the remaining abilities to the two broad abilities that are weakest in association with *g* (i.e., *Gs* and *Gt*).[2]

Carroll (1997) believes that the evidence for *g* is overwhelming. Horn disagrees with Carroll (see Horn, 1991; Horn & Noll, 1997), and instead posits what Jensen (1998) calls a *truncated hierarchical model*, a model that does not contain a single *g* factor at the apex. Debates about the nature and existence of *g* have waxed and waned for decades and have been some of the liveliest debates in differential psychology (Gustafsson & Undheim, 1996; Jensen, 1997). Much of the debate has been theoretical in nature with definitions of *g* ranging from an index of neural cognitive efficiency, general reasoning ability or mental energy to a mere statistical irregularity (Neisser et al., 1996). After being more or less banned from the scientific scene (Gustafsson & Undheim, 1996), the prominent position of *g* in contemporary models of intelligence (e.g., Carroll's Three-Stratum model and Jensen's [1998] seminal *g* factor treatise) has helped it to once again take center stage in intelligence research and dialogue. Interested readers are directed to the writings of Carroll (1993a; 1997), Horn (1991), Horn and Noll (1997), and Jensen (1997; 1998) for further discussion of the *g*-related issues and research.

Second, in the Cattell-Horn model, quantitative knowledge and quantitative reasoning abilities together represent a distinct broad ability, as depicted by the *Gq* rectangle in the bottom half of Figure 2.3. Carroll (1993a), on the other hand, considers quantitative ability to be "an inexact, unanalyzed popular concept that has no scientific meaning unless it is referred to the structure of abilities that compose it. It cannot be expected to constitute a higher-level ability" (p. 627). Therefore, Carroll classifies quantitative reasoning as a nar-

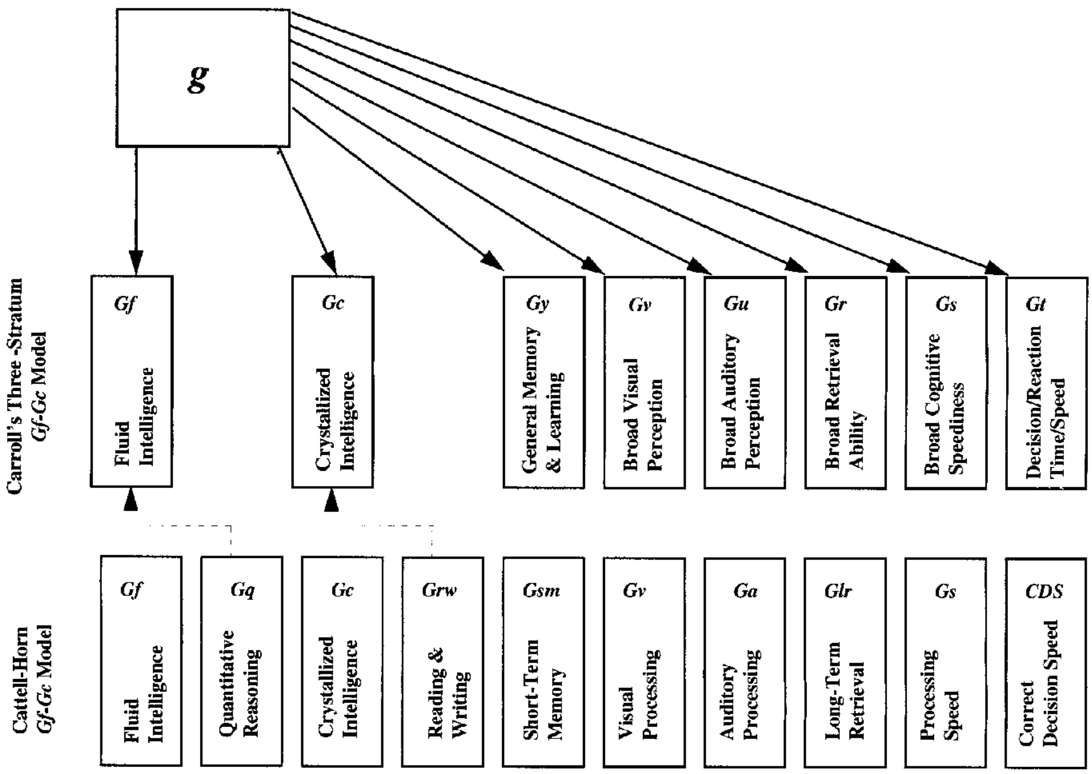

FIGURE 2.3 The Broad and General Strata of the Cattell-Horn and Carroll *Gf-Gc* Models

row ability subsumed by *Gf,* as indicated by the arrow leading from the *Gq* rectangle in the Cattell-Horn model to the *Gf* rectangle in the Carroll model in Figure 2.3. Furthermore, Carroll included mathematics achievement and mathematics knowledge factors in a separate chapter in his book which described a variety of knowledge and achievement abilities (e.g., technical and mechanical knowledge; knowledge of behavioral content) that are not included in his theoretical model.

Third, recent versions of the Cattell-Horn model have included a broad English-language reading and writing ability *(Grw)* that is depicted in the bottom half of Figure 2.3 (McGrew, 1997; Woodcock, 1993). Carroll, however, considers reading and writing to be narrow abilities subsumed by *Gc,* as reflected by the arrow leading from the *Grw* rectangle in the Cattell-Horn model to the *Gc* rectangle in the Carroll model in Figure 2.3.

Finally, the Carroll and Cattell-Horn models differ in their treatment of certain narrow memory abilities. Carroll combines both short-term memory and the narrow abilities of associative, meaningful, and free recall memory (defined later in this chapter) with learning abilities under his General Memory and Learning factor *(Gy).* Horn (1991) makes a distinction between immediate apprehension (e.g., short-term memory span) and storage and retrieval abilities, while Carroll combines them into a single broad ability *(Gy).* However, Horn (1988) indicated that it is often difficult to distinguish short-term memory and storage from retrieval abilities. For example, in some of his writings Horn (1991) referred to associative memory as a narrow ability subsumed by short-term memory. However, Horn (1988) listed the Delayed Recall tests of the WJ-R (Woodcock & Johnson, 1989), which are measures of associative memory (McGrew, 1997), under the long-term storage and retrieval ability *(Glr).*

Seeking a Standard Nomenclature: An Integrated Cattell-Horn-Carroll *Gf-Gc* Model

Notwithstanding the important differences between the Cattell-Horn and Carroll models, in order to realize the practical benefits of the calls for more theory-based interpretation (Kaufman, 1979; Kamphaus et al., 1997), it would be useful if a single *Gf-Gc* taxonomy is used to classify the individual tests in intelligence batteries. A first effort to create a single *Gf-Gc* taxonomy for use in the evaluation and interpretation of intelligence batteries was the integrated Cattell-Horn-Carroll model (McGrew, 1997). McGrew and Flanagan (1998) subsequently presented a slightly revised integrated model, which has been further revised in the current work via two changes (i.e., the splitting of Phonetic Coding into separate analysis and synthesis abilities under *Ga* and the inclusion of working memory under *Gsm;* see Figure 2.4).

The exclusion of *g* in Figure 2.4 does not mean that the integrated model used in this text does not subscribe to a separate general human ability or that *g* does not exist. Rather, it was omitted by McGrew (1997) and McGrew and Flanagan (1998) as it was judged to have little practical relevance to *Gf-Gc* cross-battery assessment and interpretation. That is, their cross-battery approach was designed to improve psychological and psychoeducational assessment practice by describing the unique *Gf-Gc* pattern of abilities of individuals. This pattern of abilities can then be related to important occupational and achievement outcomes as well as other human traits (McGrew & Flanagan, 1998).

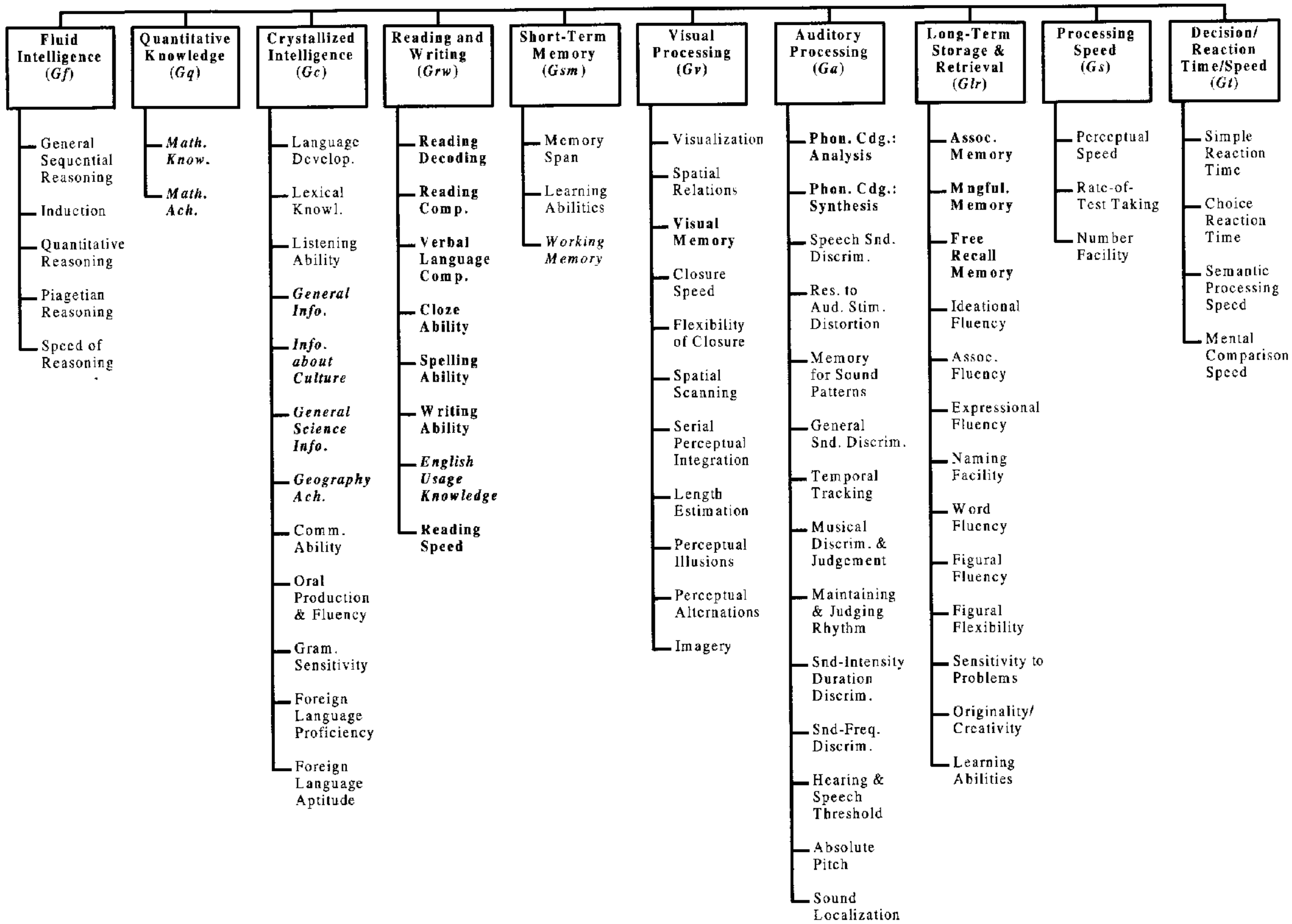

FIGURE 2.4 An Integrated Cattell-Horn-Carroll *Gf-Gc* Model of the Structure of Cognitive Abilities

Note: Italic font indicates abilities that were not included in Carroll's three-stratum model but were included by Carroll in the domains of knowledge and achievement. **Bold font** indicates abilities that are placed under different *Gf-Gc* broad abilities than in Carroll's model. These changes are based on the Cattell-Horn model and/or recent research (see McGrew, 1997, and McGrew & Flanagan, 1998). See Table 2.1 for definitions of narrow abilites.

Broad and Narrow *Gf-Gc* Ability Definitions

In this section the definitions of the broad and narrow abilities included in the *Gf-Gc* model are presented. These definitions are consistent with those presented in McGrew and Flanagan (1998). They were derived from an integration of the writings of Carroll (1993a), Gustafsson and Undheim (1996), Horn (1991), McGrew (1997), McGrew and colleagues (1991), and Woodcock (1994). The narrow ability definitions are presented in Table 2.1.

Fluid Intelligence (Gf). *Fluid Intelligence* refers to mental operations that an individual may use when faced with a relatively novel task that cannot be performed automatically. These mental operations may include forming and recognizing concepts, drawing inferences, comprehending implications, problem solving, and extrapolating. Inductive and deductive reasoning are generally considered to be the hallmark narrow-ability indicators of *Gf*. What may come as a surprise to experienced Wechsler users is the finding that, with the exception of the WAIS-III Matrix Reasoning test, the Wechsler Intelligence Scales do not measure much in the way of *Gf* abilities, one of the primary indicators of intelligent behavior. Definitions of the narrow abilities subsumed by *Gf* are presented in Table 2.1.

Crystallized Intelligence (Gc). *Crystallized Intelligence* refers to the breadth and depth of a person's acquired knowledge of a culture and the effective application of this knowledge. This store of primarily verbal or language-based knowledge represents those abilities that have been developed largely through the "investment" of other abilities during educational and general life experiences (Horn & Noll, 1997). Almost half of the tests in the Wechsler Scales measure various aspects of *Gc*.

Schematically, *Gc* might be represented by the interconnected nodes of a fishing net. Each node of the net represents an acquired piece of information, and the filaments between nodes (with many possible filaments leading to and from multiple nodes) represent links between different bits of stored information. A person high in *Gc* abilities would have a rich "fishing net" of information with many meaningfully organized and interconnected nodes. *Gc* is one of the abilities mentioned most often by lay persons when asked to describe an intelligent person (Horn, 1988). The image of a sage captures the essence of *Gc*.

Gc includes both declarative (static) and procedural (dynamic) knowledge. Declarative knowledge is held in long-term memory *(Glr)* and is activated when related information is in working memory *(Gsm)*. Declarative knowledge includes factual information, comprehension, concepts, rules, and relationships, especially when the information is verbal in nature. *Declarative knowledge* refers to knowledge "that something is the case, whereas procedural knowledge is knowledge of how to do something" (Gagne, 1985, p. 48). *Procedural knowledge* refers to the process of reasoning with previously learned procedures in order to transform knowledge. For example, a child's knowledge of his or her street address would reflect declarative knowledge, while a child's ability to find his or her way home from school would require procedural knowledge (Gagne, 1985). The breadth of *Gc* is apparent from the number of narrow abilities (i.e., 12) that it subsumes (see Table 2.1).

Quantitative Knowledge (Gq). *Quantitative Knowledge* represents an individual's store of acquired quantitative declarative and procedural knowledge. The *Gq* store of acquired knowledge represents the ability to use quantitative information and manipulate numeric symbols. The Wechsler Arithmetic test is an indicator of an aspect of *Gq*.

It is important to understand the difference between *Gq* and the Quantitative Reasoning (RQ) ability that is subsumed by *Gf*. *Gq* represents an individual's store of acquired mathematical knowledge, while RQ represents the ability to reason inductively and deductively when solving quantitative problems. *Gq* would be evident when a task requires mathematical skills and general mathematical knowledge (e.g., knowing what the square root symbol means). RQ would be required in order to solve for a missing number in a number series task (e.g., 2 4 6 8 ___). Two narrow abilities are listed and defined under *Gq* in Table 2.1.

Reading/Writing Ability (Grw). *Reading/Writing Ability* is an acquired store of knowledge that includes basic reading and writing skills required for the comprehension of written language and the expression of thought via writing. It includes both basic (e.g., reading decoding, spelling) and complex abilities (e.g., reading comprehension and the ability to write a story). Currently, this ability domain has been neither well defined nor extensively researched within the *Gf-Gc* framework. Also, in typical practice, *Grw* (and *Gq*) are considered to be *achievement* domains and are therefore measured by achievement tests and not intelligence tests. In Carroll's (1993a) three-stratum model, eight narrow reading and writing abilities are subsumed by *Gc* in addition to other abilities. In the *Gf-Gc* models presented by McGrew (1997), McGrew and Flanagan (1998), and the current authors (see Figure 2.4), these eight narrow abilities define the broad *Grw* ability. These *Grw* narrow abilities are defined in Table 2.1.

Short-Term Memory (Gsm). *Short-Term Memory* is the ability to apprehend and hold information in immediate awareness and then use it within a few seconds. *Gsm* is a limited capacity system, as most individuals can only retain seven "chunks" of information (plus or minus two chunks) in this memory system at one time. The ability to remember a telephone number long enough to dial it, or the ability to retain a sequence of spoken directions long enough to complete a task specified in the directions, are examples of *Gsm*. Given the limited amount of information that can be held in short-term memory, information is typically retained for only a few seconds before it is lost. As most individuals have experienced, it is difficult to remember an unfamiliar telephone number for more than a few seconds unless one consciously employs a cognitive learning strategy (e.g., continually repeating or rehearsing the numbers). Once a new task requires an individual to use their *Gsm* abilities to store new information the previous information held in short-term memory is either lost or must be stored in the acquired stores of knowledge (i.e., *Gc, Gq, Grw*) through the use of *Glr*.

More recently, the related construct of working memory has received considerable attention in the cognitive psychology literature, resulting in the inclusion of practical tests of working memory in recent revisions of two intelligence batteries (viz., WAIS-III; WJ-III). Working memory is considered to be the "mechanism responsible for the temporary storage and processing of information" (Richardson, 1996, p. 23). However, the integration of working memory into the *Gf-Gc* framework is hindered by the lack of a universally accepted definition of the construct (Logie, 1996). Working memory has been referred to

(text continued on page 42)

TABLE 2.1 Narrow *Gf-Gc* Stratum I Ability Definitions and Task Examples

	Gf-Gc Broad Stratum II Ability	
Narrow stratum I name (code)	**Definition**	**Task Example**
	Fluid Intelligence (Gf)	
General Sequential Reasoning (RG)	Ability to start with stated rules, premises, or conditions, and to engage in one or more steps to reach a solution to a novel problem.	An examinee is presented with an incomplete logic puzzle and must deduce the missing components following careful analysis of the presented stimuli.
Induction (I)	Ability to discover the underlying characteristic (e.g., rule, concept, process, trend, class membership) that governs a problem or a set of materials.	An examinee is presented with a certain pattern of related stimuli and must select one of several stimuli that would complete or continue the pattern.
Quantitative Reasoning (RQ)	Ability to inductively and deductively reason with concepts involving mathematical relations and properties.	An examinee is presented with an incomplete series of related numbers and must select the number(s) that best complete the series.
Piagetian Reasoning (RP)	Seriation, conservation, classification and other cognitive abilities as defined by Piaget's developmental theory.	An examinee must demonstrate knowledge of conservation of mass or volume when presented with transformations of either the actual state of the object or items extraneous to the object, such as a container holding the object (e.g., When 5 ounces of water is transformed to ice is there a change in the amount of water?).
Speed of Reasoning (RE)	(Not clearly defined by existing research.)	An examinee must say the days of the week while counting by 3's as quickly as possible (e.g., Monday, 3, Tuesday, 6, Wednesday, 9, etc.).
	Quantitative Knowledge (Gq)	
Mathematical Knowledge (KM)	Range of general knowledge about mathematics.	An examinee is asked to demonstrate knowledge of basic mathematical facts and operations.
Mathematical Achievement (A3)	Measured mathematics achievement.	An examinee is required to perform simple mathematical calculations using pencil and paper.

	Crystallized Intelligence (Gc)	
Language Development (LD)	General development, or the understanding of words, sentences, and paragraphs (*not* requiring reading), in spoken native language skills.	An examinee is presented with two words and must describe the common relation or similarity between them.
Lexical Knowledge (VL)	Extent of vocabulary that can be understood in terms of correct word meanings.	An examinee must provide oral definitions for words of increasing difficulty.
Listening Ability (LS)	Ability to listen and comprehend oral communications.	An examinee is presented with an incomplete verbal passage and must provide a word that completes the passage.
General (verbal) Information (K0)	Range of general knowledge.	An examinee must provide specific responses to questions of general factual information (e.g., In what direction does the sun rise?).
Information about Culture (K2)	Range of cultural knowledge (e.g., music, art).	An examinee is presented with pictorial depictions of major artistic works (e.g., the Mona Lisa) and must correctly identify the name of the work or the artist.
General Science Information (K1)	Range of scientific knowledge (e.g., biology, physics, engineering, mechanics, electronics).	An examinee must correctly respond to questions demonstrating general knowledge of basic scientific ideas or facts (e.g., What is the largest planet in our solar system? What is the ozone layer?).
Geography Achievement (A5)	Range of geographic knowledge.	An examinee must identify capitals of countries around the world.
Communication Ability (CM)	Ability to speak in "real life" situations (e.g., lecture, group participation) in an adult-like manner.	An examinee is required to view a picture of a small town with stores and streets and must describe the scene and give directions from one store in the picture to another.
Oral Production and Fluency (OP)	More specific or narrow oral communication skills than reflected by Communication Ability (CM).	An examinee is presented with a starting stimulus word and must use the word properly in a sentence.
Grammatical Sensitivity (MY)	Knowledge or awareness of the grammatical features of the native language.	An examinee must correctly label the parts of speech contained in a sentence or correct those parts of speech that are utilized incorrectly (e.g., disparate tenses in a sentence).

(continued)

TABLE 2.1 Continued

	Gf-Gc Broad Stratum II Ability	
Narrow stratum I name (code)	**Definition**	**Task Example**
Foreign Language Proficiency (KL)	Similar to Language Development (LD) but for a foreign language.	An examinee is presented with two words in a foreign language and must describe the common relation or similarity between them.
Foreign Language Aptitude (LA)	Rate and ease of learning a new language.	An examinee is presented with several translated words that are paired with pictorial stimuli and must pair the words with the pictures following a single presentation.
	Reading/Writing (Grw)	
Reading Decoding (RD)	Ability to recognize and decode words or pseudowords in reading.	An examinee is required to accurately pronounce a list of nonsense words.
Reading Comprehension (RC)	Ability to comprehend connected discourse during reading.	An examinee is required to read a short passage and respond to questions about the passage.
Verbal (printed) Language Comprehension (V)	General development, or the understanding of words, sentences, and paragraphs in native language, as measured by *reading* vocabulary and *reading* comprehension tests.	An examinee must read a list of four vocabulary words and choose two of the four words that belong together in some meaningful way.
Cloze Ability (CZ)	Ability to supply words deleted from prose passages that must be read.	An examinee is required to read a short passage and supply a missing word that best corresponds to the theme or content of the passage.
Spelling Ability (SG)	Ability to spell.	An examinee must spell a series of increasingly difficult orally presented words.
Writing Ability (WA)	Ability to write with clarity of thought, organization, and good sentence structure.	An examinee is given a starting stimulus and must write a well-organized story that adheres to the structural rules of writing (e.g., examinee starts a new paragraph when he or she presents a new idea).

English Usage Knowledge (EU)	Knowledge of writing in the English language with respect to capitalization, punctuation, usage, and spelling.	An examinee must correct sentences with respect to capitalization, punctuation, spelling, and usage errors.
Reading Speed (RS)	Time required to silently read a passage or series of sentences as quickly as possible.	An examinee is asked to silently read a passage for one minute. Reading speed reflects words per minute read.
	Short-Term Memory (Gsm)	
Memory Span (MS)	Ability to attend to and immediately recall temporally ordered elements in the correct order after a single presentation.	An examinee is presented with a series of numbers or words and must repeat them orally in the same sequence as presented.
Working Memory (MW)	Ability to temporarily store and perform a set of cognitive operations on information that requires divided attention and the management of the limited capacity of short-term memory.	An examinee is presented a series of numbers and words in a mixed-up order and is then required to reorder and say the complete list of numbers first in order followed by the words in order.
Learning Abilities (L1)	A number of factors that are specific to particular kinds of learning situations and memory [Also listed under *Glr*]. (Not clearly defined by existing research.)	An examinee must learn paired-associate material to a criterion during a study phase that is followed by an intervening task and finally a "relearning" testing phase.
	Visual Processing (Gv)	
Spatial Relations (SR)	Ability to rapidly perceive and manipulate relatively simple visual patterns or to maintain orientation with respect to objects in space.	An examinee is required to view a stimulus pattern or design and reproduce the design using blocks or cubes.
Visual Memory (MV)	Ability to form and store a mental representation or image of a visual stimulus and then recognize or recall it later.	An examinee is required to reproduce or recognize a previously presented visual stimulus that has been removed.
Visualization (Vz)	Ability to mentally manipulate objects or visual patterns and to "see" how they would appear under altered conditions.	The examinee is presented with a visual image and must draw how the image would look upside down.
Closure Speed (CS)	Ability to quickly combine disconnected, vague, or partially obscured visual stimuli or patterns into a meaningful whole, *without knowing in advance* what the pattern is.	An examinee is required to identify an object from a line drawing that has portions of the lines missing.

(continued)

TABLE 2.1 Continued

	Gf-Gc Broad Stratum II Ability	
Narrow stratum I name (code)	**Definition**	**Task Example**
Flexibility of Closure (CF)	Ability to find, apprehend, and identify a visual figure or pattern embedded in a complex visual array, *when knowing in advance* what the pattern is.	An examinee must identify ten animals that are embedded in a complex visual scene.
Spatial Scanning (SS)	Ability to accurately and quickly survey a spatial field or pattern and identify a path through the visual field or pattern.	An examinee is required to complete a series of increasingly difficult mazes within a specified time period.
Serial Perceptual Integration (PI)	Ability to apprehend and identify a pictorial or visual pattern when parts of the pattern are presented rapidly in serial or successive order.	An examinee is required to correctly identify or name a stimulus when portions of the stimuli are presented serially (e.g., portions of a line drawing of a cat are passed through a small "window").
Length Estimation (LE)	Ability to accurately estimate or compare visual lengths and distances without using measurement instruments.	An examinee is presented with a series of paired double-arrow lines of differing orientations and determine whether they are the same length or different.
Perceptual Illusions (IL)	Ability to resist being affected by perceptual illusions involving geometric figures.	An examinee is presented with pictures of geometric shapes that have superimposed patterns and must correctly identify the dominant geometric shape present in the picture.
Perceptual Alternations (PN)	Consistency in the rate of alternating between different visual perceptions.	An examinee is asked to view a series of flashing bars with one constant bar in the middle and must indicate whether or not the bars are going behind the middle bar or are flashing simultaneously at its side.
Imagery (IM)	Ability to vividly mentally manipulate abstract spatial forms. (Not clearly defined by existing research.)	An examinee is given a starting stimulus (e.g., a square) and must follow a series of verbal transformations to determine the resultant stimuli (e.g., a triangle)

	Auditory Processing (Ga)	
Phonetic Coding: Analysis (PC:A)	Ability to segment larger units of speech sounds into smaller units of speech sounds.	An examinee is presented with the pronunciation of a word and must identify the beginning and ending sounds.
Phonetic Coding: Synthesis (PC:S)	Ability to blend smaller units of speech together into larger units of speech.	An examinee is presented with the isolated sounds for a word and must blend the sounds together and identify the word.
Speech Sound Discrimination (US)	Ability to detect differences in speech sounds under conditions of little distraction or distortion.	An examinee is presented with a series of tape-recorded phonetically nonmeaningful sounds and must identify whether the sounds are the same or different.
Resistance to Auditory Stimulus Distortion (UR)	Ability to understand speech and language that has been distorted or masked in one or more ways.	An examinee must identify monosyllabic and multisyllabic words while listening to an increasing level of noise presented through earphones.
Memory for Sound Patterns (UM)	Ability to retain on a short-term basis auditory events such as tones, tonal patterns, and voices.	An examinee is presented with a series of tone patterns and later must identify whether subsequently presented patterns were among those originally heard.
General Sound Discrimination (U3)	Ability to discriminate tones, tone patterns, or musical materials with regard to pitch, intensity, duration, and rhythm.	An examinee is presented with two short musical patterns and must identify whether the patterns are similar or different and, if different, how they differ (e.g., by duration, intensity)
Temporal Tracking (UK)	Ability to track auditory temporal events so as to be able to count, rearrange, or anticipate them.	An examinee is presented with a steady pattern of musical beats and must identify the note that is to come next after the music has stopped.
Musical Discrimination and Judgment (U1, U9)	Ability to discriminate and judge tonal patterns in music with respect to melodic, harmonic, and expressive aspects (e.g., phrasing, tempo, and intensity variations).	An examinee is presented with tape-recorded samples of musical pieces from different musical genres presented in either major or minor keys and must describe the differences in the music in terms of its harmonic complexity and mood.

(continued)

TABLE 2.1 **Continued**

Narrow stratum I name (code)	Definition	Task Example
	Gf-Gc **Broad Stratum II Ability**	
Maintaining and Judging Rhythm (U8)	Ability to recognize and maintain a musical or equal time beat.	An examinee is presented with a tape-recorded metronome keeping 4/4 time (i.e., one measure) that is comprised of quarter notes and must demonstrate knowledge of equal time beat by tapping out eighth notes.
Sound-Intensity/Duration Discrimination (U6)	Ability to discriminate sound intensities and to be sensitive to the temporal/rhythmic aspects of tonal patterns.	An examinee must listen to a series of tape-recorded sounds and indicate by raising their hand when one sound becomes more intense than the previously presented sound.
Sound-Frequency Discrimination (U5)	Ability to discriminate frequency attributes (pitch and timbre) of tones.	An examinee is presented with random tape-recorded notes played on the high, middle, and low ends of a piano keyboard and must describe the relationship of the second note played to the first note played (e.g., higher, lower).
Hearing and Speech Threshold Factors (UA, UT, UU)	Ability to hear pitch and varying sounds over a range of audible frequencies.	An examinee is presented with a series of 15 tape-recorded sounds and must indicate by writing a check mark in a response booklet whenever they hear a sound.
Absolute Pitch (UP)	Ability to perfectly name or identify the pitch of tones.	An examinee is presented with a tape-recorded note of a piano key (e.g., C or F sharp) and is required to name the note.
Sound Localization (UL)	Ability to localize heard sounds in space.	An examinee is presented with earphones and must indicate whether a presented sound was heard in the left, right, or both sides of the headset.
	Long-Term Storage and Retrieval (Glr)	
Associative Memory (MA)	Ability to recall one part of a previously learned but unrelated pair of items when the other part is presented (i.e., paired-associative learning).	An examinee is presented with a set of visual stimuli paired with nonsense words and must correctly identify the nonsense word that had been presented with a certain visual stimulus.

Meaningful Memory (MM)	Ability to recall a set of items where there is a meaningful relation between items or the items comprise a meaningful story or connected discourse.	An examinee is presented with a short story and must retell the story as accurately as possible immediately following a single presentation.
Free Recall Memory (M6)	Ability to recall as many unrelated items as possible, in any order, after a large collection of items is presented.	An examinee is presented with a series of objects and, after they are removed, must recall the objects in any order.
Ideational Fluency (FI)	Ability to rapidly produce a series of ideas, words, or phrases related to a specific condition or object. Quantity not quality is emphasized.	An examinee must rapidly name as many square objects as he can within a specified time limit.
Associational Fluency (FA)	Ability to rapidly produce words or phrases associated in meaning (semantically associated) with a given word or concept.	An examinee must name as many examples of objects that fit into a specified category (e.g., name as many fruits as you can think of) within a specified time limit.
Expressional Fluency (FE)	Ability to rapidly think of and organize words or phrases into meaningful complex ideas under high general or more specific cueing conditions.	An examinee must rapidly name a category that bests represents a series of presented words (e.g., Pattern, material, thread . . . things to make clothing.)
Naming Facility (NA)	Ability to rapidly produce names for concepts when presented with a pictorial or verbal cue.	An examinee must rapidly provide the general name of a category when shown specific pictorial stimuli (e.g., a picture of an apple, shirt, and bus, would require the reply: fruit, clothing, transportation)
Word Fluency (FW)	Ability to rapidly produce words that have specific phonemic, structural, or orthographic characteristics (independent of word meanings).	An examinee must name as many words as he can think of that start with the "sh" sound within a specified time limit.
Figural Fluency (FF)	Ability to rapidly draw or sketch several examples or elaborations when given a starting visual or descriptive stimulus.	An examinee must draw as many things as he can when presented with a nonmeaningful starting visual stimulus.
Figural Flexibility (FX)	Ability to quickly change set in order to generate new and different solutions to figural problems.	An examinee is presented with five geometric shapes and must manipulate those shapes to create objects described by the examiner (e.g., build a house, build a car).

(continued)

TABLE 2.1 Continued

	Gf-Gc Broad Stratum II Ability	
Narrow stratum I name (code)	**Definition**	**Task Example**
Sensitivity to Problems (SP)	Ability to identify and state practical problems in a given situation or rapidly think of and state various solutions to, or consequences of, such problems.	An examinee is required to answer questions such as "What is the thing to do if you lock your keys in the car?"
Originality/Creativity (FO)	Ability to rapidly produce original, clever, or uncommon verbal or ideational responses to specified tasks.	An examinee is given a starting stimulus word such as "car" and must construct as many words as he can using those three letters in the word (e.g., carrot, care, carton, cartoon, racecar, ma<u>car</u>oon)
Learning Abilities (L1)	A number of factors that are specific to particular kinds of learning situations and memory [Also listed under *Gsm*]. (Not clearly defined by existing research).	An examinee must learn paired-associate material to a criterion during a study phase that is followed by an intervening task and finally a "relearning" testing phase.
	Processing Speed (Gs)	
Perceptual Speed (P)	Ability to rapidly search for and compare known visual symbols or patterns presented side by side or separated in a visual field.	An examinee must rapidly view rows of stimuli and cross out those stimuli that are similar within the presented row within a specified time limit.
Rate of Test Taking (R9)	Ability to rapidly perform tests which are relatively easy or that require very simple decisions.	An examinee is required to pair numbers with symbols according to a presented key as rapidly as possible.
Number Facility (N)	Ability to rapidly and accurately manipulate and deal with numbers, from elementary skills of counting and recognizing numbers to advanced skills of adding, subtracting, multiplying, and dividing numbers.	An examinee is required to complete a series of arithmetic problems using paper and pencil in a specified time limit.

	Decision/Reaction Time or Speed (Gt)	
Simple Reaction Time (R1)	Reaction time to the presentation of a single visual or auditory stimulus.	An examinee is required to quickly depress a computer key when presented with a specific geometric figure (e.g., a square) that appears intermittently in a series of other figures on the computer screen.
Choice Reaction Time (R2)	Reaction time to one of two or more alternative stimuli, depending on which alternative is signaled.	An examinee is required to rapidly depress one of two computer keys depending on the type of stimulus presented (e.g., press green key when square is presented and red key when a circle is presented).
Semantic Processing Speed (R4)	Reaction time when the decision requires some encoding and mental manipulation of stimulus content.	An examinee is required to rapidly depress a key when the stimulus viewed is opposite to a stimulus description provided (e.g., the examinee would depress a key when he/she viewed a star beneath a cross after initially viewing the star above the cross).
Mental Comparison Speed (R7)	Reaction time where the stimuli must be compared for a particular attribute.	An examinee is required to rapidly depress a key when presented with two identical geometric shapes and refrain from pressing when the shapes differ.

Note: Most all definitions were derived from Carroll (1993a). Two-letter factor codes (e.g., RG) are from Carroll (1993a). Information in this table was adapted from McGrew (1997) with permission from Guilford Press.

as the "mind's scratchpad" (Jensen, 1998, p. 220) and most models of working memory postulate a number of subsystems or temporary *buffers*. The phonological or articulatory loop processes auditory-linguistic information while the visuospatial sketch/scratchpad (Baddeley, 1986; 1992; Logie, 1996) is the temporary buffer for visually processed information. Most working memory models posit a central executive or processor mechanism that coordinates and manages the activities and subsystems in working memory.

Carroll (1993a) is skeptical of the working memory construct as reflected in his conclusion that "although some evidence supports such a speculation, one must be cautious in accepting it because as yet there has not been sufficient work on measuring working memory, and the validity and generality of the concept have not yet been well established in the individual differences research" (p. 647). Respecting the judgement of one of the primary architects of the *Gf-Gc* model (i.e., Carroll), we have chosen not to elevate working memory to the status of a broad stratum II ability. Instead, we feel that at this time, current knowledge argues for the classification of working memory as a narrow ability under *Gsm*. This makes logical sense given that working memory shares a number of cognitive processes with Memory Span (MS) yet also includes additional processes that differentiate the two abilities. Given that Carroll includes Learning Abilities under his *Gsm* factor (which he calls *Gy*), it is clear that the *Gsm* portion of the *Gf-Gc* framework requires additional research to elucidate the relations between the various narrow memory abilities that currently comprise *Gsm*.

To allow working memory "membership" in the *Gf-Gc* taxonomy, we suggest herein that the *Gf-Gc* taxonomic code system (e.g., MS = Memory Span; see Table 2.1) be expanded to include *MW* for working memory. Given that Carroll questions the validity of the working memory construct, we propose this MW code primarily for practical use and ease of communication. Additional research is necessary before a consensus can be reached about the inclusion (or exclusion) of working memory in the *Gf-Gc* model of intelligence. Practitioners need to recognize the somewhat tenuous nature of the *Gsm* domain when using the cross-battery assessment approach outlined in subsequent chapters.

Visual Processing (Gv). *Visual Processing (Gv)* is the ability to generate, perceive, analyze, synthesize, store, retrieve, manipulate, transform, and think with visual patterns and stimuli (Lohman, 1994). These abilities are measured frequently by tasks that require the perception and manipulation of visual shapes and forms, usually of a figural or geometric nature (e.g., the Wechsler Block Design and Object Assembly tests). An individual who can effectively mentally reverse and rotate objects, interpret how objects change as they move through space, perceive and manipulate spatial configurations, and maintain spatial orientation would be regarded as having a strength in *Gv* abilities. The various narrow abilities subsumed by *Gv* are listed and defined in Table 2.1.

Auditory Processing (Ga). In the broadest sense, auditory abilities "are cognitive abilities that depend on sound as input and on the functioning of our hearing apparatus" (Stankov, 1994, p. 157) and reflect "the degree to which the individual can cognitively control the perception of auditory stimulus inputs" (Gustafsson & Undheim, 1996, p. 192). *Auditory Processing* is the ability to perceive, analyze, and synthesize patterns among auditory stimuli, and discriminate subtle nuances in patterns of sound (e.g., complex musical structure) and speech when presented under distorted conditions. While *Ga* abilities do not require the comprehension of language (*Gc*), they may be very important in the development of language skills. *Ga* subsumes most of those abilities referred to as phonological awareness/processing. However, as can be seen from the list of narrow abilities subsumed by *Ga* (Table 2.1), this domain is very broad.

A change from McGrew and Flanagan's (1998) discussion of *Ga* is the splitting of Carroll's Phonetic Coding (PC) narrow ability into separate analysis (PC:A) and synthesis (PC:S) abilities. Support for two different PC abilities comes primarily from four sources. First, in a sample of kindergarten students, Yopp (1988) reported evidence in favor of two phonemic awareness factors: simple phonemic awareness (required one operation to be performed on sounds) and compound phonemic awareness (required holding sounds in memory while performing another operation on them). Second, in what appears to be the most comprehensive *Ga* factor-analytic study to date, Stankov and Horn (1980) presented evidence for seven different auditory abilities, two of which had tests of sound blending (synthesis) and incomplete words (analysis) as factor markers. Third, the WJ-R Sound Blending and Incomplete Words tests (which are almost identical in format to the tests used by Stankov and Horn) correlated only moderately (0.37 or 13.7% shared or common variance) across the kindergarten to adult WJ-R norm sample, a correlation that suggests that these tests are measuring different aspects of PC. Finally, using confirmatory factor-analytic methods, Wagner, Torgesen, Laughton, Simmons, and Rashotte (1993) presented a model of phonological processing that included separate auditory analysis and synthesis factors.

Although the features of these different auditory factors across respective studies are not entirely consistent, there are a number of similarities. For example, Yopp's (1988) simple phonemic factor appears to be analogous to Wagner and colleagues' (1993) synthesis factor and the factor Stankov and Horn (1980) identified with the aid of sound blending tasks. Also, Yopp's (1988) compound phonemic factor bears similarities to Wagner and colleagues' (1993) analysis factor and the Stankov and Horn (1980) factor, identified in part by an incomplete words task. At this time, we conclude that Wagner and colleagues' (1993) analysis/synthesis distinction is probably the most useful. According to Wagner and colleagues (1993), analysis and synthesis can be defined as "the ability to segment larger units of speech into smaller units" and "the ability to blend smaller units of speech to form larger units" (p. 87), respectively. As a result, in the current model (Figure 2.4) we proposed that PC be split into separate PC:A and PC:S narrow abilities.

Long-Term Storage and Retrieval (Glr). *Long-Term Storage and Retrieval* is the ability to store information in and fluently retrieve new or previously acquired information (e.g., concepts, ideas, items, names) from long-term memory. *Glr* abilities have been prominent in creativity research where they have been referred to as idea production, ideational fluency, or associative fluency. It is important to not confuse *Glr* with *Gc, Gq,* and *Grw,* a person's stores of acquired knowledge. *Gc*, *Gq*, and *Grw* represent *what* is stored in long-term memory, while *Glr* is the *efficiency* by which this information is initially stored in and later retrieved from long-term memory. Using the *Gc* fishing net analogy discussed earlier in this chapter (where the nodes and links of the net represent the knowledge that is stored in long-term memory), *Glr* is the process by which individuals efficiently add new nodes and links to their "fishing net" of stored knowledge and then later retrieve information from their net of knowledge.

Different processes are involved in *Glr* and *Gsm*. Although the word *long-term* frequently carries with it the connotation of days, weeks, months, and years in the clinical literature, long-term storage processes can begin within a few minutes or hours of performing a task. Therefore, the time lapse between the initial task performance and the recall of information related to that task is not of critical importance in defining *Glr*. More important is the occurrence of an intervening task that engages short-term memory during the interim before the attempted recall of the stored information (e.g., *Gc;* Woodcock, 1994). In the present *Gf-Gc* model, 13 narrow memory and fluency abilities are included under *Glr* (see Table 2.1).

Processing Speed (Gs). *Processing Speed* or mental quickness is often mentioned when talking about intelligent behavior (Nettelbeck, 1994). Processing speed is the ability to fluently and automatically perform cognitive tasks, especially when under pressure to maintain focused attention and concentration. *Attentive speediness* encapsulates the essence of *Gs*. *Gs* is measured typically by fixed-interval timed tasks that require little in the way of complex thinking or mental processing (e.g., the Wechsler Animal Pegs, Symbol Search, and Digit Symbol/Coding tests).

Recent interest in information processing models of cognitive functioning has resulted in a renewed focus on *Gs* (Kail, 1991; Lohman, 1989). A central construct in information processing models is that of limited processing resources (e.g., the limited capacities of short-term or working memory). That is, "many cognitive activities require a person's deliberate efforts and that people are limited in the amount of effort they can allocate. In the face of limited processing resources, the speed of processing is critical because it determines in part how rapidly limited resources can be reallocated to other cognitive tasks" (Kail, 1991, p. 152). Woodcock (1994) likens *Gs* to a valve in a water pipe. The rate at which water flows in the pipe (i.e., *Gs*) increases when the valve is wide open and it decreases when the valve is partially closed. Three different narrow speed of processing abilities are subsumed by *Gs* in the present *Gf-Gc* model (see Table 2.1).

Decision/Reaction Time or Speed (Gt). In addition to *Gs*, both Carroll and Horn include a second broad speed ability in their respective *Gf-Gc* models. Processing Speed (Decision/Reaction Time or Speed; *Gt*), the ability proposed by Carroll, subsumes narrow abilities that reflect an individual's quickness in reacting (reaction time) and making decisions (decision speed). Correct Decision Speed (CDS), what Horn proposed as the second speed ability (*Gs* being the first) is typically measured by recording the time an individual requires to provide an answer to problems on a variety of tests (e.g., letter series, classifications, vocabulary; Horn, 1988; 1991). Because CDS appears to be a much narrower ability than *Gt*, it is subsumed by *Gt* in the *Gf-Gc* model used in this book.

Gt should not be confused with *Gs*. *Gt* abilities reflect the immediacy with which an individual can react (typically measured in seconds or parts of seconds) to stimuli or a task, while *Gs* abilities reflect the ability to work quickly over a longer period of time (typically measured in intervals of 2 to 3 minutes). Being asked to read a passage (on a self-paced scrolling video screen) as quickly as possible and, in the process, touch the word *the* with a stylus pen each time it appears on the screen, is an example of *Gs*. The individual's *Gs* score would reflect the number of correct responses (taking into account errors of omission and commission). In contrast, *Gt* may be measured by requiring a person to read the same text at their normal rate of reading and press the space bar as quickly as possible whenever a light is flashed on the screen. In this latter paradigm, the individual's score is based on the average response latency or the time interval between the onset of the stimulus and the individual's response.

Recent *Gf-Gc* Theory Developments

Research and development regarding the *Gf-Gc* theory is not static. Two general lines of research that are germane to the current work are providing insights into potential future refinements and extensions of the *Gf-Gc* theory.

Refining the Structural Taxonomy

Recent research (Roberts & Stankov, 1998; Roberts, Stankov, Pallier, & Dolph, 1997) has suggested possible modifications to the *Gf-Gc* model of cognitive abilities. Roberts and associates (1997) presented evidence that suggests that tactile-kinesthetic (TK) abilities may represent either a narrow stratum I ability influenced by both *Gv* and *Gf,* or a broad stratum II ability closely related to *Gv* and *Gf.* According to Roberts and associates, the TK factor consists of a complex set of abilities that are related to *Gf* and may involve tactile, visual, and working memory processes.

TK measures are not included in any of the major intelligence batteries. However, instruments such as the Wechsler Scales are often supplemented with a variety of TK tests (e.g., examinees must identify numbers written on their fingertips; examinees must place geometrically shaped blocks into a form board while blindfolded). This is a common practice in neuropsychological assessment. For example, the recently published NEPSY (Korkman, Kirk, & Kemp, 1997) includes a test that requires an examinee to identify the finger or fingers the examiner touches (Finger Discrimination). Further research in the TK domain is important given the salient role these abilities have played in neuropsychological assessment and aging research (i.e., cognitive decline due to *decrements* in sensory processes).

The other major *Gf-Gc* domain that has been the subject of active exploration recently is cognitive speed (*Gs* and *Gt*). Roberts and Stankov (1998), for example, have presented evidence for a four-level hierarchical structure of cognitive speed in which they identify speed factors at the narrow (stratum I) and broad (stratum II) levels, and suggest the presence of intermediate levels between and below these strata. Roberts and Stankov identified a single *broad* speed ability (*Gt*—Cognitive Speed) and five *narrow* speed abilities (viz., Clerical/Perceptual Speed, Induction Speed, Visual/Auditory [Perceptual] Test-Taking Speed, General Decision Speed, and Movement Time). In addition, these researchers posited a Psychometric Speed ability "in limbo" between the narrow and broad stratum speed abilities. Finally, below the five narrow (stratum I) abilities exists the possibility of three more specific decision speed and movement time abilities, respectively. Currently, the Wechsler Intelligence Scales and other intelligence batteries only include tests that tap the Clerical/Perceptual *(Gs)* portion of the Roberts and Stankov model (e.g., WISC-III Coding).

The addition of a TK ability and a more complex and specific cognitive speed hierarchy to the *Gf-Gc* model has primarily theoretical implications at this time. Additional research is needed to determine the replicability of these findings in additional samples of varying ages and, more importantly, to determine their practical implications for psychological assessment.

Gf-Gc and Information Processing Theory: Pushing the Edge of the Theory Envelope

As mentioned earlier in this chapter (see Figure 2.1 and related discussion), a number of researchers have proposed theoretical models that describe and explain cognitive performance as a composition of both cognitive and noncognitive variables within an information processing framework. Most prominent within the context of *Gf-Gc* theory are Woodcock's (1993; 1997; 1998) *Gf-Gc Cognitive Performance Model (CPM)* and *Information Processing Model (IPM)*. According to Woodcock (1998, p. 143) these models are not intended to extend *Gf-Gc* theory, but rather, to "nudge current theory further into clinical and research practice." The basis of the CPM model is that a person's cognitive performance is a complex interaction of many different components that can be differentiated by function. The most recent revision of the CPM is presented schematically in Figure 2.5.

Briefly, *Acquired Knowledge,* represented by the *Gc, Gq,* and *Grw* abilities, includes knowledge stores of factual (declarative) and automatized cognitive processes and procedures (procedural knowledge) within a domain. For example, being able to identify Σ as the statistical summation symbol is a type of *Gq* declarative knowledge. Knowing how *to do* the summation process is a form of *Gq* procedural knowledge. With the exception of the Digit Span test, the entire Wechsler Verbal scale is comprised of acquired knowledge indicators (viz., *Gc* and *Gq*).

Thinking Abilities (viz., *Gv, Ga, Glr, Gf*) are involved in the cognitive processing of information that is placed in short-term memory *(Gsm)* but cannot be processed automatically. These abilities are often what many professionals think of when talking about intelligence, because they are involved in new learning or performing tasks that an individual cannot complete or solve automatically. All of the Wechsler performance tests (except Symbol Search, Animal Pegs, and Digit-Symbol/Coding) are measures of *Gv* thinking abilities.

Cognitive Efficiency includes abilities that influence the speed *(Gs)* or automaticity of cognitive functioning through the efficient allocation of mental resources within the limited capacity short-term and working memory systems *(Gsm).* Using the *Gq* example described above, two individuals who have the same overall level and pattern of thinking abilities and stores of acquired *Gq* knowledge may vary in the speed and accuracy by which they can mentally sum a series of five numbers, due to differences in overall speed of cognitive processing, automaticity of mathematical facts, and ability to hold information in working memory via the use of cognitive strategies. The various Wechsler processing speed tests (i.e., Symbol Search, Animal Pegs, Digit Symbol/Coding) are considered measures of one aspect of cognitive efficiency (i.e., *Gs*) while the short-term and working memory tests (i.e., Digit Span, Sentences, Letter-Number Sequencing) are classified as *Gsm.*

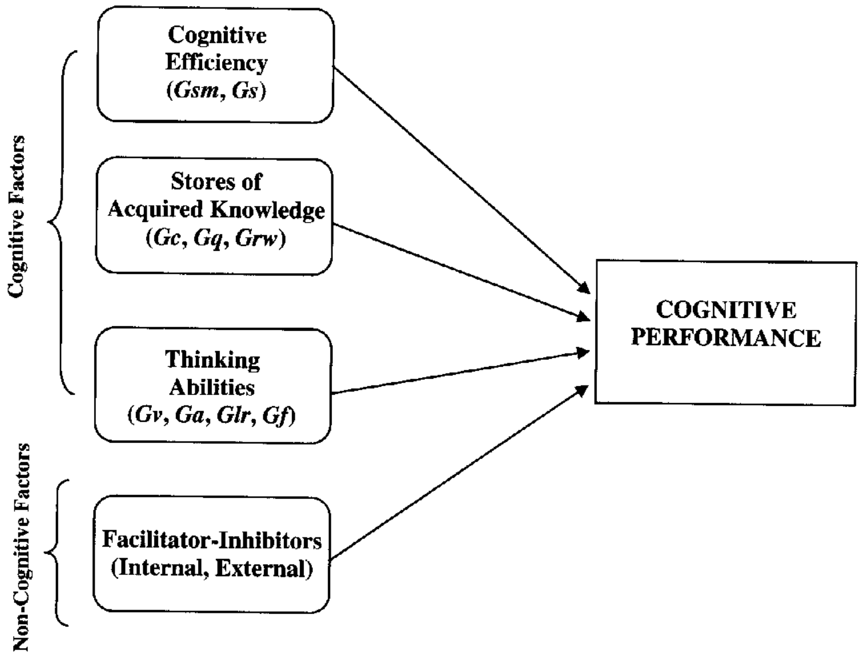

FIGURE 2.5 A Schematic Representation of Woodcock's Cognitive Performance Model

Facilitators-Inhibitors are internal (e.g., health, attention and concentration, cognitive style, emotional state) and external (e.g., distracting stimuli) variables that can "modify cognitive performance for better or for worse, often overriding the effects of strengths and weaknesses in the previously described cognitive abilities" (Woodcock, 1998, p. 146). For example, two individuals with similar overall levels and patterns of thinking abilities and cognitive efficiency may differ in their ability to sum the previously discussed five numbers, due to differences in concentration, intrinsic motivation, or state of health at the time of performance. The collective participation of these four functional cognitive and noncognitive components aids in explaining intraindividual and interindividual variation in the performance of complex cognitive activities.

Woodcock (1993; 1998) extended the CPM into a more complex *Gf-Gc Information Processing Model (IPM)* that integrates *Gf-Gc* abilities and other aspects of cognition within an information processing framework. Woodcock's grounding of the CPM and IPM in the Cattell-Horn *Gf-Gc* theory is consistent with Taylor's (1994) conclusion that a "positive feature of the Cattell model is that it is amenable to dynamic, learning, or developmental interpretations" (p. 185). The IPM is admittedly complex and cannot be described in sufficient detail in this book. The interested reader is referred to Woodcock (1997; 1998) for a discussion of the model and a description of a practical *Gf-Gc Diagnostic Worksheet* that could be applied to the Wechsler Scales when they are supplemented via the Wechsler-based *Gf-Gc* cross-battery approach described in Part III of this book.

Another notable effort along these lines is Snow's (1989) attempt to use new forms of psychometric theory to measure information processing constructs such as declarative knowledge acquisition, proceduralization, and automatization (Anderson, 1985). Snow's work was not specifically designed to extend *Gf-Gc* theory, but it shares the same general goal of integrating cognitive and noncognitive constructs to better understand cognitive performance.

Briefly, Snow suggested that cognitive constructs need to be combined with affective and conative (i.e., noncognitive) constructs to fully measure an individual's aptitudes for learning (Snow, Corno, & Jackson, 1996). According to Snow (1989), learners approach tasks with previously developed *conceptual structures* and *procedural skills* (i.e., initial states) that subsume *Gf-Gc*-type cognitive constructs (e.g., *Gc, Gv, Gsm*). In addition, conative personal characteristics in the broad domains of *learning strategies, self-regulatory functions,* and *motivational orientation* are viewed as interacting with the cognitive constructs during cognitive performance. The important common feature of Woodcock and Snow's attempts to explain cognitive performance is that both dynamic models include cognitive and noncognitive constructs.

Although the Wechsler Intelligence Scales (and for that matter all major intelligence batteries) do not directly measure noncognitive constructs, David Wechsler was at the forefront in highlighting the important link between cognitive and noncognitive factors when trying to explain real-world intelligent behavior. Throughout his career, Wechsler was interested in how a variety of *nonintellectual factors* (e.g., persistence, curiosity, and motivation) influenced the expression of intelligent behavior (Zachary, 1990). In 1943 Wechsler stated:

> When our scales measure the nonintellective as well as intellectual factors in intelligence, they will more nearly measure what in actual life corresponds to intelligent behavior. (Wechsler, 1944, p. 103)

Even Spearman, who is almost exclusively associated with intelligence and cognition, recognized the importance of nonintellectual factors. In his seminal book "The Abilities of Man," Spearman (1927) stated that: "the process of cognition cannot possibly be treated apart from those of conation and affection, seeing that all these are but inseparable aspects in the instincts and behavior of a single individual, who himself, as the very name implies, is essentially indivisible" (p. 2). The work of Woodcock and Snow suggests that current theory and research are catching up with the pioneering ideas of David Wechsler and Charles Spearman.

Gardner, Sternberg, and Guilford's Theories: Relevance to Contemporary Psychometric Theories

In addition to the increased interest in the psychometric *Gf-Gc* theory of intelligence, there has been considerable attention (particularly in the popular press) in Gardner's theory of *Multiple Intelligences* (Chen & Gardner, 1997; Gardner, 1983; 1993; 1994) and Sternberg's *Triarchic* theory of intelligence (Sternberg, 1994; 1997). Guilford's *Structure-of-*

Intellect model, although now more of historical interest, also continues to receive prominent treatment in most books or book chapters on intelligence. Given the movement toward theory-based intelligence test interpretation, the utility of these three theories for practical test development and interpretation warrant comment. These three theories are briefly described here with a particular emphasis on their relationship to contemporary *Gf-Gc* theory.

Gardner's Theory of Multiple Intelligences

The description of *Gf-Gc* theory as a multiple intelligences theory often causes confusion when individuals try to reconcile this model with Gardner's multiple intelligences (MI) theory (Chen & Gardner, 1997; Gardner, 1983; 1993; 1994). Although Gardner's MI theory has yet to serve as the foundation for an individually administered norm-referenced battery of tests, the concepts have received considerable attention in the popular press.

Gardner (1983) originally proposed seven intelligences which included (together with a brief example): (1) linguistic—used in reading or writing of poetry; (2) logical-mathematical—used by scientists or mathematicians to solve problems; (3) spatial—used by architects to visualize and draw building plans; (4) musical—used by musicians to compose songs; (5) bodily-kinesthetic—used by athletes or dancers when performing or competing; (6) interpersonal—used by therapists to understand and interact with clients; and (7) intrapersonal—used by individuals to gain insight about themselves. Recently, Gardner (1998) has added an eighth intelligence dealing with the ability to discern patterns in nature (i.e., the *naturalist*) (Meyer, 1997). He also suggested the possibility of two additional types of intelligence (viz., *spiritual* and *existential*), although the evidence he presented for the latter two is preliminary in nature. The terms Gardner uses to label his eight intelligences are dramatically different from the terminology of *Gf-Gc* theory. What are the differences and similarities between the *Gf-Gc* and Gardner multiple intelligences theories?

Sternberg (1997) suggested that Gardner's theory (and his own triarchic theory of intelligence) differs from traditional psychometric theories in that it specifies a "system of interacting abilities rather than just specifying a set of abilities" (p. 1134). McGrew (1993; 1995) suggested that the fundamental differences between the two theories is that *Gf-Gc* theory is concerned with *describing the basic domains or building blocks of intelligent behavior* in the cognitive domain, while Gardner's theory focuses on *how these different domains or building blocks are combined,* along with other personal competencies (e.g., motor and social skills), in patterns representing different forms of aptitude or expertise (i.e., adult end-states valued by a culture) (Chen & Gardner, 1997).

Using Greenspan's model of personal competence (Greenspan & Driscoll, 1997), a model that includes the broad domains of physical and emotional competence and conceptual, practical, and social intelligence as an overarching framework, McGrew (1994) suggested that Gardner's seven intelligences represent unique combinations or patterns of human cognitive abilities that, together with other personal competencies, help to explain, understand, or predict aptitude, expertise, or talent. For example, Gardner's logical-mathematical intelligence reflects a sensitivity to and capacity for processing logical and/or numerical patterns and the ability to manage long sequences or chains of reasoning. Scientists and mathematicians would most likely be high on logical-mathematical intelligence. An individual who has high logical-mathematical intelligence may have high fluid intelligence, quantitative knowledge and reasoning, and visual-spatial abilities, abilities that are central in contemporary *Gf-Gc* theory. It is the specific combination of *Gf-Gc* strengths that a person exhibits that defines him or her as being high in logical-mathematical intelligence. Furthermore, individuals who are high in Gardner's bodily-kinesthetic intelligence, for example, may have specific *Gf-Gc* strengths (e.g., visual-spatial), *plus* strengths in other personal competence domains such as physical competence that help to explain their overall level of bodily-kinesthetic performance.

Thus, according to McGrew (1994), Gardner's theory is not an attempt to isolate the basic domains or elements of intelligence (a function performed by *Gf-Gc* theory), rather, it describes different patterns of expertise or aptitude based on specific combinations of *Gf-Gc* abilities *and other personal competencies*. In this regard, Gardner's different intelligences are conceptually similar to Snow's (1989; 1991; 1992) *aptitude complexes* which define aptitudes in the broadest sense (i.e., including both cognitive and conative structures).

Although Gardner's MI theory has considerable appeal, there has been no empirical evaluation of the validity of the theory as a whole (Sternberg & Kaufman, 1998), and the available empirical evaluations have been found wanting. In a review of Gardner's (1983) *Frames of Mind,* the book that first outlined his MI theory, Lubinski and Benbow (1995) concluded that there is "little empirical support for or against the unique features of Gardner's ideas. Before MI theory can be taken seriously by the scientific community and policy makers, Gardner's (1983) bold theoretical skeleton is in need of empirical flesh" (p. 937). According to Carroll (1993a), Gardner "discounts multifactorial theories of intelligence... because, he claims, they fail to account for the full diversity of abilities that can be observed. Generally, Gardner has neglected the evidence on the basis of which the present three-stratum theory has been constructed" (p. 641). Furthermore, in a review and comparison of structural *Gf-Gc* theory, Gardner's multiple intelligences theory, and Sternberg's Triarchic theory (Sternberg, 1985), Messick (1992) characterized Gardner's (as well as Sternberg's) theory as appealing selectively to factor-analytic research while ignoring or downplaying research that challenged his model.

It seems clear that the descriptions of Gardner's seven multiple intelligences "do not derive from any consistent set of empirical data and can be tied to data only in piecemeal fashion, thereby being constantly threatened by the perverse human tendency to highlight results that are consonant with the theory's logic over findings that are dissonant" (Messick, 1992; p. 368). Bouchard (1984), Gustafsson and Undheim (1996), Scarr (1985), and Snow (1985) also questioned the empirical support for Gardner's theory. Despite the lack of a strong program of validity research, Gardner's theory has produced many school-based educational interventions. According to Sternberg and Kaufman (1998), evidence to support these interventions is also limited.

Sternberg's Triarchic Theory

The Triarchic theory of intelligence (Sternberg, 1994; 1997) is an attempt to describe the processes that underlie intelligent thought by understanding the way in which intelligence relates to the internal and external world and experiences of individuals (Messick, 1992). Sternberg suggested that three major elements or *components* influence intelligent thought.

First, three sets of processing components or mental processes (i.e., knowledge acquisition, performance, metacomponents) allow individuals to solve problems (Sternberg, 1994; 1997). *Knowledge acquisition components* allow individuals to learn new information, while *performance* and *metacomponents* are involved in working with problems to produce solutions and executing and monitoring the problem-solving processes, respectively (Eggen & Kauchak, 1997). Second, *experiential components* are involved in relating new experiences and knowledge to old experiences and knowledge and recognizing and creating new patterns of information. Third, *contextual components* are concerned with adaptation—that is, explaining how a person's intelligence allows him or her to select new environments or adapt or modify existing environments (Eggen & Kauchak, 1997; Travers, Elliott, & Kratochwill, 1993).

According to Messick (1992), Sternberg's focus on culturally relevant conceptions of intelligence in relation to individual experiences results in a "focus on five critical aspects of intelligence—problem solving, verbal ability, social and practical competence, coping with novelty, and the automatization of performance" (p. 376). A consideration of these concepts suggests that the Triarchic theory includes parts of the *Gf-Gc* model, namely *Gf, Gc, Glr,* and *Gs,* respectively (Messick, 1992).

More recently, Sternberg (1996) proposed a theory of *successful intelligence* which "involves an individual's discerning his or her pattern of strengths and weaknesses, and the figuring out ways to capitalize upon the strengths and at the same time to compensate for weaknesses" (Sternberg & Kaufman, 1998, p. 494). Much like Gardner's theory of multiple intelligences, Sternberg's successful intelligence is comprised of abilities drawn from a broader array of personal competencies than those typically associated with traditional notions of intelligence. These include analytical, creative, and practical abilities. As is the case with much of Sternberg's theoretical concepts, analytical abilities are broadly defined (e.g., identifying a problem, defining the nature of the problem, devising a strategy to solve the problem, and monitoring the solution process). Analytical abilities more than likely include some that are *Gf-Gc* abilities and others that are yet to be determined. In contrast, creative and practical abilities are best thought of as including abilities from *other* competence domains (e.g., practical intelligence in Greenspan's model of personal competence).

To date, little validity evidence exists in support of Sternberg's theory of successful intelligence, due, in part, to its recency. However, Sternberg's Triarchic theory, like Gardner's MI theory, has not fared well when evaluated against established standards of validity. Messick (1992) indicated that "several aspects of Sternberg's theory are simply nonfactual . . . the theory is construct dense. . . . In the process, he [Sternberg] forgoes relations of strict deductibility and tends to rely on metaphorical descriptions . . . but they are not conducive to the derivation of empirical consequences instrumental to theory testing" (p. 379). Messick's less than positive treatment of Sternberg's Triarchic theory is echoed in Cronbach's (1986) response to some of Sternberg's claims:

> We don't have much theory, and I don't favor using the word loosely for almost any abstraction or point of view. . . . I would reserve the word *theory* for substantial, articulated, somewhat validated constructions. Rather than an emperor with no clothes, we have theory being used as an imperial cloak that has no emperor inside. (p. 23)

Guilford's Structure-of-Intellect Model

Anyone who has read any psychology book that covers the topic of intelligence has probably seen the "Rubik's cube" of intelligence theories. As a result of an ambitious and largely factor analytically based program of research, J. P. Guilford and associates proposed a three-dimensional *Structure-of-Intellect (SOI)* model (which has been presented typically as a three-dimensional cube) where all cognitive abilities were classified in terms of *operations* (cognitive, memory, divergent and convergent production, evaluation), *contents* (visual, auditory, symbolic, semantic, behavioral), and *products* (units, classes, relations, systems, transformations, implications). Based on the SOI model, understanding intelligence would require the classification and measurement of well over 100 discrete abilities. The SOI model probably represents the most detailed and comprehensive attempt to develop a precise and systematic taxonomy of cognitive abilities.

Although SOI interpretations of the Wechsler Intelligence Scales have been offered (Meeker, 1969; 1975), they no longer are used with much frequency due to both practical constraints (i.e., too time consuming and eclectic; Kaufman, 1994), and more importantly, a lack of supporting empirical evidence (Carroll, 1993a; Cronbach & Snow, 1977; Gustafsson & Undheim, 1996; Messick, 1992; Vernon, 1961). Carroll (1993a) rendered a particularly harsh judgement when he concluded that the SOI model "is fundamentally defective" (p. 59) and should be "marked down as a somewhat eccentric aberration in the history of intelligence models; that so much attention has been paid to it is disturbing, to the extent that textbooks and other treatments of it have given the impression that the model is valid and widely accepted, when clearly it is not" (p. 60).

Gardner, Sternberg, and Guilford: Concluding Comments

Currently, Gardner's MI, Sternberg's Triarchic, and Guilford's SOI models of intelligence are not likely to have significant impact on intelligence test interpretation. The SOI model had its day in the court of intelligence research and theory and was judged to be lacking in the validity evidence necessary for sound intelligence test interpretation. Although Gardner's and Sternberg's theories are receiving much attention in the popular press, they have been found to be data-poor. They both attend selectively to or ignore features of the extensive *Gf-Gc* research literature. Hence, if Sternberg (and Gardner) "had treated factorial theories and research on human abilities in more depth, their empirical and scholarly efforts might have systematically built upon (or undercut) these structural formulations and advanced the science of intellect in cumulative rather than idiosyncratic fashion" (Messick, 1992, p. 382). Although the MI and Triarchic theories may appear to be judged too harshly here and by others we cite, this does not diminish the possibility that these theories may eventually help broaden our understanding and measurement of intelligence. The MI and Triarchic theories are *different* from traditional psychometric theories of intelligence and, therefore, may require the development of different measurement approaches if they are to have a signficant influence on the practice of measuring intelligence.

Moving the Wechslers from Then to Now: Narrowing the Theory-Practice Gap

As depicted in Figure 2.1, there currently exists a significant theory-practice gap in the field of intellectual assessment. This is particularly true in the case of the WPPSI-R, K-ABC, and KAIT batteries which measure only two to three broad *Gf-Gc* abilities adequately (McGrew & Flanagan, 1998). The K-ABC primarily measures *Gv* and *Gsm,* and to a much lesser extent, *Gf,* while the KAIT primarily measures *Gf, Gc,* and *Glr,* and to a much lesser extent *Gv* and *Gsm.* And while the CAS, DAS, SB:IV, and WJ (1997) batteries do not provide sufficient coverage to narrow the theory-practice gap, their comprehensive measurement of four to five *Gf-Gc* abilities is nonetheless an improvement over the above mentioned batteries (McGrew & Flanagan, 1998).

Although the most recent versions of two Wechsler Scales (i.e., WAIS-III and WISC-III) measure a greater breadth of *Gf-Gc* abilities than their predecessors, a significant theory-practice gap remains. The WPPSI-R provides for adequate coverage of *Gc* and *Gv,* and partial measurement of *Gs, Gq,* and *Gsm.* The WISC-III and WAIS-III represent improvements over the WPPSI-R as both provide for good coverage of *Gs* in addition to *Gc* and *Gv.* Like the WPPSI-R, the WISC-III allows for only partial measurement of *Gq* and *Gsm.* However, the WAIS-III includes a single *Gf* test (viz., Matrix Reasoning) and expanded coverage of *Gsm.* These recent developments, although much welcomed and necessary, reflect slow incremental progress. As such, even when the WISC-III and WAIS-III are combined with the two memory batteries to which they are statistically linked (i.e., Children's Memory Scale and Wechsler Memory Scale—Third Edition, respectively), the end result does not represent an effective narrowing of the theory-practice gap. This is not unexpected given that the third generation of the Wechsler Scales have not been influenced overtly by contemporary *Gf-Gc* theory and research.

In defense of the Wechsler Scales, the serious attention that has focused on developing tests that are firmly grounded in empirically supported theories of intelligence is a relatively recent trend in the history of intelligence test development (Kamphaus, 1998; Sternberg & Kaufman, 1998). In terms of a single battery, currently the WJ-III comes closest to narrowing the gap between practice and contemporary *Gf-Gc* theory. As stated earlier, however, most major intelligence batteries need to be supplemented with other measures in order to narrow the *Gf-Gc* theory-practice gap.

Conclusion

There is little doubt that the Wechsler Intelligence Scale verbal/nonverbal (performance) model has exerted, and continues to exert, a significant influence on the measurement, classification, and interpretation of intellectual behavior. Because of the historical dominance of the Wechsler Scales in psychological assessment, it is understandable that many assessment professionals have internalized the belief that a verbal/nonverbal taxonomy is one of the best models for understanding intelligence test performance. However, the Wechsler verbal/nonverbal model is not based on an empirically derived theory of intelligence. Therefore, it should come as no surprise that comprehensive reviews of the extant cognitive abilities factor analysis research reveal a convergence on a hierarchical *Gf-Gc* model of intelligence and not a dichotomous verbal/nonverbal model. From the perspective of *Gf-Gc* theory, the Wechsler-based verbal/nonverbal model measures only a small portion of the 10 empirically supported broad *Gf-Gc* abilities. Carroll (1993a; 1993b) concluded that the Wechsler Verbal scale is an approximate measure of crystallized intelligence *(Gc)* and the Performance scale is an approximate measure of both broad visual perception *(Gv)* and, somewhat less validly, fluid intelligence *(Gf)*. Recent cross-battery factor analysis studies of the Wechsler Scales and other intelligence batteries (see Chapter 3) support Carroll's (1993b) *Gf-Gc* analysis of the Wechsler Scales. However, the Wechsler Performance scale is now being viewed as a measure of predominantly *Gv,* and not *Gf,* abilities (Elliott, 1994; Kaufman & Kaufman, 1993; McGrew & Flanagan, 1996; 1998; Woodcock, 1990). Carroll (1993a) provided a succinct judgement regarding the Wechsler Scales when he concluded that "presently available knowledge and technology would permit the development of tests and scales that would be much more adequate for their purpose than the Wechsler Scales" (p. 702).

We believe that the less than positive results from the second and third waves of intelligence test interpretation of the Wechsler Scales (i.e., clinical and psychometric profile analysis) were largely due to the use of either clinical, empirical, or theoretical taxonomics that were based on little or no sound empirical evidence. We believe that *Gf-Gc* theory is currently the *best* and most advanced taxonomy from which to understand human cognitive abilities and from which to improve the practice of theory-based intelligence test interpretation. *Gf-Gc* theory has provided a useful framework from which to analyze and interpret intelligence tests (e.g., the *Intelligence Test Desk Reference (ITDR): Gf-Gc Cross-Battery Assessment* [McGrew & Flanagan, 1998] is organized around the *Gf-Gc* taxonomy; Prentky, 1994). Furthermore, the *Gf-Gc* framework has provided a standard set of names or terms for the components of the entity (i.e., a standard nomenclature), an important feature of good taxonomies. Such an established nomenclature increases the effective communication among researchers and practitioners so that a "knowledge base can be accumulated" (Reynolds & Lakin, 1987).

Based on our review, the Wechsler verbal/nonverbal model (as well as a number of other theoretical models of intelligence) cannot be considered a *good* taxonomic system from which to organize thinking regarding intelligence tests. *The Wechsler verbal/nonverbal model does not represent a theoretically or empirically supported model of the structure of intelligence.* Unfortunately, a by-product of the Wechsler Scales' success is that many assessment professionals have grounded their practices in a largely atheoretical taxonomy of human cognitive abilities, a practice that was found to be seriously wanting in the second and third waves of intelligence test interpretation and a practice that constrains interpretation from benefiting from the now recognized need for theory-based interpretation. The staying power of this venerable (and out-of-date) taxonomy is at variance with viable taxonomies that are flexible and evolving. The premature hardening of the taxonomic categories can result in a deformation of the scientific process by "hermetically sealing of the boundaries of knowledge" (Prentky, 1994, p. 507).

We recognize that other theories of intelligence, including many of those reviewed briefly in this chapter, have either made important contributions to the intelligence knowledge base or posses new and/or unique features that will make new contributions and perhaps illuminate the limitations of the *Gf-Gc* theory. Contemporary *Gf-Gc* theory was presented as the most researched, empirically supported, and comprehensive descriptive hierarchical psychometric framework from which to organize thinking about intelligence test interpretation. According to Gustafsson and Undheim (1996), "the empirical evidence in favor of a hierarchical arrangement of abilities is overwhelming" (p. 204). An integrated Cattell-Horn-Carroll *Gf-Gc* model was presented here as the taxonomic framework around which the practice of intelligence testing and interpretation should be organized. The remainder of this book is devoted to describing how the interpretation of the Wechsler Intelligence Scales can be moved into the current emphasis on more theory-based intelligence test interpretation. In particular, we seek to modernize the Wechsler Scales by wrapping them in the *Gf-Gc* theory and taxonomy of human intelligence.

ENDNOTES

1. Piaget's theory of cognitive development is not included in this discussion since his work focuses primarily on universal cognitive changes and not on individual differences in cognitive development (Niesser et al., 1996).
2. It is important to not confuse low correlations with *g* as meaning those abilities that are furthest to the right in Carroll's model are not important. For example, Kamphaus (1997) has stated that the Wechsler Coding test is not important since it represents an ability *(Gs)* which is to the far right in Carroll's model. Abilities (and thus tests) may be low in *g* but may correlate significantly with other criteria. For example, a review of the literature has shown the *Gs* abilities are important for reading, math, and writing during the elementary school years (see Chapter 3). Claims that certain abilities are not important because of lower internal validity (i.e., not a strong measure of *g*) fail to recognize that internal and external validity are different forms of validity and that both are important for different reasons.

CHAPTER

3 Contemporary *Gf-Gc* Theory and Wechsler Test Interpretation

Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy *and* appropriateness *of* inferences *and* actions *based on test scores or other modes of assessment.*
—Messick (1989, p. 13)

In Chapter 2 it was concluded that most individually administered intelligence batteries, including the Wechsler Scales, fall short in the valid measurement of the full range of known *Gf-Gc* abilities. This conclusion is grounded in the notion that the ability to draw valid inferences about theoretical constructs from observable measures (e.g., tests) is a function of the extent to which the underlying program of validity research attends to both the *theoretical* and *empirical* domains of the focal constructs (Bensen, 1998; Bensen & Hagtver, 1996). The purpose of this chapter is to describe the characteristics of strong programs of test validation research, explain why parts of the Wechsler Intelligence Scales (as organized and typically interpreted) are judged *not* to be based on a strong validation research program, and finally, to demonstrate how the application of the *Gf-Gc* theory can strengthen the validity of the inferences drawn from Wechsler test scores.

Strong Theory: A Necessary Prerequisite for Strong Construct Validity

Similar to the calls for theory-based intelligence test interpretation in Chapter 1, leading scholars in the area of test validity (Cronbach, 1971; Cronbach & Meehl, 1955; Loevinger, 1957; Messick, 1989; Nunnally, 1978) stress the prominent role theory should play in the construction, validation, and interpretation of psychological tests (Bensen, 1998). Therefore, it is essential that the underlying theory of any psychological test be based on a solid foundation of evidence. In Chapter 2, *Gf-Gc* theory was described as the most comprehensive and empirically supported psychometric theory of intelligence and the theory around which intelligence tests should be developed and interpreted. Given the salient role theory plays in test validation, it is first necessary to provide evidence which supports the use of *Gf-Gc* theory in the interpretation of the Wechsler Intelligence Scales.

Supporting Evidence for *Gf-Gc* Theory

Extensive and robust factor-analytic evidence supports the validity of the *Gf-Gc* theory of intelligence. Unfortunately, since the factor-analytic literature on cognitive abilities is typically the only evidence cited in support of *Gf-Gc* theory, there is a common misconception that *Gf-Gc* theory is only a factor-analytic-based theory. However, support for the hierarchical *Gf-Gc* theory has been documented through five major forms of validity evidence (Gustafsson & Undheim, 1996; Horn, 1994; Horn & Noll, 1997). Each is discussed briefly in the text following.

Structural Evidence

Structural evidence, or evidence based on the individual differences, factor-analytic research tradition, has been the most prominent evidential base for the *Gf-Gc* constructs (Taylor, 1994). This source of evidence is based on the principle of concomitant variation. That is, if measures covary repeatedly across studies that differ in sample characteristics, time, and place, then this covariation suggests the plausibility of a common underlying function (Horn & Noll, 1997). The extant factor-analytic research over the past 50 to 60 years has converged consistently on models of intelligence similar to the *Gf-Gc* models presented in Figure 2.4 (in Chapter 2). Furthermore, the *Gf-Gc* structure presented in Figure 2.4 has been found to be invariant across different gender, ethnic, and racial groups (Carroll, 1993a). Carroll (1993a) concluded 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" (p. 687).

Given the wide age range covered by the family of Wechsler Intelligence Scales, it is important to know if the *Gf-Gc* model presented in Figure 2.4 is invariant across ages. Historically, both logical and theoretical considerations have suggested that cognitive abilities become more differentiated with age (the age-differentiation hypothesis) (Carroll, 1993a). Carroll's massive review of the cognitive ability factor-analytic research addressed the age-differentiation issue given that it included studies with subjects from as young as 6 to 11 months to 70 years. Carroll (1993a) stated that "my general conclusion on age-differentiation of cognitive ability factors is that it is a phenomenon whose existence is hard to demonstrate . . . the question of the age differentiation is probably of little scientific interest except possibly at very young ages . . . the same factors are found throughout the life span" (p. 681). The apparent invariance of the *Gf-Gc* factors across the life span, male and female groups, and different cultures and racial groups supports the application of the *Gf-Gc* cross-battery approach to intelligence test interpretation for most of the population.

Developmental Evidence

The validity of the *Gf-Gc* constructs is supported also by differential developmental changes in the growth and decline of cognitive abilities across the life span (Carroll, 1993a; Dixon, Kramer, & Baltes, 1985). This type of evidence typically takes the form of comparing *Gf-Gc* growth curves. *Developmental evidence* has shown that different *Gf-Gc* abilities follow divergent developmental trajectories with increasing age (Horn, 1982; 1985; Horn & Cattell, 1967; Schaie, 1979; 1983; 1994). For example, from young adulthood to old age, increases in *Gc* and *Glr* (maintained abilities) and decreases in *Gf, Gs,* and *Gsm* (vulnerable abilities) have been reported (see Horn & Noll, 1997 for a summary). The finding of maintained and vulnerable abilities and differential ability growth curves across the lifespan suggests that different mechanisms or determinants (e.g., education, genes, injuries, lifestyle factors) operate differentially in the development and decline of *Gf-Gc* abilities, evidence that supports the validity of the different *Gf-Gc* constructs (Carroll, 1983; Horn & Noll, 1997).

Neurocognitive Evidence

Neurocognitive evidence exists in the form of empirical relations between measures of *Gf-Gc* abilities and physiological and neurological functioning (Horn & Noll, 1997). For example, the norepinephrine system of the brain has been associated with neurological arousal that is characteristic of *Gf* abilities (Horn, 1982; 1985; Iverson, 1979). The localization of specific cognitive abilities in certain regions of the brain (e.g., verbal or *Gc*-type abilities are often reported to be localized mainly in the left hemisphere) is another example of neurocognitive evidence (see Kaufman, 1990 and Lezak, 1995 for summaries). Differential declines in different *Gf-Gc* abilities that are associated with age-related central nervous system deterioration (e.g., the development of senile plaques apparent in Alzheimer's patients) suggest that different *Gf-Gc* abilities are supported by different underlying brain structures and functions—a finding that further supports the construct validity of the different abilities.

Heritability Evidence

Another form of support for the different *Gf-Gc* abilities is *heritability evidence,* evidence concerned with the "proportion of phenotypic (observed) differences among individuals in a population that can be attributed to genetic differences among them" (Plomin & Petrill, 1997, p. 57). Although at times controversial and theoretically and methodologically complex (see McArdle & Prescott, 1997), behavioral-genetic research has suggested that different sets of genes may determine different structures and functions of the brain (Plomin & Petrill, 1997). Although no definitive conclusions have been reached, different heritability estimates have been reported for different cognitive abilities in some studies (Carroll, 1993; McGue & Bouchard, 1989; Plomin, DeFries, & McClearn, 1990; Scarr & Carter-Saltzman, 1982; Vandenberg & Volger, 1985). For example, Vanderberg and Volger (1985) cite parent-offspring research that suggests that spatial (*Gv*-Spatial Relations) and verbal *(Gc)* abilities have higher heritabilities than visual memory (*Gv*-Visual Memory) and perceptual speed (*Gs*-Perceptual Speed). In studies of twins, McGue and Bouchard (1989) reported that genetic influences were largest for spatial (*Gv*-Spatial Relations) abilities and smallest for visual memory (*Gv*-Visual Memory) abilities. Some behavioral genetic research has suggested that different *Gf-Gc* abilities may be influenced by separate genetic and environmental factors. When combined with research that has reported the differentiation of cognitive abilities at early ages (Carroll, 1993), Horn and Noll (1997) concluded that "the outlines for different intelligences can be seen in early childhood" (p. 81).

Outcome-Criterion Evidence

Finally, differential achievement or *outcome-criterion evidence* supports the existence of separate *Gf-Gc* abilities. Supporting outcome-criterion evidence is found in research studies that have investigated the relations between *Gf-Gc* abilities and academic achievement, occupational success, and other human traits. This outcome-criterion evidence is important, as *Gf-Gc*-based intellectual assessments will be of little practical value if they fail to produce valid interpretations of intellectual functioning that contribute to improved diagnostic and classification decisions, predictions about performance, and interventions. A first step on the road to using *Gf-Gc* theory to improve the practice of intellectual assessment is to understand the relations between *Gf-Gc* abilities and other variables.

McGrew and Flanagan (1998) presented a summary of more than a decade of research that has examined the relations between different *Gf-Gc* constructs and measures and other non-*Gf-Gc* constructs and measures. An abstracted summary of their review is presented in Table 3.1. In general, the information presented in Table 3.1 indicates that different *Gf-Gc* constructs (and valid measures of the constructs) are significantly and differentially related to different academic, occupational, interest, and personality variables.[1] This form of evidence provides additional support for the validity of the different *Gf-Gc* constructs.

Supporting Evidence: Concluding Comments and Cautions

The validity evidence present for the *Gf-Gc* theory approximates the desired standard of validity evidence—a nomological network of different types of validity evidence (viz., structural, developmental, neurocognitive, heritability, and achievement or outcome criteria). This conclusion is similar to that reached by Messick (1992), who, after comparing the validity evidence for the *Gf-Gc* theory and two other theories of multiple cognitive abilities (i.e., Gardner's and Sternberg's theories), concluded that the *Gf-Gc* theories of intelligence "fare somewhat better . . . because they reflect many decades of programmatic research" (p. 382). Messick went as far as to state that *Gf-Gc* "multifactor theory and measurement provide a partial standard of validity for both Gardner and Sternberg" (1992, p. 366). It seems clear that *Gf-Gc*-organized Wechsler-based cross-battery assessments (described in Chapters 7 and 8) have the potential to contribute meaningfully to research studies and reviews on the relations between cognitive abilities and many different outcome criteria, because they are organized within this well-articulated and researched theoretical framework.

TABLE 3.1 Select *Gf-Gc* External Validity Evidence Summarized by McGrew and Flanagan (1998)

	Reading Achievement	Math Achievement	Occupational Performance	Interest Traits	Personality Traits
Gf	Inductive (I) and general sequential reasoning (RG) abilities play a moderate role in reading comprehension.	Inductive (I) and general sequential (RG) reasoning abilities are consistently very important at all ages.	Mathematician Scientist	Realistic Investigative Enterprising (–)	Openness to Experiences
Gc	Language development (LD), lexical knowledge (VL), and listening ability (LS) are important at all ages. These abilities become increasingly important with age.	Language development (LD), lexical knowledge (VL), and listening abilities (LS) are important at all ages. These abilities become increasingly more important with age.	Accountant Leader (military)/Soldier Lawyer Poet Scholar Scientist	Realistic Investigative Artistic Conventional (–)	Typical Intelligence Engagement Test Anxiety
Gsm	Memory span (MS) is important especially when evaluated within the context of working memory.	Memory span (MS) is important especially when evaluated within the context of working memory.		Realistic Enterprising (–)	Openness to Experiences Test Anxiety (–)
Gv		May be important primarily for higher level or advanced mathematics (e.g., geometry, calculus).	Carpenter Engineer Leader (military)/Soldier Machinist Photographer Teacher Architect/Draftsperson Artist/Sculptor Electrician Navigator/Pilot Air traffic controller Designer Mathematician Scientist		Typical Intelligence Engagement
Ga	Phonetic coding (PC) or "phonological awareness" is very important during the elementary school years.		Musician Musical composer Sonar Operator		
Glr	Naming facility (NA) or "rapid automatic naming" is very important during the elementary school years. Associative memory (MA) may be somewhat important at select ages (e.g., age 6 years).				
Gs	Perceptual speed (P) abilities are important during all school years, particularly the elementary school years.	Perceptual speed (P) abilities are important during all school years, particularly the elementary school years.	Accountant Leader (military)/Soldier Clerk/Typist Proofreader		Conventional Test Anxiety (–)

Note: (–) Indicates a negative correlation between the *Gf-Gc* ability and the interest and/or personality trait.

Although there is considerable support for the *Gf-Gc* theory of intelligence, it should not be considered *the* definitive theory. *Gf-Gc* theory is not without limitations. Horn and Noll (1997) summarized four major limitations of *Gf-Gc* theory. Carroll (1995b; 1997) also provided appropriate words of caution about the limits of *Gf-Gc* theory.

First, *Gf-Gc* theory is more a descriptive, empirical generalization of research findings than a deductive explanation of these findings. A research tradition has evolved in which the *Gf-Gc* variables included in successive studies are based on the variables included in prior studies, a situation that does not include the requisite *a priori* theoretical basis for validating a theory. Horn and Noll (1997) acknowledge this limitation, but point out that all scientific theory is the result of a research history and culture, and it "evolves out of a repetitive spiral of building on what is known (induction), which leads to deductions that generate empirical studies and more induction, which leads to further deductions, which spawn further induction, and so on" (p. 83).

Second, the structure implied in the *Gf-Gc* theory is a limitation. Although the statistical method of factor analysis can produce neat hierarchically organized factors, these empirically based frameworks most likely do not represent accurately the organization of actual human cognitive abilities. That is, *Gf-Gc* theory is largely a product of linear equations (viz., factor analysis). Natural phenomena most likely are nonlinear in nature. As Horn and Noll (1997) stated, "The equations that describe the outer structure and convolutions of brains must be parabolas, cycloids, cissoids, spirals, folliums, exponentials, hyperboles, and the like. It is likely that the equations that best describe the inner workings of brains—human capabilities—are of the same forms, not those that describe city blocks and buildings" (p. 84).

Third, *Gf-Gc* theory provides little information on how the *Gf-Gc* abilities develop or how the cognitive processes work together. The theory is largely product oriented and provides little guidance on the dynamic interplay of variables (i.e., the processes) that occur in human cognitive processing (Gustafsson & Undheim, 1996). However, as described in Chapter 2, recently Woodcock (1993, 1997) has articulated a *Cognitive Performance Model* and a *Gf-Gc Information Processing Model* of intellectual performance that specify relations between and among *Gf-Gc* abilities, information processing constructs, and noncognitive variables. Currently these models are largely speculative and need further study.

Fourth, Carroll (1997), one of the primary architects of the *Gf-Gc* taxonomy, humbly pointed out that additional work needs to be completed in the factor-analytic study of human cognitive abilities. "The map of abilities provided by the three-stratum theory undoubtedly has errors of commission and omission, with gaps to be filled in by further research" (Carroll, 1997, p. 128). Carroll (1995) stated that certain aspects of the hierarchical structure may need to be refined and/or revised, including the identification of additional narrow abilities, the clarification of already identified narrow abilities, and the clarification of the number and structure of broad abilities. Although Carroll's wise words should temper the tendency to believe that we have now discovered the "holy taxonomic grail" of human cognitive abilities, the *Gf-Gc* taxonomy is currently the most comprehensive and empirically supported psychometric framework from which to understand the structure of human intelligence.

Linking *Gf-Gc* Theory and Applied Measurement: The Importance of a Strong Program of Construct Validation Research

The benefits accrued from the identification of a valid theory of intelligence (i.e., *Gf-Gc* theory) are wasted if the development of operational measures of the relevant cognitive constructs is not based on a strong program of construct validation research (Benson, 1998; Cronbach, 1989). But just what is a "strong" program of construct validation research and how does it relate to the development and interpretation of the Wechsler Intelligence Scales?

Many leading researchers and scholars in the area of psychological measurement have, at one time or another, described similar validation stages or components that are necessary for strong test validity (see, for example, Cronbach, 1971; Cronbach & Meehl, 1955; Loevinger, 1957; Messick, 1989; Nunnally, 1978). In general, a common theme among these voices is that the test validation process is similar to that used to develop and evaluate scientific theories and involves the evaluation of both theory and measures in a concurrent iterative process. Benson's (1998) synthesis of the various validity concepts and components is particularly useful for understanding the interplay between theory and measurement. According to Benson (1998), the steps that allow for valid inferences to be drawn from test scores are those illustrated in Figures 3.1 and 3.2.

The Substantive Stage of Construct Validation: A Conceptual Explanation

The information presented in Figures 3.1 and 3.2 conveys the message that a strong theory is a necessary foundation from which to develop and interpret valid measures of intelligence. A strong theory of intelligence is needed in order to specify, define, and circumscribe both the theoretical and empirical domains of the focal constructs. This point will be illustrated within the domain of visual intelligence or processing *(Gv)*.

As illustrated in Figure 3.1, the first step toward developing valid measures of the *Gv* construct is the identification of a strong theory. We have already presented ample evidence to support the conclusion that *Gf-Gc* theory is a strong theory. Therefore, the *Gf-Gc* framework and the *Gv* domain (which is part of this framework) constitute the *theoretical domain* (see Figure 3.1) from which to develop measures during the substantive stage (see Figure 3.2) in this example.

The next step in the substantive stage of construct validation is to develop and evaluate possible measures of the theoretical constructs (i.e., the *empirical domain*) (see Figure 3.1). During this step a variety of methods and concepts should be employed to "flesh out" the potential measures in the empirical domain (see Figure 3.2). For example, definitions of the theoretical constructs must be specified in order to circumscribe the constructs to be measured. Typically, this is followed by the development of test formats, the generation of

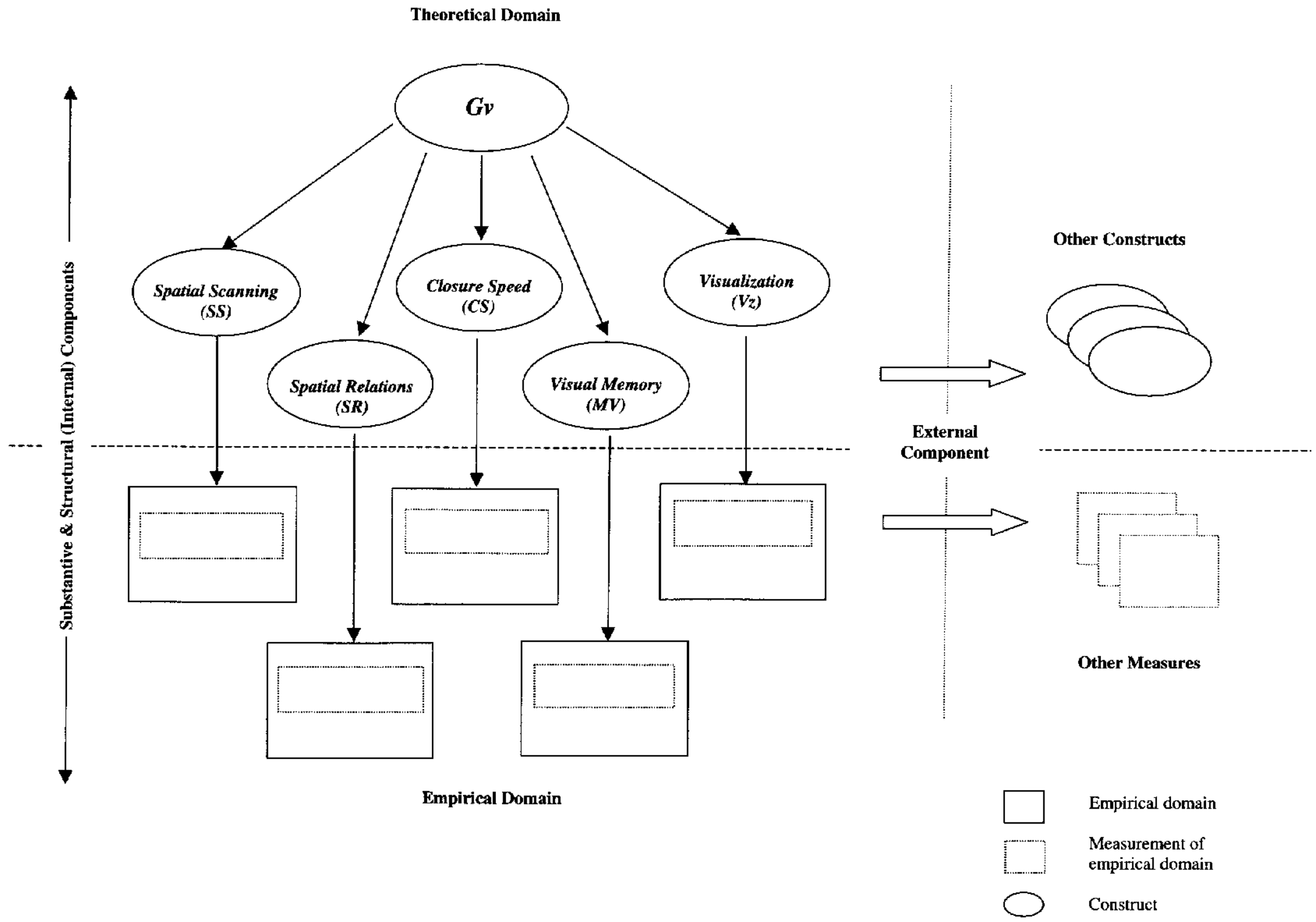

FIGURE 3.1 Relationship between Three Stages/Components of Strong Construct Validation Research Programs: A *Gv* Example

Note: The *Gv* narrow abilities of Flexibility of Closure (CF), Serial Perceptual Integration (PI), Length Estimation (LE), Perceptual Illusions (IL), Perceptual Alternations (PN), and Imagery (IM) are excluded from this figure.

	Substantive Stage/Component	Structural Stage/Component	External Stage/Component
Purpose	• Define the *theoretical* and *empirical* domains of intelligence	• Examine the *internal* relations among the measures used to operationalize the theoretical construct domain (i.e., intelligence)	• Examine the *external* relations among the focal construct (i.e., intelligence) and other constructs and/or subject characteristics
Question Asked	• How should intelligence be defined and operationally measured?	• Do the observed measures "behave" in a manner consistent with the theoretical domain definition of intelligence?	• Do the focal constructs and observed measures "fit" within a network of expected construct relations (i.e., the nomological network)?
Methods & Concepts	• Theory development & validation • Generate definitions • Item and scale development • Content validation • Evaluate construct underrepresentation and construct irrelevancy	• Internal domain studies • Item/subscale intercorrelations • Exploratory/confirmatory factor analysis • Item response theory (IRT) • Multitrait-Multimethod matrix • Generalizability theory	• Group differentiation • Structural equation modeling • Correlation of observed measures with other measures • Multitrait-Multimethod matrix
Characteristics of *strong* validation programs	• A strong psychological theory plays a prominent role • Theory provides a well-specified and bounded domain of constructs • The empirical domain includes measures of all potential constructs (i.e., adequate construct representation) • The empirical domain includes measures that only contain reliable variance related to the theoretical constructs (i.e., construct relevance)	• Moderate item internal consistency • Measures covary in a manner consistent with the intended theoretical structure • Factors reflect trait rather than method variance • Items/measures are representative of the empirical domain • Items fit the theoretical structure • The theoretical/empirical model is deemed plausible (especially when compared against other competing models) based on substantive and statistical criteria	• Focal constructs vary in theorized ways with other constructs • Measures of the constructs differentiate existing groups that are known to differ on the constructs • Measures of focal constructs correlate with other validated measures of the same constructs • Theory-based hypotheses are supported, particularly when compared to rival hypotheses

FIGURE 3.2 A Continuum of Stages for Strong Construct Validation Research Programs

Note: Information in this table is based on Benson, J. (1988). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice.*

test items, the gathering of item calibration data, and the use of various item analysis procedures (e.g., Rasch scaling) to evaluate the adequacy of the items and scales. Essential to the process are activities directed at insuring adequate *content validity* of the measures that will be used to represent the theoretical construct (Benson, 1998; Messick, 1989).

Briefly, *content* or *construct representation* is concerned with the extent to which the items/tests within an empirical domain adequately reflect the major aspects of the theoretical domain of constructs (Bensen, 1998). In Figure 3.1, adequate *Gv* construct representation would be insured through the development of tests for the abilities of Spatial Scanning, Spatial Relations, Closure Speed, Visual Memory, and Visualization. The development and use of a single test indicator (e.g., a test of Visual Memory) or two similar indicators (e.g., two different tests that measure Visual Memory) to represent the *Gv* construct would not result in accurate and valid inferences regarding the complete *Gv* construct. In general, the most valid measures of intellectual ability constructs (i.e., *Gf-Gc* abilities) would be those that include the largest number of different tests (within practical constraints), each measuring a unique aspect of every one of the broad theoretical constructs.

Substantive Validity of the Wechsler Scales

With the benefit of hindsight, it is now clear that the Wechsler Intelligence Scales are not grounded in a strong theory of intelligence and thus, have a weak substantive foundation. In addition to the material summarized in previous chapters, evidence that supports this conclusion is the paucity of substantive Wechsler Scale validity research studies. A comprehensive literature review of validity studies conducted with the Wechsler Intelligence and Memory Scales since 1989 was completed by the present authors. A summary of this review is provided in Appendix A (Network of Validity Evidence of the Wechsler Scales). The studies included in this appendix were classified by the present authors as fitting into one of these categories: substantive, structural, and external. The information presented in Figure 3.2 guided our classifications.[2]

Because the *Gf-Gc* taxonomy "provides what is essentially a 'map' of all known cognitive abilities . . . [it] can be used in interpreting scores on the many tests used in individual assessment" (Carroll, 1993a, p. 127). In other words, the *Gf-Gc* taxonomy can serve as a blueprint from which to evaluate the content validity of intelligence tests. The first systematic attempt to map the major intelligence batteries to the *Gf-Gc* structural framework was presented by Woodcock (1990). Woodcock's (1990) classifications, which were only at the broad (stratum II) level, were extended to the narrow (stratum I) level by McGrew (1997). McGrew and Flanagan (1998) subsequently refined McGrew's classifications and extended the classifications to over 200 cognitive ability measures. The Wechsler *Gf-Gc* test classifications presented by McGrew and Flanagan are used here to evaluate the content or substantive validity of the Wechsler Intelligence Scales.

The Wechsler *Gf-Gc* classifications required that each test be classified at both the broad (stratum II) and narrow (stratum I) ability levels. The broad abilities measured by tests coincide with the *broad Gf-Gc* abilities in the Integrated Cattell-Horn-Carroll model described in Chapter 2. Figure 3.3 presents an example of a *Gf-Gc* broad- and narrow-ability test mapping for five Wechsler tests.

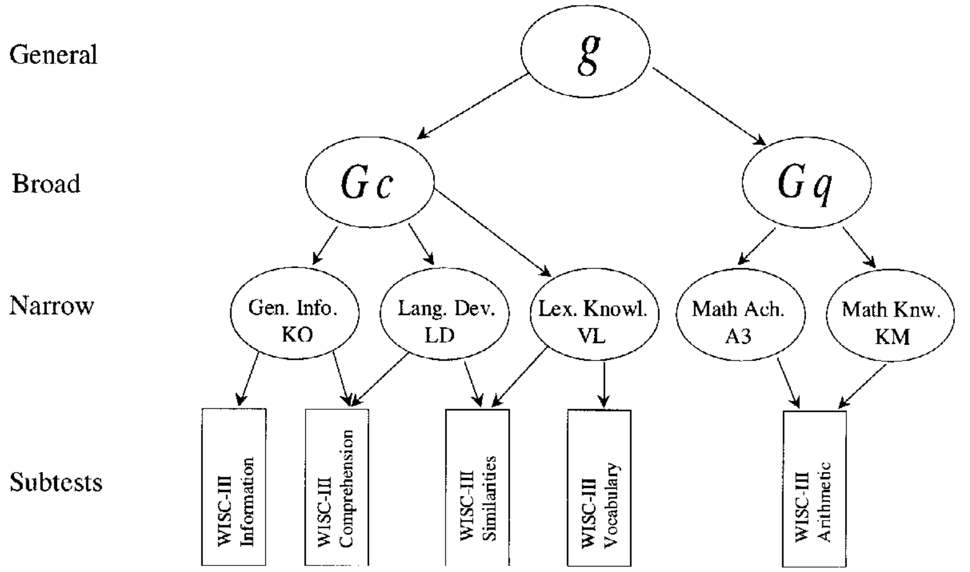

FIGURE 3.3 The Relations between Five WISC-III Tests and the Three Strata (Narrow, Broad, and General) of the *Gf-Gc* Model of Intelligence

Note: There are additional narrow abilities in the domains of *Gc* and *Gq* that are not included in this figure.

Figure 3.3 shows that the broad ability of *Gc* subsumes the specialized narrow abilities of General Information, Language Development, Lexical Knowledge, and others not included in the figure. The rectangles in Figure 3.3 represent five tests from the WISC-III. In this figure, the Information, Comprehension, Similarities, and Vocabulary tests are all classified as indicators of the broad *Gc* ability. The Arithmetic test is classified as an indicator of the broad *Gq* ability. The complete *Gf-Gc* content validity evaluation of the Wechsler Scales is presented in Figure 3.4. The broad *Gf-Gc* test classifications presented in Figure 3.4 were derived from a series of confirmatory cross-battery intelligence test factor analysis studies (described below). The narrow-ability classifications of the Wechsler subtests, as well as the expert consensus procedures on which they are based, are described in McGrew (1997) and McGrew and Flanagan (1998) and are summarized in Chapter 4.

A review of Figure 3.4 reveals that the Wechsler Intelligence Scales have weak content validity when evaluated according to the *Gf-Gc* theoretical model. This is not a surprising conclusion given that the Wechsler scales have, by and large, maintained their ancestral link with David Wechsler's largely atheoretical notions of intelligence. Of the *Gf-Gc* domains that are represented across the Wechsler Intelligence batteries (viz., *Gf, Gc, Gsm, Gv, Ga, Glr, Gs*), *Gc* and *Gv* are the only abilities that have strong content or construct representation (defined by the measurement of three or more narrow abilities under a broad ability). In addition to *Gc* and *Gv*, the WISC-III adequately represents the construct of *Gs* (i.e., it measures at least two distinct narrow abilities that define this broad ability) while the WAIS-III adequately represents *Gs* and *Gsm*.

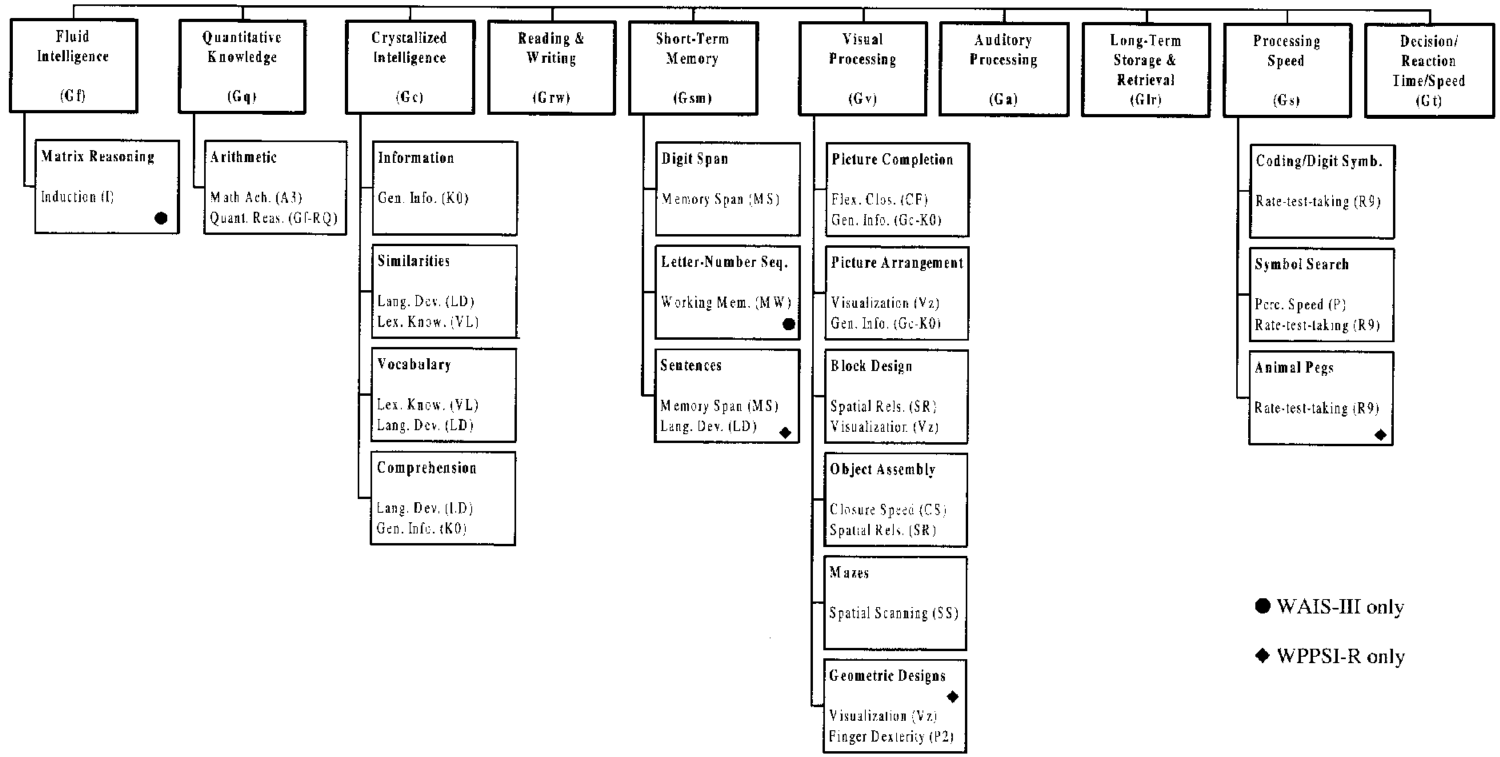

FIGURE 3.4 Wechsler *Gf-Gc* Empirical-Theoretical Domain Mapping Based on Strong Substantive and Structural Validity Evidence: The Foundation of the Wechsler *Gf-Gc* Cross-Battery Approach

The Wechsler Intelligence Scales do not contain any measures of *Ga* and *Glr*. In the *Gf* domain, the WAIS-III has weak construct representation (i.e., it only measures one narrow ability within this domain), while the WPPSI-R and WISC-III do not contain any strong measures of *Gf*. Similarly, *Gsm* is underrepresented on the WPPSI-R and WISC-III (i.e., each scale provides only one measure of this construct) and *Gs* is underrepresented on the WPPSI-R.

This content validity analysis suggests that, from the perspective of *Gf-Gc* theory, the Wechsler Intelligence Scales have a weak substantive foundation. This is primarily due to the fact that during the development and subsequent revisions of the Wechsler Intelligence Scales, no explicit attempt was made to specify and circumscribe the theoretical domain (and subdomains) of intelligence. As a result, the Wechsler Intelligence Scales have not benefited from a test development process where measures were operationalized according to a theory-based "blueprint" similar to that portrayed in Figure 3.1. The continuum represented in Figure 3.2 suggests that the lack of a strong substantive component is the Achilles heel of the Wechsler Intelligence Scales. The largely atheoretical foundation of the Wechsler Scales places a number of constraints on the validity of inferences that can be drawn from all of the Wechsler tests and composite scores. As will be demonstrated in the remainder of this chapter, the lack of a sound theory constrains the degree to which interpretation of the Wechsler Scales can benefit from current efforts to make test interpretation more theory based.

The Structural Stage of Construct Validation: A Conceptual Explanation

As highlighted in Figure 3.2, the *structural stage* of test validation has an internal focus on the collection of tests designed to measure the theoretical domain (e.g., *Gv*). A variety of internal domain studies and methods are used to evaluate whether the tests measure the construct(s) they are purported to measure. That is, does empirical evidence support a strong link between the empirical and theoretical domains of the focal constructs? Although a theory of intelligence may have considerable popular appeal or supporting evidence, the lack of psychometrically sound measures of the focal constructs of a theory will limit its practical utility.

As is the case with all major intelligence batteries, Appendix A reveals that the primary structural validity tool of the Wechsler Scales has been *factor analysis,* a statistical procedure that groups together measures (tests) that intercorrelate positively with one another. The use of factor-analytic methods in the investigation of the structural validity of intelligence batteries can be organized along two dimensions. The first dimension is the *method* of factor analysis that is employed (i.e., exploratory or confirmatory, both of which subsume a variety of different procedures). *Exploratory* factor analysis allows the data to "speak for themselves," since it identifies the factor structure of an intelligence battery with no *a priori* model in mind. In contrast, *confirmatory* factor analysis typically examines the extent to which an *a priori* hypothesized factor structure or model fits the data, as well as how the fit of the model compares to alternative models, and how the model might be modified to fit the data better. Confirmatory factor methods are becoming particularly prominent in theory-based test development and research activities (Keith, 1997).

The second dimension of factor-analytic methods involves the breadth of cognitive ability indicators (or tests) that are analyzed. *Within-battery* factor analysis is confined to the tests from a single intelligence battery (e.g., the 13 WISC-III tests). *Cross-battery* (or joint) factor analysis, includes tests from more than one intelligence battery (e.g., WISC-III and WJ-III).

In the *Gv* example presented in Figure 3.1, if tests of Spatial Scanning, Spatial Relations, Closure Speed, Visual Memory, and Visualization are factor analyzed together with indicators of other *Gf-Gc* constructs, structural or internal construct validity evidence would emerge in the grouping (i.e., loading) of these five specific tests on a *Gv* factor while other non-*Gv* tests would load on separate and uniquely different factors. In the case of intelligence tests like the Wechsler Scales, positive structural evidence would be inferred if the tests selected or developed from the empirical domain are determined to be valid measures of the corresponding constructs in the theoretical domain (see Figure 3.1).

Structural Validity of the Wechsler Scales

David Wechsler appeared to have strong convictions regarding the tasks in his scales and their ability to measure a range of cognitive functions sufficient to assess general intelligence validly (Wechsler, 1958; Zachary, 1990). As such, he may have felt little need to include much information regarding validity in the technical manuals of his tests. The little evidence that was presented in support of the early Wechsler Scales' validity reflected, for the most part, a content-description orientation (e.g., some data concerning the correlation between the Wechsler Scales and other global measures of intelligence such as the Stanford-Binet, the latter of which hovered around 0.80). Although there were references to the underlying factor structure of the Wechsler Scales, the early manuals never addressed the issue empirically. Rather, it was left to the auspices of independent researchers to provide such data, which they did in abundance.

Unlike the earlier versions, the current Wechsler Scales come with manuals that are replete with validity data. For example, nearly half of the entire WAIS-III manual (not including appendices) is devoted to the topic of validity. Like their predecessors, the validity data reported in the current Wechsler manuals continue to be supported by extensive independent investigations (see Appendix A). It is clear that validity has emerged as the preeminent concern for the publishers of the Wechsler Scales. Below we present a summary of the factor-analytic construct validity evidence that has been reported for the Wechsler Scales.

Within-Battery Structural Research. The preponderance of structural validity studies included in Appendix A support the internal validity of the Wechsler Scales. Within-battery factor-analytic studies of the early Wechsler Scales invariably revealed a two-factor solution involving a verbal and nonverbal (or perceptual-organizational) factor (Gutkin & Reynolds, 1981; Kaufman, 1975; Leckliter, Matarazzo, & Silverstein, 1986; Silverstein, 1982). A number of the WAIS-R factor analysis studies listed in Appendix A (e.g., Gutkin, Reynolds & Galvin, 1984; Parker, 1983; Silverstein, 1982) were included in a recent literature review in which the reviewers concluded that the WAIS-R was characterized best by three factors—namely, Verbal Comprehension (VC), Perceptual Organization (PO), and Working Memory/Freedom From Distractibility (WMFFD) (Leckliter, Matarazzo, & Silverstein, 1986). A similar but more recent example listed in Appendix A is Keith and Witta's (1997) multisample hierarchical confirmatory factor analysis of the WISC-III standardization data. These researchers found support for similar VC and PO factors as well as a similar FFD-like factor they called Quantitative Reasoning. Consistent with the WISC-III factor-based index interpretation system, Keith and Witta found support for a separate Processing Speed (PS) factor.

The large number of within-battery factor-analytic studies reported in Appendix A supports the current consensus that a three- or four-factor structure characterizes the various permutations of the 11 common core Wechsler subtests. These factors include Verbal Comprehension (VC), Perceptual Organization (PO), Processing Speed (PS), and Working Memory (WM) (or Freedom From Distractibility, depending on the Scale). In lieu of a detailed and lengthy discussion of all the within-battery factor analysis studies of the various versions of the Wechsler Scales, Figure 3.5 is used to summarize and illustrate how the extant factor analysis research findings have changed the face of the Wechsler interpretive framework over time.

Figure 3.5 shows that the Verbal and Performance two-factor structure was believed to best represent the structural characteristics of the Wechsler Scales until the late 1970s. Alan Kaufman's research and writings on the WISC-R (Kaufman, 1979) were particularly instrumental in the acceptance of the familiar three-factor (VC, PO, FFD) structure. As demonstrated in Figure 3.5, this more differentiated WISC-R factor structure occurred through the splitting off of the Arithmetic and Digit Span subtests from the Verbal Scale, and their marriage with the Coding subtest (which split off from the Performance Scale) to form the new FFD triad.

Movement toward an even more differentiated factor structure occurred with the publication of the WISC-III (Wechsler, 1991). The addition of the Symbol Search test resulted in the splitting of the FFD factor into separate FFD and PS factors in factor analysis studies. Likewise, the addition of two new tests (i.e., Letter-Number Sequencing and Matrix Reasoning) to the Wechsler subtest family with the publication of the WAIS-III resulted in a "fine-tuning" of the four-factor structure of the Wechsler Scales. In addition to the WISC-III-like VC and PS factors, the latest Wechsler battery has expanded the composition of the PO factor with the addition of the Matrix Reasoning test. In addition, the new Letter-Number Sequencing test has resulted in a reconceptualization of the FFD factor into a three-test Working Memory (WM) factor.

The movement away from the original Verbal and Performance Scales in the direction of the four factor indexes (VC, PO, PS, WM), a change that is supported by a substantial body of systematic within-battery factor analysis research within the past 10 to 15 years (see Appendix A), reflects incremental improvement in the internal or structural validity of the factor-based interpretation schemes of the Wechslers as manifest in the most recent Wechsler Scale (i.e., WAIS-III). This improvement in the structural validity of the WAIS-III, in particular, suggests a harbinger of changes in the next revisions of the WPPSI-R and WISC-III.

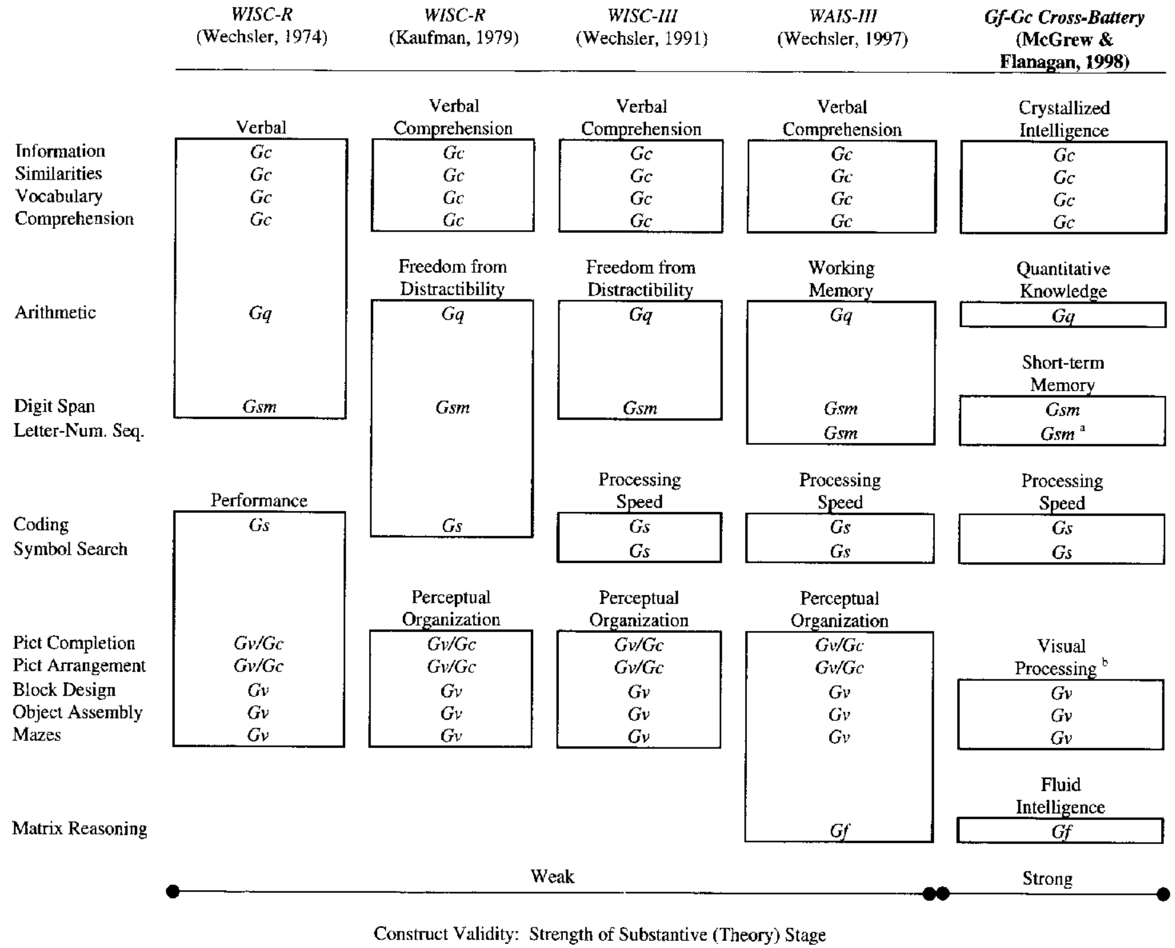

FIGURE 3.5 A Comparison of Wechsler Intelligence Scale Interpretation Schemes Based on Within-Battery and Cross-Battery Structural Validity Research

[a]Working memory (MW) has been integrated into the *Gsm* factor domain in the current work (see Chapter 2).

[b]Picture Completion and Picture Arrangement are not included as they are *mixed* measures of two constructs (*Gv* and *Gc*).

Although the within-battery factor analysis research of the most recent Wechsler Scales generally support the internal validity of the batteries, the ever changing interpretation of two of the original FFD tests (viz., Digit Span and Arithmetic) should give pause for concern (McGrew, 1999). As reflected in Figure 3.5, within-battery factor analyses of the various Wechsler Scales have shown that the Arithmetic and Digit Span subtests have changed their factor allegiance from their original Verbal factor to that of FFD, and more recently, WM. The ever changing factorial nature of the Arithmetic and Digit Span subtests raises concerns about the validity of the factor-based interpretations that have been offered for these subtests. Not unexpectedly, the factorial ambiguity of these two tests resulted in independent researchers offering a wide range of interpretations for the FFD factor (Kamphaus, 1993). For example, this factor has been suggested to measure short-term memory, math achievement, attention and concentration, sequential processing, and executive problem-solving strategies, as well as a variety of personality and emotional constructs (Kaufman, 1994; Wielkiewicz, 1990). How can this be? Even Alan Kaufman, whose name has become synonymous with the FFD label, has expressed frustration with the interpretative quagmire as reflected in his statement, "I cringe whenever I read 'Kaufman's Freedom from Distractibility factor.' It's not mine, and I don't want it" (Kaufman, 1994, p. 212).

The authors believe that the variable interpretations for the Arithmetic, Digit Span, and Coding/Digit Symbol tests reflect, in large part, the fallout from the weak substantive or theoretical foundation of the Wechsler Scales. It seems clear that the structural validity of the Wechsler Scales can be improved on significantly through the application of theory-based cross-battery factor analysis research. The application of *Gf-Gc* organized cross-battery factor analysis to the Wechsler Scales represents, in a sense, a post-hoc structural validity repair to the substantive foundation of the Wechsler Scales.

Cross-Battery Structural Research Defined. Cross-battery (or joint) factor analysis, as stated earlier, includes tests from *more than one* intelligence battery in the analysis. Theory-based cross-battery factor analysis intelligence test studies are consistent with the focus of the fourth wave of intelligence test interpretation since these investigations organize the analyses from the perspective of a particular theory of intelligence. That is, the measures from the various intelligence batteries are freed from the confines of their own battery and are allowed to, or are specified to, load on the various cognitive factors included in the theoretical domain. Thus, the tests from the different individual batteries are allowed to *cross* battery boundaries in order to mingle with other tests and subsequently load on the theoretical factors specified by the theory of intelligence.

Consistent with the call for more theory-based intelligence test interpretation, the authors believe that the *Gf-Gc* theory of intelligence, if imposed on the analysis of the complete collection of Wechsler subtests together with measures from other intelligence batteries, will facilitate understanding of the structural characteristics and validity of the Wechsler Intelligence Scales. Fortunately, there have been a series of recent *Gf-Gc*-designed cross-battery factor analysis studies with a number of major intelligence batteries. These confirmatory cross-battery studies, which collectively will be referred to as the *CB studies* hereafter, included Woodcock's (1990) analyses of the WJ-R, Wechsler's, SB:IV (Thorndike, Hagen, & Sattler, 1986), and K-ABC (Kaufman & Kaufman, 1983); McGhee's (1993) analyses of the WJ-R, Differential Ability Scales (DAS; Elliott, 1990a), and Detroit Tests of Learning Aptitude-3 (DTLA-3; Hammill & Bryant, 1991); Flanagan and McGrew's (1998) WJ-R and Kaufman Adult Intelligence Test (KAIT: Kaufman & Kaufman, 1993) analyses, and joint WJ-R and DAS analyses (Laurie Ford, personal communication, August, 1998). McGrew and Flanagan (1998) recently synthesized the resulting *Gf-Gc* classifications for all the tests from all the major intelligence batteries based on Flanagan and McGrew's (1997) and McGrew's (1997) summaries of the extant *Gf-Gc* CB research.

The CB studies differ substantially from the within-battery structural validity studies described previously in a number of important ways. First, the Horn-Cattell *Gf-Gc* model of intelligence was the theoretical model used to organize each confirmatory factor analysis study. Second, all the data sets included tests from the WJ-R, a battery that has empirically validated indicators of eight broad *Gf-Gc* abilities (McGrew, 1994; McGrew et al., 1991; Reschly, 1990; Woodcock, 1990; Ysseldyke, 1990). As a result, each of the major intelligence batteries was analyzed together with a common set of empirically supported *Gf-Gc* test indicators from the WJ-R. That is, the *Gf-Gc* analyses of all the major intelligence batteries used the WJ-R *Gf-Gc*-designed battery as a common reference point. Thus, most of the broad *Gf-Gc* ability test classifications assigned by the authors of this textbook were based on empirical research studies. These *Gf-Gc* classifications provide evidence for the construct validity of the tests within individual intelligence batteries. A summary of these broad *Gf-Gc* classifications for nine major intelligence batteries is presented in Table 3.2. Table 3.2 represents an update of the broad ability classifications presented in McGrew and Flanagan (1998).

As summarized in Table 3.2, most intelligence batteries measure a rather limited range of *Gf-Gc* abilities. When using a criterion of at least two qualitatively different indicators (a cross-battery assessment guiding principle) for adequate construct representation (see Chapter 6 for a discussion), it is clear that most major intelligence batteries measure only *three* or *four* *Gf-Gc* abilities well. These data demonstrate that it is necessary to *cross* batteries (i.e., supplement an intelligence test with tests from another battery or other batteries) in order to obtain a comprehensive evaluation of cognitive function (e.g., when determining appropriate early intervention services; Wilson, 1992). The information presented in Table 3.2 is discussed further in Chapter 4.

Cross-Battery Structural Research Applied to the Wechsler Scales. The importance of the CB study approach to understanding the structural validity of all intelligence batteries, and the Wechsler Scales in particular, is demonstrated when a comparison is made between the CB study results and a typical within-battery exploratory factor analysis. For example, as summarized in Figure 3.5, three factors (VC, PO, FFD) typically were reported for the WISC-R (Kaufman, 1979). The previously mentioned three-factor structure has evolved more recently into a VC, PO, PS, and FFD/WM structure. The latter four-factor structure, which is based on within-battery structural validity research, contrasts with the *Gf-Gc* cross-battery-based interpretation of the Wechsler Scales also presented in Figure 3.5. A comparison of the within-battery and *Gf-Gc* cross-battery Wechsler interpretations reveals a number of important insights.

When considered within the context of *Gf-Gc* theory, the VC and PO factors are interpreted as measuring mainly *Gc* and *Gv* abilities, respectively (Carroll, 1993b; McGrew

TABLE 3.2 *Gf-Gc* Broad (Stratum II) Empirical Classifications of Intelligence Batteries (Based on Cross-Battery Studies Since 1990)

Gf-Gc Factor	WJ-R/III	Wechslers	SB:IV	DAS[*]	K-ABC[*]	KAIT
Long-Term Retrieval *(Glr)*	**Memory for Names** **Visual-Auditory Learning** **Delayed Recall-MN** Delayed Recall-VAL Retrieval Fluency (*Gs*) Rapid Picture Naming (*Gs*)	. . .	. . .	†	. . .	**Rebus Learning** **Delayed Recall Rebus Learning** Delayed Recall Auditory Comprehension
Short-Term Memory *(Gsm)*	**Memory for Words** **Numbers Reversed** Auditory Working Memory Memory for Sentences (*Gc*)	**Digit Span** **Letter-Number Sequence**	**Memory for Digits** Memory for Objects Memory for Sentences (*Gc*) Bead Memory (*Gv*)	**Recall of Designs**	**Number Recall** **Word Order** Hand Move (*Gq*)	**Memory for Block Design**
Processing Speed *(Gs)*	**Visual Matching** Cross Out (*Gv*) Decision Speed (*Gv*) Pair Cancellation	**Coding/Digit Symbol** **Symbol Search**	. . .	†	. . .	. . .
Auditory Processing *(Ga)*	**Incomplete Words** **Sound Blending** **Auditory Attention** **Sound Patterns**	. . .	. . .	. . .	. . .	. . .
Visual Processing *(Gv)*	**Spatial Relations** Picture Recognition Visual Closure	**Block Design** **Object Assembly** Mazes Picture Comprehension (*Gc*) Picture Arrangment (*Gc*)	**Pattern Analysis** Copying Paper Folding (*Gq*)	Pattern Construction	**Triangles** Gestalt Closure Spatial Memory Math Analogies (*Gf*) Photo Series (*Gf*)	. . .
Crystallized Intelligence *(Gc)*	**Picture Vocabulary** **Oral Vocabulary** **Verbal Comprehension** **General Information** **Oral Comprehension**	**Information** **Similarities** **Vocabulary** **Comprehension**	**Vocabulary** **Verbal Relations** Comprehension Absurdities	**Word Definitions** **Similarities**	**Faces and Places** **Riddles** Expressive Vocabulary	**Famous Faces** Definitions (*Grw*) Double Meanings (*Grw*) Auditory Comprehension (*Gsm*)
Fluid Reasoning *(Gf)*	**Analysis-Synthesis** **Concept Formation** Planning (*Gv*) Verbal Analogies (*Gc*)	**Matrix Reasoning**	**Matrices**	**Matrices** Seq-Quant Reasoning (*Gq*)	. . .	**Logical Steps** **Mystery Codes**
Quantitative Ability *(Gq)*	**Calculation** **Applied Problems** Math Concepts (*Gf*) Math Fluency *(Gs)*	**Arithmetic**	**Quantitative** **Number Series** **Equation Building**	†	**Arithmetic**	. . .

Note: Strong measures of *Gf-Gc* factors are reported in bold-faced type; measures not in bold type are moderate or mixed indicators of *Gf-Gc* abilities. Primary measures of *Gf-Gc* factors for the WJ-R are based on the empirical analyses of Woodcock (1990) and McGrew and Flanagan (1998); *Gf-Gc* factor classifications of the SB:IV, Wechslers, and K-ABC are reported in Woodcock (1990); classifications for the DAS and KAIT are reported in McGhee (1993) and Flanagan and McGrew (1998), respectively. For additional information on *Gf-Gc* factor classifications of major intelligence test batteries see McGrew and Flanagan (1998). *Gf-Gc* codes reported in parentheses next to subtest names identify the factor(s) on which the subtest has a secondary factor loading. These subtests are mixed measures of abilities. Information in this table was adapted from Flanagan and McGrew (1997) with permission from Guilford Press.

[*]Only a subset of DAS and K-ABC tests were joint factor analyzed by McGhee (1993) and Woodcock (1990), respectively. Therefore, only the tests that were included in these analyses are reported in this table. The *Gf-Gc* factor classifications of *all* DAS and K-ABC tests following a logical task analysis and expert consensus are reported in McGrew and Flanagan (1998).

†Tests measuring these abilities (i.e., *Glr, Gs, Gq*) are included on the DAS battery but have yet to be included in *Gf-Gc* organized confirmatory cross-battery factor analyses.

& Flanagan, 1996; 1998). It is also important to note that the FFD factor was not consistent with or similar to any broad ability within the *Gf-Gc* taxonomy. Is the taxonomy wrong, or was the FFD factor suspect?

Based on logical content analysis, it seems clear that the three WISC-R FFD tests (Arithmetic, Digit Span, Coding) are likely "loner" indicators of *Gq, Gsm,* and *Gs,* respectively. This hypothesis has been supported in *Gf-Gc* CB studies of the WISC-R and WJ-R (Woodcock, 1990) and the WISC-III and WJ-III (Phelps, Bowen, Chaco, Howard, Leahy, & Lucenti, 1999). For example, when the WISC-R tests were analyzed together with empirically validated indicators of a broad range of *Gf-Gc* abilities from the WJ-R (Woodcock, 1990; 1998), the results revealed that the traditional FFD tests loaded on three separate factors. The WISC-R FFD tests abandoned one another to "hang out" (i.e., correlate) with indicators of other *Gf-Gc* abilities with which they had more in common. Specifically, Arithmetic, Digit Span, and Coding loaded strongly on the *Gq, Gsm,* and *Gs* factors, respectively. Thus, in within-battery factor analysis studies, the so-called FFD factor appears to have been an invalid factor that consisted of tests that had no counterparts that measured similar cognitive ability constructs with which they could correlate to form valid factors (McGrew, 1999).

This conclusion is supported further by comparing the within-battery WISC-R and within-battery WISC-III analyses. With the addition of one new test (Symbol Search) in the third edition of the WISC, Coding abandoned its tried-and-true counterparts (Arithmetic and Digit Span) and formed a separate and, indeed, more appropriate *Gs* factor, as both Coding and Symbol Search involve perceptual speed and rate of test taking. Because the WISC-R FFD factor consists of tests that are indicators of *different Gf-Gc* abilities, it did not represent a valid theoretical cognitive construct. Given that two indicators (and preferably three or more) are needed to define a factor in factor analysis (Zwick & Velicer, 1986), it is not surprising that the Wechsler's loner tests loaded together on the so-called FFD factor. There simply were not enough indicators present in the within-battery WISC-R and WISC-III factor analyses (i.e., one or two more tests each of *Gq, Gsm,* and *Gs*) to identify clearly the separate abilities measured by the FFD tests. Furthermore, as is reflected in Figure 3.5, contemporary *Gf-Gc* research suggests that even the most recent recommended WAIS-III interpretation of the Arithmetic and Digit Span subtests as measures of working memory is suspect. *Gf-Gc* organized CB research suggests that Arithmetic and Digit Span subtests are primarily measures of *Gq* and *Gsm,* respectively. The ever changing nature of the original FFD factor, and the chameleon nature of Arithmetic and Digit Span in particular, demonstrate how valid test interpretation can be compromised when tests are developed and/or interpreted on the basis of a weak substantive foundation (McGrew, 1999).

The importance of the *Gf-Gc* CB study approach is also apparent when one examines the Wechsler PO tests. The PO tests have a long history of being interpreted as measures of *Gf* abilities (e.g., Kaufman, 1979; 1994). However, as summarized in Figure 3.5, cross-battery analyses with the WISC-R (Woodcock, 1990) and WISC-III (Phelps et al., 1999) show that the PO tests do not load on the *Gf* factor (defined by the WJ Analysis-Synthesis and Concept Formation tests). These results, as well as other exploratory cross-battery analyses of the Wechsler's with the DAS (Elliot, 1994; Stone, 1992) and KAIT (Kaufman & Kaufman, 1993), call into question the traditional *Gf* interpretation of the Wechsler PO tests (McGrew & Flanagan, 1996).

The recent addition of a test of *Gf* (i.e., Matrix Reasoning) to the WAIS-III, and hopefully a similar addition to the next versions of the WPPSI-R and WISC-III, is a step in the direction of better representation of the theoretical domain of intelligence. However, the inclusion of the Matrix Reasoning test in the PO index is a step backward in the empirical-theory match of the Wechsler index scores. As will be discussed in Chapter 4, matrix reasoning tests are strong measures of *Gf,* not *Gv*. In fact, matrix analogy or reasoning tests are considered to be one of (if not) the best indicators of *Gf* (Carroll, 1993a). The inclusion of the Matrix Reasoning test in the WAIS-III PO index results in the PO index now being comprised of measures of *Gv, Gf,* and to a lesser extent, *Gc.* We believe that this further muddying of the construct validity waters of the PO index is due to the lack of a strong theoretical or substantive foundation to the Wechsler program of validity research.

The Structural Validity of the Wechsler Scales: Concluding Comments

Systematic and strong programs of theory-based construct validity research are becoming increasingly important during the fourth wave of theory-based intelligence test interpretation. The *construct validity* of a test "is the extent to which the test may be said to measure a theoretical construct or trait" (Anastasi & Urbina, 1997, p.126). In the case of intelligence tests, the constructs of interest are the different cognitive abilities included in the theoretical domain of intelligence. The appropriate evaluation of the construct validity of tests requires the embedding of the constructs of interest in a conceptual framework (APA, 1985; Messick, 1995) and the examination of tests via research studies that include substantive, structural (internal), and external components. We believe that *Gf-Gc* theory is currently the best supported framework of cognitive abilities on which a strong program of test construct validation can be based (Bensen, 1998).

The primary conclusion reached from the preceding discussion is that *the Wechsler Intelligence Scales' continued allegiance to a largely practical and atheoretical model of intelligence limits the validity of the inferences that can be drawn from some of the Wechsler subtest and composite scores.* Figure 3.5 illustrated how the lack of a strong theory (which is the cornerstone of the substantive stage of validation) can result in the completion of structural or internal validity studies with a collection of tests that have significant content validity limitations, which in turn can significantly confound attempts to draw valid inferences from some of the test and index scores (e.g., Wechsler FFD tests). Although most of the internal structural validity studies reported for the Wechsler Scales have provided evidence for a within-battery factor structure of these instruments (see Appendix A), this evidence does not portray accurately their underlying theoretical constructs, as was evidenced through an evaluation of the individual Wechsler tests within a strong construct validation framework.

The Wechsler *Gf-Gc* cross-battery interpretations proposed by McGrew and Flanagan (1998) (summarized in the far right portion of Figure 3.5), interpretations which are based on structural analysis studies of a collection of measures (from many intelligence batteries) that collectively have strong substantive or content validity, provides a strong theory-based and construct-validated approach to interpreting the Wechsler Intelligence Scales. This is the beginning of the *Gf-Gc* empirical-theoretical domain match that is the foundation for the Wechsler Scale cross-battery interpretive approach described in subsequent chapters.

The External Stage of Construct Validation: A Conceptual Explanation

Although positive structural evidence is a necessary condition for establishing construct validity for tests, it does not meet the sufficient condition (Nunnally, 1978). The necessary and sufficient conditions for construct validity are both met when structurally valid measures demonstrate expected convergent and divergent relations with measures of constructs external to the focal measures (Benson, 1998; Benson & Hagtvet, 1996). This "looking outside" of the focal theoretical and empirical domains represents the *external* stage or component of a strong program of construct validity research.

As depicted in Figure 3.1, the external component of test and construct validation examines the nature and strength of significant relations between the focal theoretical constructs and a network of "other" constructs and measures "outside" of or external to the theoretical and empirical domains of primary interest. In the *Gv* example depicted in Figure 3.1, evidence for the construct validity of individual *Gv* tests (e.g., in the case of the Wechsler Scales, tests like Block Design and Object Assembly) or composite scores (e.g., Wechsler PO index) might take the form of moderate to high correlations between the focal *Gv* tests and composites and other *Gv* tests (e.g., SB:IV Pattern Analyses; WJ-R Spatial Relations) or composites (e.g., WJ-R Visual Processing Cluster) concurrent with low or nonsignificant relations with valid measures of distinctly different constructs (e.g., the WJ-R Auditory Processing Cluster consisting of the Sound Blending and Incomplete Words tests). Additional evidence might take the form of the finding that the *Gv* tests add important incremental information to the prediction of success in certain occupations that require strong visual-spatial skills (e.g., mathematics, sculpture, architecture). In other words, valid tests of cognitive constructs should (1) correlate significantly with "other" (external) tests of the same constructs; (2) either not correlate or correlate at low levels with tests of distinctly different constructs; and (3) make theoretically consistent predictions of a number of outcome variables (see Figure 3.2).

In light of the preceding discussion, ideally, all intelligence test authors and publishers should strive to achieve the comprehensive validity evidence. That is, they should provide substantive, structural, and external validity evidence that supports the recommended inferences that can be made from their individual intelligence test and composite scores. A review of Appendix A indicates that a significant body of research literature supports the external validity of the Wechsler Scales. This conclusion is similar to Zimmerman and Woo-Sam's (1997) recent synthesis of the criterion-related validity of the WISC-III. Zimmerman and Woo-Sam reported an average correlation (across 55 samples) of 0.75 between the WISC-III Full Scale score and 11 different ability measures. They also reported correlations ranging from 0.50 to 0.65 between the WISC-III Full Scale score and various achievement tests. A review of all the external validity evidence listed in Appendix A produces conclusions similar to those of Zimmerman and Woo-Sam.

The positive external validity studies reported for the Wechsler Scales would, on initial inspection, raise doubts about the validity of the claim that a strong substantive or theoretical foundation is required in order for an intelligence test to possess strong construct validity. In other words, how can the Wechsler Scales demonstrate significant external validity evidence in the absence of a strong substantive foundation? The answer lies in the fact that although positive external validity evidence has been reported, some of which provides strong evidence for some of the Wechsler tests and indexes, a portion of this evidence is not what it appears to be. Problems with the external validity evidence provided in support of the Wechsler Scales are discussed briefly below.

External Validity of Confounded Measures

Space limitations do not allow a detailed discussion and evaluation of the Wechsler external validity research presented in Appendix A. Many of the research studies reported in Appendix A provide external validity evidence for many of the Wechsler tests and scores. For example, consistent with the *Gf-Gc* outcome-criterion validity evidence summarized in Table 3.1, many of the studies listed in Appendix A report significant relations between the Wechsler VC index and other measures of verbal or *Gc* abilities. Furthermore, a number of these studies report that the Wechsler verbal subtests are significantly related to school achievement. Overall, considerable evidence supports the validity of inferences drawn from the tests comprising the Wechsler VC index. This is most likely due to the strong Wechsler VC-*Gc* empirical-theoretical domain match.

The same, however, cannot be said for the Wechsler FFD index—the primary example of poor construct validity research in this chapter. At first glance, the extant external validity research would appear to support the validity of the FFD index. Numerous studies have reported significant relations between the Wechsler FFD index and many external variables ranging from reading and math achievement to personality and emotional disturbance (Wielkiewicz, 1990). Even more impressive is Kaufman's (1994a) list of 15 different diagnostic categories of children who score low on the FFD tests. The list includes children with reading and learning disabilities, children with leukemia who received cranial irradiation therapy, children with epilepsy, and "believe it or not—normal children, especially girls, living in the Western part of the United States" (Kaufman, 1994a, p. 213), to name a few.

Although a diverse and impressive number of studies, Kaufman's (1994a) comment that the "FD subtests are like a land mine that explodes on a diversity of abnormal populations but leaves most normal samples unscathed" (p. 213) should serve as a reminder that the mere presence of significant relations between intelligence test scores and external variables is not always indicative of strong validity. In fact, patterns of significant correlations that do not "behave" in a manner consistent with theory may cast doubt on the construct validity of the test scores.

The indiscriminate nature of the FFD external evidence is difficult to explain. Just what is the nature of the construct underlying the FFD score that is related to both reading and math achievement, developmental language disorders, attention deficit disorder, schizophrenia, and Duchenne muscular dystrophy (Kaufman, 1994a)? According to Wielkiewicz (1990), "the studies available so far do not define any single construct that could connect this wide range of findings" (p. 93). When evaluated within a validity framework similar to the *Gv* example presented in Figure 3.1, the authors believe that the promiscuous statistical tendencies of the FFD index are due to a breakdown in the substantive and structural stages of construct validation. That is, the FFD factor is not a valid indicator of a valid theoretical construct within a valid theoretical model of intelligence (Carroll, 1993a). As stated earlier, the original FFD tests are best considered to be indicators of three separate *Gf-Gc* abilities (i.e., Arithmetic *[Gq]*; Digit Span *[Gsm]*; Coding *[Gs]*). The lack of a substantive theoretical foundation for the Wechsler tests has resulted in the post-hoc emergence of the FFD and its relations to a variety of disorders, the interpretations of which are similarly not grounded in any substantive theoretical domain. Therefore, it is not surprising that the external validity evidence for the FFD index is limited and confusing. In many ways a score on the FFD index is like a thermometer. Its wide range of significant external validity correlations suggests that it may help identify when something is amiss or atypical, but it contributes little in the way of explanation or prescription and, in fact, may result in the selection of the wrong interpretative hypotheses or subsequent recommendations (McGrew, 1999).

The meaning of the Wechsler PO index score, both from internal and external validity perspectives, also suffers from a poor theoretical-empirical domain match. For example, the current WISC-III PO index score, although predominately a measure of *Gv* abilities, also includes two tests (i.e., Picture Completion and Picture Arrangement) that the extant factor analytic research has consistently found to be measures of both *Gv* and *Gc*. Furthermore, although the addition of a measure of *Gf* (i.e., Matrix Reasoning) is a much welcomed improvement in the WAIS-III, the positive impact of this addition is minimized by its inclusion in the PO index. The result is a WAIS-III PO index that is a mixed measure of *Gv*, *Gc*, and *Gf* abilities. Although such a factorially complex measure may have important purposes in certain situations, factorial complexity makes it difficult to understand and interpret significant relations between the PO index and external validity criteria. Thus, significant correlations between the Wechsler PO index and other external criteria, notwithstanding the interpretation, is indeterminate.

Strong Construction Validation Models Do Make a Difference

Recently, Flanagan (1999) presented results from a series of causal models where the relations between the WISC-R factors and reading achievement factors were examined in a sample of 166 normal elementary school-aged subjects. Figure 3.6 summarizes the major findings from two causal models that were consistent with the atheoretical Wechsler model, a model that has been described in this chapter as being based on a "weak" program of construct validity research. The simple *g* model revealed a strong and significant relation between the WISC-R *g* score (i.e., Full Scale) and Reading (which included three different aspects of reading). The 0.64 structural path between the WISC-R *g* factor and Reading, a relation that indicates that the Wechsler *g* factor accounted for approximately 41 percent of the Reading factor variance, is strong external validity evidence for the WISC-R.

Even more impressive are the results of the WISC-R *g*+specific abilities causal model in the bottom half of Figure 3.6, a model that included, in addition to the *g*/Reading

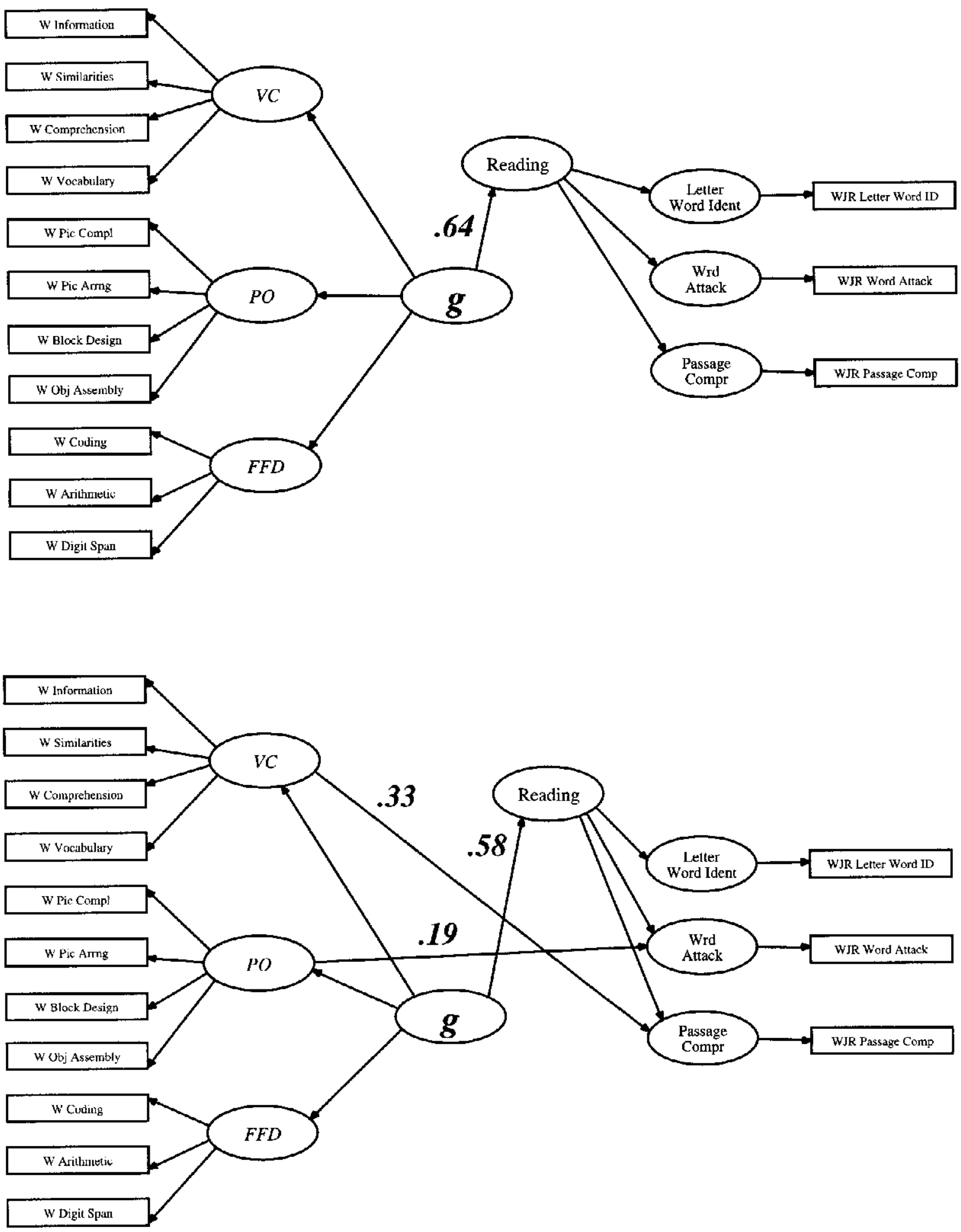

FIGURE 3.6 Example of External Validity Evidence for Atheoretical Within-Battery-Based Wechsler Factor Models: Summary of Flanagan's (1999) *g* and *g*+Specific Abilities and Reading Achievement Causal Models

Note: Factor loadings are omitted for readability purposes.

causal path, significant paths between specific cognitive abilities (i.e., VC, PO, and FFD) and specific reading abilities (i.e., Letter Word Identification, Word Attack, and Passage Comprehension)[3]. In addition to a significant *g*/Reading path (0.58), significant structural paths were identified between VC and Passage Comprehension (0.33) and PO and Word Attack (0.19). These results indicate that even though the WISC-R factor scores may be based on a weak atheoretical construct validity model, positive external validity evidence still exists for the WISC-R Full Scale and VC and PO scores. On initial inspection, these findings might argue against the central tenet of this chapter, namely, the need for measures of intelligence to be grounded in a strong program of construct validation. However, support for this central tenet is found in the significant improvement in external validity reported by Flanagan (1999) when the *Gf-Gc* cross-battery framework was superimposed on the WISC-R. These results are summarized in Figure 3.7.

Briefly, the Wechsler Scales can be "modernized" by taking the *Gf-Gc* interpretation of the Wechsler scales (as summarized in Figures 3.4 and 3.5) and "fleshing out" a more valid interpretation of the scales via the cross-battery approach (McGrew & Flanagan, 1998). As summarized in Figure 3.7, Flanagan (1999) used this approach and specified a causal model where the four-factor WISC-R model in Figure 3.6 was replaced with a *Gf-Gc* WISC-R cross-battery cognitive model (accomplished by supplementing the strong WISC-R *Gf-Gc* test indicators with strong WJ-R *Gf-Gc* test indicators). A number of important differences in the findings reported in Figure 3.6 and Figure 3.7 should be noted.

In the simple *g*/Reading model, the structural coefficient increased from 0.64 (Figure 3.6) to 0.81 (Figure 3.7), reflecting an increase in prediction/explanation of 24 percent of the total Reading variance (65% to 41%). A similar pattern is reported in the *g*+specific abilities models where a stronger *g*/Reading path was present for the WISC-R *Gf-Gc* cross-battery model (0.71) when compared to the simple WISC-R model (0.58). Also, the number and nature of the significant specific cognitive abilities to specific reading abilities paths changed. More importantly, the significant specific paths in the WISC-R *Gf-Gc* cross-battery model are more consistent with the extant *Gf-Gc* reading research literature (see Table 3.1).

In both figures the significant *Gc*/Passage Comprehension (0.42) and WISC-R VC/ Passage Comprehension (0.33) paths are consistent with the *Gc*/reading research reported in Table 3.1. In contrast, the significant WISC-R PO/Word Attack path (0.19) in Figure 3.6 is difficult to interpret in light of the extant research literature which provides little, if any, support for a significant relation between word attack skills and a factor (PO) that is primarily *Gv* in nature (see Table 3.1). The lack of a significant relation between the valid *Gv* factor and reading achievement in Figure 3.7 is more consistent with the research literature. A possible explanation, and one that would further support the tenet that the atheoretical basis of the Wechsler Scales has constrained and confounded the external validity evidence for the scales, is that the significant PO/Word Attack path (0.19) in Figure 3.6 may be due to the *Gc* variance present in the Wechsler Picture Completion and Picture Arrangement subtests.

Even more interesting is the identification of the significant *Gs*/Passage Comprehension (0.14) and *Ga*/Word Attack (0.26) paths in Figure 3.7 and their absence in Figure 3.6 (WISC-R data). The significant *Ga*/Word Attack path is consistent with research that has demonstrated a strong relation between phonological awareness and beginning reading (see

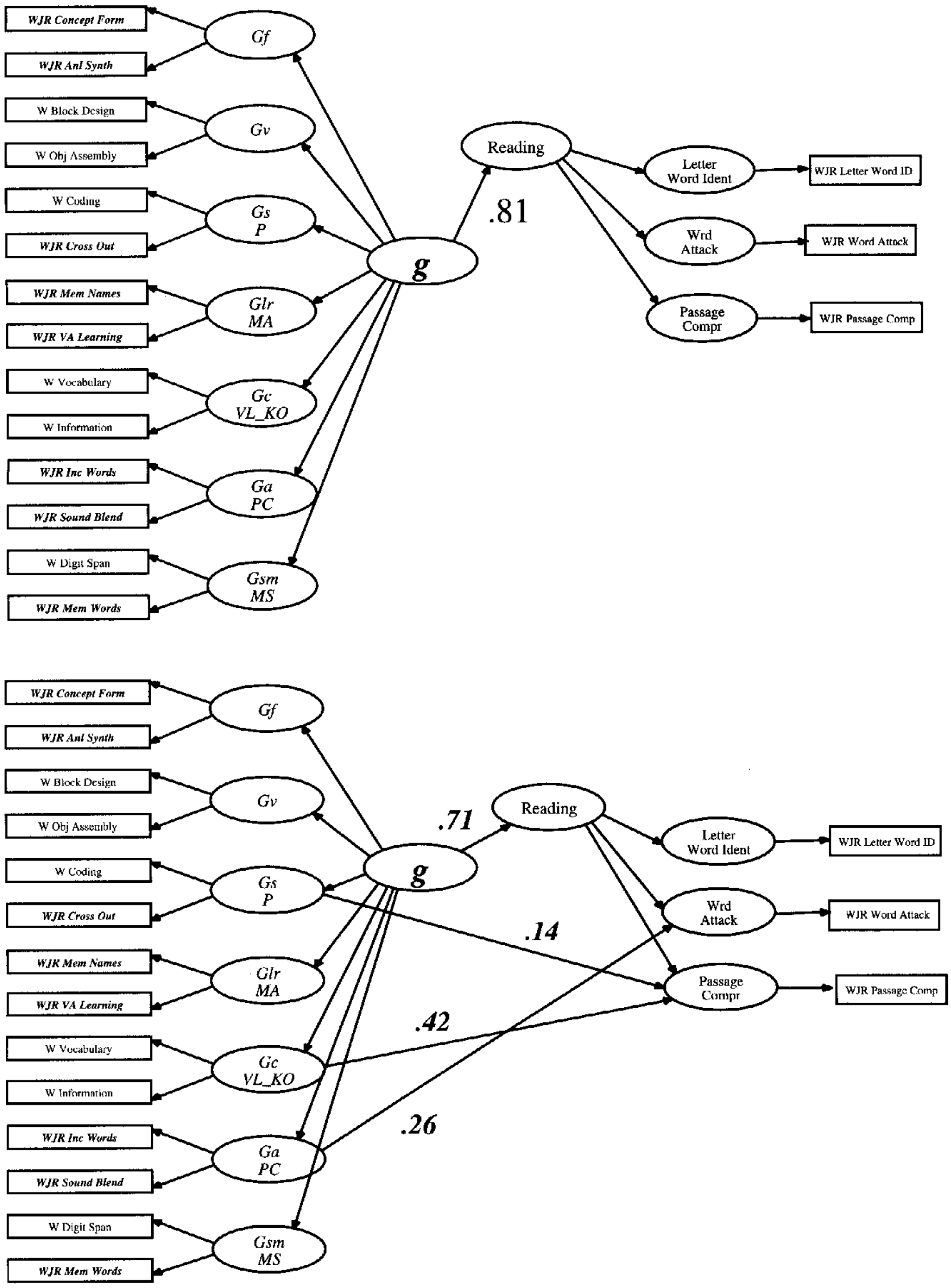

FIGURE 3.7 Example of External Validity Evidence for *Gf-Gc* Cross-Battery-Based Wechsler Factor Models: Summary of Flanagan's (1999) *g* and *g*+Specific Abilities and Reading Achievement Causal Models

Note: Factor loadings are omitted for readability purposes.

Table 3.1). This path did not surface in the traditional WISC-R model (Figure 3.6) simply because the Wechsler Scales do not include measures of this construct (see Figures 3.4 and 3.5, and Table 3.2), a situation that is largely due, in our judgement, to the limited attention paid to the substantive stage of construct validation in the development and revision of the Wechsler Scales.

Finally, the absence of a significant *Gs*/Passage Comprehension path in Figure 3.6 and the presence of a significant *Gs*/Passage Comprehension path (0.14) in Figure 3.7, is most likely due to the previously discussed problems with the WISC-R FFD factor. Contemporary research has shown that the WISC-R Coding test is a strong indicator of one aspect of *Gs* (see Table 3.2). However, as seen in Figure 3.6, it's potential contribution to prediction/explanation of reading is degraded in the WISC-R via its combination of two other tests (i.e., Arithmetic and Digit Span) that measure other cognitive abilities (i.e., *Gq* and *Gsm*). Fortunately this situation has been rectified in the WISC-III and WAIS-III through the addition of another measure of *Gs* (i.e., Symbol Search), a revision that has made possible the emergence of a separate and valid *Gs* factor index in these two batteries.

External Validity and the Wechsler Scales: Concluding Comments

In summary, it should be clear from the difficulties associated with the interpretation of the confounded Wechsler FFD factor-based index and the research results reported by Flanagan (1999) that a number of conclusions regarding the external validity of certain aspects of the Wechsler Scales may be inaccurate. Although significant correlations have been reported between the Wechsler FFD and PO indexes and important outcome criteria (e.g., school achievement), the inferences that can be drawn from these correlations are suspect due to the confounded (factorially complex) nature of these index scores. Furthermore, the failure to include indicators of cognitive abilities (e.g., *Ga* and *Glr*) that substantive analysis has identified as important abilities to include in an intelligence test that is used to predict and understand school achievement, reveals a significant limitation of the atheoretical Wechsler Scales. Finally, the significant improvement in the prediction/explanation of total reading (24 percent increase in reading variance explained) which occurred when the WISC-R substantive and structural shortcomings were "ameliorated" via a *Gf-Gc* cross-battery approach, suggests that a more substantive theory-based approach to constructing and interpreting test batteries can result in stronger external validity.

The remainder of this book outlines a theoretically and empirically based Wechsler *Gf-Gc* cross-battery approach that provides the Wechsler Scales with a much needed substantive foundation, which, in our judgement, will improve the internal and external validity of intellectual assessments that are based on the Wechsler Scales. Ultimately, this approach will provide a viable means of propelling the Wechsler Scales into the theory-based fourth wave of intelligence test interpretation.

Conclusion

"Those who are enamoured of the practice without science are like a pilot who goes into a ship without a rudder or compass and never has any certainty where he is going. Practice should always be based on a sound knowledge of theory" (Leonardo da Vinci). Similarly, we believe that the applied science of intelligence testing must be based on solid empirical and theoretical knowledge. This stated belief, which is a cornerstone of this entire book, should not be misconstrued as an indictment of the original Wechsler Intelligence Scales. Historical hindsight is 20/20. It must be kept in mind that David Wechsler's original test was intentionally grounded in practical and clinical considerations rather than theoretical deliberations. Also, only portions of the *Gf-Gc* terrain were emerging at the time when David Wechsler was developing his first scales. The original Verbal and Performance factors made sense at the time and were quite valuable and useful.

In order to improve the validity of inferences drawn from the various Wechsler scores, it is important to understand that "that was then and this is now." The recent application of *Gf-Gc*-organized cross-battery factor analyses to the major intelligence batteries (McGrew & Flanagan, 1998) has produced a closer correspondence between the empirical and theoretical domains of intelligence. The left and right sides of Figure 3.5 represent the *then* and *now* of Wechsler Scale test interpretation, respectively. An even more refined *now* *Gf-Gc* interpretation of the Wechsler Scales was presented in Figure 3.4.

The imposition of the *Gf-Gc* framework and the breaking down of the secular within-battery confines of the Wechsler Intelligence Scales (i.e., adopting a cross-battery perspective) reflects a necessary post-hoc attempt to shore up the weak substantive foundation of the Wechsler interpretative system. Research presented in this chapter demonstrated how the lack of a strong substantive foundation has undermined certain aspects of the internal and external validity of the Wechsler Scales. Furthermore, research showed that the imposition of a valid substantive foundation (i.e., contemporary *Gf-Gc* theory) can significantly improve this state of affairs. The empirical-theoretical construct mapping presented in Figure 3.4 is intended to "right" the Wechsler construct validity ship. It is our belief that the imposition of a strong substantive framework to the interpretation of the Wechsler Intelligence Scales via McGrew and Flanagan's (1998) *Gf-Gc* cross-battery approach will result in a greater alignment of the Wechsler *Gf-Gc* empirical-theoretical domains. The end result should be the derivation of more valid inferences from Wechsler test scores.

ENDNOTES

1. For a detailed explanation of this table, see McGrew and Flanagan (1998).
2. At least two of the three authors classified every study according to the criteria summarized in Figure 3.2. In addition to the authors' classifications, two upper-level doctoral students in a clinical psychology program classified every study included in Appendix A. This process resulted in four ratings for each study. In the few instances in which disagreement was found among the raters, the consensus reached by the authors constituted the final classification. The positive (+) and negative (–) signs throughout Appendix A indicate whether the individual investigations provided or failed to provide validity evidence for the Wechsler Scales, respectively. A review of this Appendix demonstrates a lack of substantive validity support for the Wechsler Scales.
3. The rationale and methods used by Flanagan (1999) mirror those that are described in detail in McGrew, Flanagan, Keith, and Vanderwood (1997). The reader is directed to this latter source for specific details of this type of methodology.