STRUCTURAL AND INCREMENTAL VALIDITY OF THE FULL-SCALE IQ: DOES SCATTER MATTER?

by

MARY HALL SLAUGHTER

(Under the Direction of Jennifer H. Lindstrom)

ABSTRACT

Since the development of modern cognitive assessments, psychologists have been instructed by assessment "experts" and taught in graduate training programs that significant variability among index scores, or scatter, renders the general intelligence score unsound for clinical interpretation and diagnostic decision-making. The existing empirical research, however, challenges the validity of this interpretive heuristic known as the variability hypothesis. To investigate the credibility of the variability hypothesis, the present study examined the impact of significant factor score variability on the structural and incremental validity of the WAIS-IV with a sample of 1,211 individuals referred for evaluation at a university-based clinic. The sample included those both with (n = 616) and without (n = 595) significant variability in their WAIS-IV index scores. Exploratory factor analysis with multiple factor extraction criteria supported the extraction of four factors consistent with the WAIS-IV theoretical structure for the Scattered Group. The higher-order general factor also accounted for the largest portions of total and common variance. Incremental validity analyses indicated that the FSIQ accounted for large and statistically significant effects across measures of academic achievement for the Scattered Group. Although the four first-order factors combined accounted for significant incremental

predictive contributions, individual WAIS-IV factor scores contributed trivial amounts of achievement variance beyond the FSIQ. Implications for clinical practice and the interpretation of the WAIS-IV in the presence of significant scatter are discussed.

INDEX WORDS: cognitive scatter, factor score variability, structural validity, incremental validity

STRUCTURAL AND INCREMENTAL VALIDITY OF THE FULL-SCALE IQ: DOES SCATTER MATTER?

by

MARY HALL SLAUGHTER

B.S., University of Georgia, 2017

B.A., University of Georgia, 2017

M.A., University of Georgia, 2019

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2023

STRUCTURAL AND INCREMENTAL VALIDITY OF THE FULL-SCALE IQ: DOES SCATTER MATTER?

by

MARY HALL SLAUGHTER

Major Professor:	Jennifer Lindstrom
Committee:	Amy Reschly
	Jason Nelson
	Scott Ardoin

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2023

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Assessment has been central to the role of school psychologists since the inception of the field. Estimates suggest that psychologists spend about half of their time engaged in assessment-related activities (e.g., Benson et al., 2019; Farmer, Goforth, et al., 2021; Hutton et al., 1992; Sotelo-Dynega & Dixon, 2014), with cognitive tests topping the list of instruments used most frequently by psychologists (Goh et al., 1981; Hutton et al., 1992; Reschly et al., 1987; Sotelo-Dynega & Dixon, 2014). Cognitive tests (also referred to as intelligence tests or measures) are administered as a part of comprehensive evaluations for individuals of all ages, most commonly for school-aged children and adolescents with known or suspected cognitive and learning challenges. Within this context, cognitive test scores are used by psychologists to help inform diagnostic decisions, special education eligibility, and treatment and intervention plans. Despite their widespread use, however, there has been a long-standing controversy regarding how psychologists should interpret cognitive measures.

Initially, interpretation of intelligence tests was limited to evaluating individual performance at the general intelligence level, using the global composite score, or the full-scale intelligence quotient (FSIQ; McGill, 2013). As theories of cognitive abilities developed and methods for factor analytic modeling advanced, subsequent versions of intelligence tests provided additional composite scores measuring cognitive domains (e.g., verbal knowledge, visual-spatial skills), along with the overall FSIQ score. During the second half of the 20th century and into the early 2000s, scholars (e.g., Kaufman, 1979; Sattler, 2008) developed

cognitive profile analysis methods that encouraged users to focus their interpretive weight on the pattern of first-order scores (e.g., composites/indexes, subtests), thus deemphasizing the FSIQ (McGill, 2018). Over the last two decades, however, numerous investigations (e.g., Canivez et al., 2017; Fletcher & Miciak, 2017; Watkins et al., 2005) have challenged long-standing assumptions about the utility of cognitive profile analysis, revealing that these methods are lacking in reliability, validity, and diagnostic and treatment utility. These researchers recommend against using cognitive profile analysis as a focal point for diagnostic and treatment decisions, and instead encourage psychologists to place their interpretive emphasis on the global FSIQ score (Dombrowski et al., 2022; Macmann & Barnett, 1997; McGill et al., 2018; Watkins, 2000). Whereas this continues to be the prevailing recommendation based on empirical research, there is an unwavering lack of consensus by scholars and psychologists regarding best cognitive assessment practices.

To add further complication, there is even greater controversy regarding cognitive test interpretation in more unique situations, such as when an examinee presents with significant variability among index scores. Within the intelligence testing literature (e.g., Groth-Marnat & Wright, 2016; Hale & Fiorello, 2004; Sattler, 2018), it is commonly suggested that significant variability among index scores, or *scatter*, renders the FSIQ an invalid measure of overall ability. When this occurs, psychologists are encouraged to forego interpretation of the FSIQ—a heuristic coined by McGill (2016) as the *variability hypothesis*. The existing empirical research, however, challenges the validity and utility of these methods.

In one of the earliest investigations of the variability hypothesis, Fiorello et al. (2002) examined the structure of the Weschler Intelligence Scales for Children, Third Edition (WISC-III; Weschler, 1991) FSIQ in a non-clinical sample of children with flat and scattered cognitive

profiles. Whereas it was found that the FSIQ for the flat subsample was primarily composed of shared variance, the FSIQ for the scattered subsample was primarily composed of unique variance. Based on these results, Fiorello et al. (2002) concluded that the FSIQ does not adequately represent global intellectual ability for individuals with significant levels of scatter. These results were later replicated with a sample of children with disabilities (Fiorello et al., 2007). However, several scholars criticized the methodology and conclusions reached by Fiorello and colleagues (Dana & Dawes, 2007; Daniel, 2007; Schneider, 2008), conducting their own investigations which found that the FSIQ remained a valid estimation of global ability for individuals with significant scatter.

In a more recent study, McGill (2016) examined the structural validity of the Kaufman Assessment Battery for Children, Second Edition (KABC-II; Kaufman & Kaufman, 2004a) for participants in the normative sample who demonstrated significant variability in their index scores. An exploratory factor analysis with multiple factor extraction criteria revealed that the five-factor structure upheld even in the presence of significant factor score variability. Given these findings, McGill concluded that the FSIQ maintains structural validity despite significant scatter. These results were later replicated with scattered and mixed comparison populations (Ayoub, 2020).

Similarly, scholars have also investigated the impact of significant factor score variability on the incremental validity of cognitive tests. In a pioneering study, Watkins et al. (2007) examined the predictive validity of the WISC-III FSIQ in clinical and mixed comparison subsamples with and without factor score variability. Findings indicated that the FSIQ was a robust predictor of scores on the Wechsler Individual Achievement Test (WIAT; Wechsler, 1992) for both the typical and clinical subsamples. As a result, the authors challenged the

practice of discounting the FSIQ score as a predictor of academic achievement when index scores vary significantly. These results were replicated in subsequent investigations using other cognitive and academic tests (Freberg et al., 2008; Kotz et al., 2008; McGill, 2013).

More recently, McGill (2016) examined the incremental validity of the KABC-II global ability score in a normative sample with significant factor score variability. Hierarchical multiple regression revealed that the FSIQ accounted for large predictive effects of scores on the Kaufman Test of Educational Achievement, Second Edition (KTEA-II; Kaufman & Kaufman, 2004b), whereas the incremental variance (after controlling for the effects of the FSIQ) accounted for by the KABC-II index scores was consistently trivial. Given these findings, McGill concluded that the FSIQ remains a valid predictor of academic achievement in the presence of significant scatter.

Overall, previous studies have consistently found that the global ability score maintains its structural and incremental validity despite significant factor score variability (e.g., Ayoub, 2020; Daniel, 2007; Kotz et al., 2008; McGill, 2013, 2016; Watkins et al., 2007). Contrary to the variability hypothesis, these findings suggest that the FSIQ should not automatically be disregarded in the presence of significant scatter. Despite the overwhelming empirical evidence, surveys continue to reveal that psychologists frequently place greater interpretive weight on the scatter of index scores (Farmer, McGill, et al., 2021; Kranzler et al., 2020; Pfeiffer et al., 2000) and these interpretive techniques continue to be encouraged in many training programs (Decker et al., 2013; Lockwood & Farmer, 2020) and in technical manuals and interpretive guidebooks (e.g., Groth-Marnat & Wright, 2016; Sattler, 2018; Weschler, 2008). One likely reason that these interpretive techniques continue to prevail in clinical practice is the limited amount of relevant literature. Whereas previous studies have examined the impact of scatter on structural validity

(e.g., Ayoub, 2020; Daniel, 2007; Fiorello et al., 2007), and others have examined the incremental validity of scores in the presence of significant factor score variability (e.g., Freberg et al., 2008; McGill, 2013; Watkins et al., 2007), only one study to date (McGill, 2016) was designed specifically to investigate the potential impact of scatter on the structural validity *and* the incremental validity of cognitive measures.

McGill's (2016) study, like most of the studies that examined the impact of scatter on either structural or incremental validity, used an archived standardization sample. Further research is needed to determine if previous findings generalize to clinical populations, as only two previous studies used a clinical sample (Freberg et al., 2008; Watkins et al., 2007). It is also important to conduct this research using other commonly used intelligence measures, such as the Weschler scales. Additionally, the existing research in this area primarily focuses on children and adolescents, indicating a need for research with postsecondary populations.

Purpose of the Current Study

To address these gaps in the literature, the present study aimed to investigate the impact of significant factor score variability on the structural and incremental validity of the Weschler Adult Intelligence Scales, Fourth Edition (WAIS-IV; Weschler, 2008) in a clinically referred sample of college students. Given that users of the WAIS-IV are encouraged to invalidate the global ability score if significant variability is observed among the index scores, additional information is needed regarding the psychometric validity of this interpretative heuristic known as the *variability hypothesis* (McGill, 2016). As such, the purpose of the current study was to determine whether significant variability in the WAIS-IV index scores renders the FSIQ an invalid measure of overall ability and thus erroneous for clinical interpretation. Put simply, this

study served to answer the question that frequently arises for many psychologists, “does scatter *really* matter?”

Chapter 2 provides a comprehensive review of the literature relevant to the current study, as well as an overview of the present study’s purpose and research question. Chapter 3 outlines the sample, measures, procedure, and data analyses for the current study. Chapter 4 details the results of the research questions, followed by a discussion of the results in Chapter 5. Chapter 5 also address the limitations of the current study and suggestions for future research and discusses the implications of the present study’s results for clinical practice.

REFERENCES

Ayoub, J. L. (2020). Does cognitive scatter attenuate the validity of the global intelligence score? (Publication No. 28091905) [Doctoral dissertation, Hofstra University]. ProQuest Dissertations & Theses Global.

Benson, N. F., Floyd, R. G., Kranzler, J. H., Eckert, T. L., Fefer, S. A., & Morgan, G. B. (2019). Test use and assessment practices of school psychologists in the United States: Findings from the 2017 National Survey. *Journal of School Psychology*, *72*, 29–48. https://doi.org/10.1016/j.jsp.2018.12.004

Canivez, G. L. (2013). Psychometric versus actuarial interpretation of intelligence and related aptitude batteries. In D. H. Saklofske, C. R. Reynolds, & V. L. Schwean (Eds.). *The Oxford handbook of child psychological assessment* (pp. 84–112). Oxford University Press.

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2017). Structural validity of the Wechsler Intelligence Scale for Children-Fifth Edition: Confirmatory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment*, *29*(4), 458–472. https://doi.org/10.1037/pas0000358

Dana, J., & Dawes, R. (2007). Comment on Fiorello et al. "Interpreting intelligence test results for children with disabilities: Is global intelligence relevant?" *Applied Neuropsychology*, *14*(1), 21–25. https://doi.org/10.1080/09084280701280379

Daniel M. H. (2007). "Scatter" and the construct validity of FSIQ: Comment on Fiorello et al. (2007). *Applied Neuropsychology*, *14*(4), 291–295. https://doi.org/10.1080/09084280701719401

Decker, S. L., Hale, J. B., & Flanagan, D. P. (2013). Professional practice issues in the assessment of cognitive functioning for educational applications. *Psychology in the Schools*, *50*, 300–313. https://doi. org/10.1002/pits.21675

Dombrowski, S. C., McGill, R. J., Watkins, M. W., Canivez, G. L., Pritchard, A. E., & Jacobson, L. A. (2022). Will the real theoretical structure of the WISC-V please stand up? Implications for clinical interpretation. *Contemporary School Psychology*, *26*, 492–503. https://doi.org/10.1007/s40688-021-00365-6

Farmer, R. L., Goforth, A. N., Kim, S. Y., Naser, S. C., Lockwood, A. B., & Affrunti, N. W. (2021). Status of school psychology in 2020, part 2: Professional practices in the NASP

membership survey. *NASP Research Reports*, *5*(3), 1–17.

Farmer, R. L., McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2021). Why questionable assessment practices remain popular in school psychology: Instructional materials as pedagogic vehicles. *Canadian Journal of School Psychology*, *36*(2), 98–114. https://doi.org/10.1177/0829573520978111

Fiorello, C. A., Hale, J. B., Holdnack, J. A., Kavanagh, J. A., Terrell, J., & Long, L. (2007). Interpreting intelligence test results for children with disabilities: Is global intelligence relevant? *Applied Neuropsychology*, *21*, 2–12. https://doi.org/10.1080/09084280701280338

Fiorello, C. A., Hale, J. B., McGrath, M., Ryan, K., & Quinn, S. (2002). IQ interpretation for children with flat and variable test profiles. *Learning and Individual Differences*, *13*, 115–125. https://doi.org/10.1016/S1041-6080(02)00075-4

Fletcher, J. M., & Miciak, J. (2017). Comprehensive cognitive assessments are not necessary for the identification and treatment of learning disabilities. *Archives of Clinical Neuropsychology*, *32*(1), 2–7. https://doi.org/10.1093/arclin/acw103

Freberg, M. E., Vandiver, B. J., Watkins, M. W., & Canivez, G. L. (2008). Significant factor score variability and the validity of the WISC-III Full Scale IQ in predicting later academic achievement. *Applied neuropsychology*, *15*(2), 131–139. https://doi.org/10.1080/09084280802084010

Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. *The Journal of Special Education*, *40*, 103–114. https://doi.org/10.1177/00224669060400020101

Goh, D. S., Teslow, J., & Fuller, G. B. (1981). The practice of psychological assessment among school psychologists. *Professional Psychology*, *12*, 696–706. https://doi.org/10.1037/0735-7028.12.6.696

Groth-Marnat, G., & Wright, A. J. (2016). *Handbook of psychological assessment* (6th ed.). Wiley.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook*. Guilford.

Hutton, J. B., Dubes, R., & Muir, S. (1992). Assessment practices of school psychologists: Ten years later. *School Psychology Review*, *21*, 271–284.

Kaufman, A. S. (1979). *Intelligent testing with the WISC-R*. Wiley.

Kaufman, A. S., & Kaufman, N. L. (2004a). *Kaufman Assessment Battery for Children* (2nd ed.). American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004b). *Kaufman Test of Educational Achievement* (2nd ed.). American Guidance Service.

Kotz, K. M., Watkins, M. W., & McDermott, P. A. (2008). Validity of the General Conceptual Ability score from the Differential Ability Scales as a function of significant and rare interfactor variability. *School Psychology Review*, *37*(2), 261–278.

Kranzler, J. H., & Floyd, R. G. (2013). *Assessing intelligence in children and adolescents: A practical guide*. Guilford Press.

Kranzler, J. H., Maki, K. E., Benson, N. F., Eckert, T. L., Floyd, R. G., & Fefer, S. A. (2020). How do school psychologists interpret intelligence tests for the identification of specific learning disabilities? *Contemporary School Psychology*. https://doi.org/10.1007/s40688-020-00274-0

Lockwood, A. B., & Farmer, R. L. (2020). The cognitive assessment course: Two decades later. *Psychology in the Schools*, *57*(2), 265–283. https://doi.org/10.1002/pits.22298

Macmann, G. M., & Barnett, D. W. (1997). Myth of the master detective: Reliability of interpretations for Kaufman's "intelligent testing" approach to the WISC-III. *School Psychology Quarterly*, *12*, 197–234. https://doi.org/10.1037/h0088959

McGill, R. J. (2013). Beyond g: Assessing the incremental validity of the Cattell-Horn-Carroll (CHC) broad ability factors on the Woodcock-Johnson III Tests of Cognitive Abilities (Publication No. 3621595) [Doctoral dissertation, Chapman University]. ProQuest Dissertations & Theses Global.

McGill, R. J. (2016). Invalidating the full scale IQ score in the presence of significant factor score variability: Clinical acumen or clinical illusion? *Archives of Assessment Psychology*, *6*(1), 49–79.

McGill, R. J. (2018). Confronting the base rate problem: More ups and downs for cognitive scatter analysis. *Contemporary School Psychology*, *22*, 384–393. https://doi.org/10.1007/s40688-017-0168-4

McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2018). Cognitive profile analysis in school psychology: History, issues, and continued concerns. *Journal of School Psychology*, *71*, 108–121. https://doi.org/10.1016/j.jsp.2018.10.007

Pfeiffer, S. I., Reddy, L. A., Kletzel, J. E., Schmelzer, E. R., & Boyer, L. M. (2000). The practitioner's view of IQ testing and profile analysis. *School Psychology Quarterly*, *15*, 376–385. https://doi.org/10.1037/ h0088795

Reschly, D. J., Genshaft, L., & Binder, M. S. (1987). *The 1986 NASP survey: Comparison of practitioners, NASP leadership, and university faculty on key issues* (ED300733). ERIC. https://files.eric.ed.gov/fulltext/ED300733.pdf

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations* (5th ed.). Sattler Publishing.

Sattler, J. M. (2018). *Assessment of children: Cognitive foundations and applications* (6th ed.). Sattler Publisher.

Schneider W. J. (2008). Playing statistical Ouija board with commonality analysis: Good questions, wrong assumptions. *Applied Neuropsychology*, *15*(1), 44–53. https://doi.org/10.1080/09084280801917566

Sotelo-Dynega, M., & Dixon, S. G. (2014). Cognitive assessment practices: A survey of school psychologists. *Psychology in the Schools*, *51*(10), 1031–1045. https://doi.org/10.1002/pits.21802

Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. *School Psychology Quarterly*, *15*(4), 465–479. https://doi.org/10.1037/h0088802

Watkins, M. W., Glutting, J. J., & Lei, P.-W. (2007). Validity of the Full-Scale IQ when there is significant variability among WISC-III and WISC-IV factor scores. *Applied Neuropsychology*, *14*(1), 13–20. https://doi.org/10.1080/09084280701280353

Watkins, M. W., Glutting, J. J., & Youngstrom, E. A. (2005). Issues in subtest profile analysis. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 251–268). Guilford Press.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children* (3rd ed.). Psychological Corporation.

Wechsler, D. (1992). *Wechsler Individual Achievement Test*. Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). Pearson.

CHAPTER 2

LITERATURE REVIEW

Since the inauguration of intelligence testing, scholars have debated how cognitive measures should be interpreted in clinical practice. Specifically, some scholars argue that the global ability score (e.g., the Full Scale Intelligence Quotient [FSIQ]) is the most parsimonious and valid predictor of important life outcomes such as achievement (e.g., Canivez, 2013b; Dombrowski & Gischlar, 2014; Gottfredson, 1997; McGill, 2016), whereas others suggest that the profile of first-order factor scores provides more useful information for diagnostic decision-making (e.g., Feifer et al., 2014; Fiorello et al., 2007; Hale & Fiorello, 2001). To add to this, there is even more contention regarding score interpretation when significant factor score variability, or *scatter*, is observed.

Since the development of modern intelligence tests, psychologists have been taught in training programs (Decker et al., 2013; Lockwood & Farmer, 2020) and encouraged by technical manuals and interpretive guidebooks (e.g., Groth-Marnat & Wright, 2016; Miller et al., 2016; Sattler, 2018) to exercise caution when an examinee presents with significant variability in index scores. Specifically, psychologists are encouraged to forego interpretation of the FSIQ in the presence of significant scatter—an interpretive heuristic known as the *variability hypothesis* (McGill, 2016). The existing empirical literature, however, challenges the validity and utility of these methods. Namely, research has consistently revealed that the global ability score remains as the most valid score, even in the presence of significant factor score variability (e.g., Daniel, 2007; Freberg et al., 2008; McGill, 2016; Watkins et al., 2007). Challenging what many

psychologists have been trained to believe, this finding leaves them speculating, does scatter *really* matter?

In order to explore this question through our own study, it is important to first review the relevant literature. As such, following is a review of the history of cognitive profile analysis, the contemporary methods for interpreting intelligence tests, and the psychometrics of intelligence test interpretation. Next is an overview of significant factor score variability, followed by a review of studies examining the impact of scatter on structural and incremental validity.

History of Cognitive Profile Analysis

A cognitive profile refers to an individual's set of scores (i.e., global ability, first-order factor, and subtest scores) derived from a standardized assessment of their cognitive ability (Ayoub, 2020). Cognitive profile analysis, then, refers to the practice of making inferences about an individual's strengths and weaknesses based on their profile of scores (McGill et al., 2018). Despite the amount of evidence that challenges the utility of these procedures, cognitive profile analysis remains common in school psychology training and practice (Bray et al., 1998). Bray et al. (1998) state that this may be because school psychologists are "simply unfamiliar with the literature" refuting these methods (p. 209). Others such as McGill et al. (2018) note that this is due, at least in part, to the number of technical manuals and texts devoted to cognitive assessment that encourage profile analysis.

Regardless, cognitive profile analysis is still popular among school psychologists (Benson et al., 2020). In order to understand why such procedures may be inadequate, as well as how cognitive measures should be interpreted, a review of previous theories and interpretive methods is necessary to provide historical context for contemporary debates within the field of school psychology.

Origin of Profile Analysis

Although the exact origin of profile analysis is difficult to pinpoint, initial uses are thought to have been for clinical purposes. That is, early researchers hypothesized that scatter within a cognitive profile was a useful predictor of pathology (Harris & Shakow, 1937). In line with this was the assumption that typically developing individuals had a stable, or flat, cognitive profile. Interest in cognitive profile analysis is thought to have increased with the publication of the first Wechsler-Bellevue Intelligence Scale (Wechsler, 1939) because the second edition, published in 1941, included a new chapter guiding psychologists on how to make diagnostic decisions based on patterns of scores (Ayoub, 2020).

Along with the rise in intelligence testing came personality assessments, which were commonly used and influenced by psychologists of a psychodynamic orientation (Ayoub, 2020; Beaujean & Benson, 2019). In particular, Rapaport et al. (1945) were early pioneers in using cognitive profiles as a diagnostic indicator, and they proposed a comprehensive process for interpreting intraindividual cognitive strengths and weaknesses for diagnosing psychiatric disorders. Their system used psychodynamic principles and was based upon the idea that variability in cognitive test performance served as a marker for the presence of behavioral pathology.

Despite the increasing popularity of cognitive assessment for diagnostic decision making, research examining its utility indicated unfavorable results (Cohen, 1952; Rabin & Guertin, 1951). By the early 1970s, psychoanalytic theory was no longer the leader in the world of cognitive assessment (Ayoub, 2020).

Intelligence Testing

Following the decline of psychoanalytic theory came the rise of cognitive and behavioral theories. As a result, cognitive profile analysis grew even more popular than ever before. Rather than using it to diagnose psychiatric disorders, however, profile analysis served a new purpose—to diagnose cognitive disorders (Ayoub, 2020). A pioneer in the use of cognitive profile analysis, Kaufman (1979) developed a method that combined clinical and psychometric approaches to cognitive test interpretation. This method, known as Intelligence Testing (IT), became the foundation of modern intelligence test interpretation in clinical practice (McGill, 2018).

Psychologists implementing the IT method were advised to interpret cognitive profiles in a stepwise, or top-down, fashion, starting with the FSIQ and ending at the subtest level (McGill et al., 2018). Test users were also encouraged to make inferences regarding the patterns of scores at all levels of the test. Interpretation of the FSIQ was deemphasized; instead, greater focus and interpretive weight was placed on the scatter of index and subtest scores (McGill, 2018). Kaufman (1994) and other proponents of the IT approach deemed that interpreting significant scatter had a number of potential implications for aiding with diagnostic decisions and treatment planning.

Of course, the IT approach was not without its critics. Namely, McDermott et al.'s (1990) review of the literature found little empirical support for the interpretation of cognitive profiles, concluding that test users should "just say no" to practices involving subtest interpretation. Other scholars argued that the IT approach had poor reliability, and thus, little clinical utility. Notably, several researchers found that the diagnostic accuracy of cognitive profiles rarely exceeded chance levels, rendering them inapt for educational decisions (e.g., Smith & Watkins, 2004; Watkins et al., 2002; Watkins & Canivez, 2004). A study by Macmann and Barnett (1997)

indicated that approximately 62% of their sample had at least one significant strength or weakness on the index score level. This finding led them to question how scatter could be a clinical indicator if it is so common in the general population. Likewise, Watkins's (2000) review of the literature indicated that interpreting significant scatter is not useful for predicting academic performance or for tailoring interventions; he concluded that interpretation of cognitive profiles lacks reliability, validity, and diagnostic utility.

Despite empirical findings failing to support cognitive profile analysis, the practice endured. To enumerate, in a survey of 354 practicing school psychologists, 89% of respondents revealed that they used subtest scores, index scores, or both in their clinical practice (Pfeiffer et al., 2000). In the same year that this survey was conducted, Watkins (2000) challenged these interpretive practices in a review of the literature entitled "Cognitive Profile Analysis: A Shared Professional Myth." Based on findings from investigations of the clinical utility of index- and subtest-level profile analysis, Watkins concluded, "psychologists should eschew interpretation of cognitive test profiles and must accept that they are acting in opposition to the scientific evidence if they engage in this practice" (p. 476).

Contemporary Cognitive Profile Analysis

Due to the lack of evidence supporting subtest analysis, proponents of IT have shifted their focus from subtest-level to composite-level scores. A series of new approaches have emerged that encourage psychologists to concentrate their interpretation on the index-level scores, stating that this improves the clinical utility of profile analysis because index scores are generally more reliable than subtests (McGill et al., 2018). Although proponents of these contemporary approaches were inspired and influenced by Kaufman's work, they were attracted to advances in theory, which guided hypotheses about how cognitive abilities relate and work

together within a comprehensive framework (McGill, 2013). In particular, scholars were drawn to the Cattell-Horn-Carroll theory of cognitive abilities, which has come to dominate the cognitive testing scene in school psychology since 2001 (Ortiz & Flanagan, 2009).

Cattell-Horn-Carroll Theory

The Cattell-Horn-Carroll (CHC) theory of intelligence conceptualizes cognitive abilities in a hierarchical structure with three strata (Schneider & McGrew, 2018). At the top of the hierarchy (Stratum III) is general intelligence (*g*). Below that are 18 broad abilities (Stratum II), including comprehension-knowledge (Gc), quantitative knowledge (Gq), fluid reasoning (Gf), short-term working memory (Gwm), visual-spatial processing (Gv), and processing speed (Gs). At the bottom of the hierarchy are narrow abilities (Stratum I), which are clusters of highly correlated specific abilities (Schneider & McGrew, 2018).

The CHC model provides the field with a standardized method to categorize and describe cognitive tasks and abilities, as it presently guides test development and interpretation in school psychology. The theory has also served as a foundation for numerous cognitive tests and systems to interpret cognitive profiles (e.g., Patterns of Strengths and Weaknesses [PSW; Flanagan et al., 2010], Cross-Battery Assessment [XBA; Flanagan et al., 2013]). These contemporary approaches to profile analysis differ significantly from earlier versions (e.g., IT), in that the current systems emphasize interpretation of composite scores and encourage psychologists to forego subtest-level interpretation (McGill et al., 2018). Although a discussion of these approaches is out of the scope of the present review, readers are encouraged to read Phipps and Beaujean (2016) and Flanagan et al. (2018) for a detailed review of PSW and XBA, respectively.

Ongoing Debate of Cognitive Profile Analysis

Despite the progressive shift from subtest- to index-level interpretation, the psychometric concerns with profile analysis did not disappear entirely. This time, however, the debate involves interpreting index scores versus the FSIQ; or, in terms of the CHC theory, broad abilities versus general intelligence.

On the one hand, researchers suggest focusing interpretation on index scores because they are more diagnostically relevant and better help to inform interventions than general intelligence (e.g., Decker et al., 2013; Hale & Fiorello, 2004; Schultz et al., 2012). Many of these scholars argue that individuals with learning disabilities have a particular PSW within their cognitive profiles, rendering index scores important for making diagnostic decisions (Feifer et al., 2014; Hale et al., 2008). Additionally, advocates of profile analysis state that the general intelligence factor fails to take into consideration that the goal of assessment is to determine how learning problems manifest and to develop effective interventions (Decker et al., 2013).

On the other hand, critics of cognitive profile analysis argue that there is little empirical support for the clinical utility of these approaches (e.g., Burns et al., 2016; Kavale & Forness, 1984; McGill & Busse, 2017; Miciak et al., 2018). Namely, two recent reviews of the literature (McGill & Busse, 2017; McGill et al., 2018) indicated that the diagnostic validity of profile analytic methods such as PSW has not been established. Other researchers have also found these methods to have low to moderate sensitivity and low positive predictive values for identifying learning disabilities (e.g., Miciak et al., 2014, 2018; Taylor et al., 2017).

Given the lack of evidence for the diagnostic accuracy of cognitive profile analysis, scholars also warn against using these methods to inform educational decisions and interventions. Namely, a meta-analysis by Burns et al. (2016) found that the effect sizes

associated with academic interventions directed by cognitive scores were mostly small; they concluded, “the current and previous data indicate that measures of cognitive abilities have little to no utility in screening or planning interventions for reading and mathematics” (p. 37).

Since the field began, numerous approaches for cognitive profile analysis have been established in the school psychology literature, and the use of these methods in clinical training and practice continues to be widespread. To elucidate, in a recent survey by Benson et al. (2019), over 50% of psychologists reported using some form of profile analysis on a routine basis. However, as articulated by McGill et al. (2018), “clinical tradition should not be confused with clinical validation” (p. 117), as a myriad of psychometric concerns regarding these methods have been raised in the literature for over 30 years. Moreover, McGill et al. (2018) also point out that contemporary cognitive profile analysis approaches may simply be an updated version of previous practices. That is, although there have been some incremental theoretical and conceptual developments of these methods over the last decade, the newer variations still suffer from the psychometric limitations that have long been associated with cognitive profile analysis in general (Canivez, 2013b; McGill et al., 2018).

Validity of Cognitive Tests

As a result of long-standing training gaps, recent efforts have been directed at promoting evidence-based assessment (EBA) in clinical practice (e.g., Youngstrom, 2014). EBA is an approach that uses empirical research (e.g., construct and diagnostic validity studies) to guide the methods and measures used in clinical assessment. Specifically, the EBA approach provides explicit guidance on necessary psychometric criteria for the application and interpretation of assessment measures. Consistent with this framework, following is an overview of structural and

incremental validity in the context of cognitive measures, as both are particularly prevalent to the present study.

Structural Validity

One of the most important criteria concerning cognitive measures is that of structural validity. Establishing structural validity of an instrument substantiates its theoretical structure, provides statistical justification for the scores produced, and allows test users to interpret scores confidently (McGill et al., 2018). Most often, factor analysis, specifically exploratory factor analysis (EFA) and/or confirmatory factor analysis (CFA), is used to evaluate the internal structure (i.e., factor structure) of cognitive measures. The results of these analyses inform the factor structure of an instrument, which is used to construct the index and composite scores that are interpreted by psychologists and other test users (McGill & Dombrowski, 2017). For this reason, the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association, & National Council on Measurement in Education, 2014) encourage test publishers to include this information in the technical manuals of these tests and suggest that users interpret scores cautiously, if at all, for tests that are not supported with appropriate factor analytic evidence. Unfortunately, however, this does not always seem to be the case. Index scores of cognitive tests often contain insufficient unique variance that psychologists are not able to disentangle at the individual level, causing concern for any clinical interpretations made from the scores (Watkins, 2017). When this is the case, the technical manuals of cognitive tests rarely contain information regarding interpreting index and subtest scores, leaving it up to the judgement of psychologists (McGill et al., 2018).

Fortunately, researchers have taken it upon themselves to investigate the structural validity of commonly used cognitive tests. Canivez et al. (2017) used CFA to examine the

structural integrity of the Weschler Intelligence Scales for Children, Fifth Edition (WISC-V; Weschler, 2014a). Contrary to the WISC-V structure promoted in the WISC-V *Technical and Interpretive Manual* (Weschler, 2014b), Canivez et al. (2017) found that a four-factor structure, rather than a five-factor structure, best fit the normative data. Their results also showed that general intelligence had a prevalent influence across all but one WISC-V subtest, whereas the broad abilities accounted for relatively little subtest variance. As such, Canivez et al. suggested that interpretive emphasis should be placed on the FSIQ because interpreting the index scores may lead to misinterpretation or overinterpretation of scores. In a similar study, Canivez and Watkins (2010a) used CFA to examine the structural validity of the Weschler Adult Intelligence Scales, Fourth Edition (WAIS-IV; Weschler, 2008a). Though their factor structure aligned with the structure proposed in the WAIS-IV *Technical and Interpretative Manual* (Weschler, 2008b), their results indicated that general intelligence accounted for the greatest amount of total and common variance. Thus, Canivez and Watkins (2010a) suggested focusing interpretation on the FSIQ.

The findings from Canivez and Watkins (2010a) and Canivez et al.'s (2017) studies are comparable to those found in EFA and CFA studies of the WISC-V (Canivez et al., 2016, 2017; Dombrowski et al., 2015) and the WAIS-IV (Canivez & Watkins, 2010b; Gignac & Watkins, 2013). Further, these results are not unique among Wechsler scales, as similar results were observed with the Differential Ability Scales, Second Edition (DAS-II; Canivez & McGill, 2016; Dombrowski, Golay, et al., 2018), Kaufman Assessment Battery for Children, Second Edition (KABC-II; McGill & Dombrowski, 2018), Reynolds Intellectual Assessment Scales (Dombrowski et al., 2009; Nelson & Canivez, 2012), and Woodcock-Johnson IV Tests of Cognitive Abilities (Dombrowski et al., 2017; Dombrowski, McGill, & Canivez, 2018).

Incremental Validity

Another key psychometric property of cognitive measures is incremental validity, which indicates whether a test provides meaningful information not yet accounted for by existing measures (Sechrest, 1963). As it relates to cognitive testing within the school and clinical setting, cognitive test variables are examined in terms of their contribution to the prediction of a clinically relevant criterion (e.g., diagnosis, treatment/intervention planning, progress monitoring; Hunsley & Meyer, 2003). Establishing the incremental validity of cognitive measures is often done using hierarchical multiple regression analysis (Canivez, 2013b). In this procedure, the FSIQ is entered first into a regression equation, followed by the first order factor scores to predict a criterion achievement variable (McGill & Busse, 2015).

Incremental validity studies using hierarchical multiple regression analysis have been conducted on various iterations of the Wechsler scales (Canivez, 2013a; Glutting et al., 2006), the KABC-II (McGill, 2015; McGill & Busse, 2013), and the Woodcock-Johnson III Tests of Cognitive Abilities (WJ-III-COG; McGill & Busse, 2015). For example, McGill and Busse (2015) examined the incremental validity of the WJ-III-COG composite scores in predicting achievement. Results showed that the FSIQ accounted for 46%–55% of the variance in the Woodcock-Johnson III Tests of Achievement (WJ-III-ACH) reading composites, 29%–54% in the WJ-III-ACH mathematics composites, and 41%–46% of the variance in the WJ-III-ACH writing composites. The WJ-III-COG composite scores combined contributed an additional 2%–23% (*Mdn* = 5%) of the incremental variance beyond general intelligence.

A similar study by Canivez (2013a) examined the incremental validity of the WAIS-IV in predicting achievement on the Weschler Individual Achievement Test-Second Edition (WIAT-II) and Third Edition (WIAT-III). The results indicated that the FSIQ accounted for 43%–77% of

the variance in WIAT-II composite scores, with the factor scores contributing an additional 1%–10% of variance. On the WIAT-III, the FSIQ accounted for 26%–63% of composite score variance, and the factor scores contributed an additional 5%–12% of incremental variance beyond general intelligence.

Across studies, findings demonstrated that general intelligence accounts for a large portion of the reliable achievement variance, and that little additional incremental variance is accounted for by factor scores after controlling for the predictive effects of the FSIQ. Given these findings, researchers maintain that interpretation of the FSIQ is more parsimonious than interpretation of composite scores, which is confounded by the fact that first-order scores are saturated with common variance (McGill, 2013).

Significant Factor Score Variability

Despite overwhelming evidence indicating that the FSIQ score has the strongest structural and incremental validity, psychologists continue to place greater interpretive weight on index scores in practice (Kranzler et al., 2020; Pfeiffer et al., 2000). Further, interpretation of cognitive test scores becomes even more ambiguous in situations such as when an examinee presents with significant factor score variability, or scatter. Technical manuals and interpretive guidebooks (e.g., *Essentials of Psychological Assessment* series) often encourage foregoing interpretation of the FSIQ in such cases, stating that the FSIQ score is invalid when factor scores vary significantly (Farmer et al., 2021). Unbeknownst to many scholars and psychologists, however, scatter is common in normative populations, occurring in over 50% of the normative sample for most cognitive measures (Binder et al., 2009; Grégoire et al., 2011; Kavale & Forness, 1984; Macmann & Barnett, 1997). In this case, then, why does the FSIQ even exist if it is supposedly invalid more than half of the time?

Before discounting the score entirely, we need to investigate the validity and interpretability of the FSIQ when there is significant factor score variability. The remainder of the present literature review does just that; first we present an overview of the theoretical underpinnings of significant scatter, then we explore studies that have examined the impact of scatter on structural and incremental validity, followed by a brief discussion of scatter research in clinical samples.

Variability Hypothesis

Within the intelligence testing literature, it is frequently suggested that significant variability among factor and index scores renders the general intelligence score (e.g., FSIQ) an invalid measure of overall ability (e.g., Courville et al., 2016; Lezak et al., 2012). In such cases, psychologists are encouraged to disregard the FSIQ for the purposes of clinical interpretation. McGill (2016) referred to this popular interpretive heuristic—so popular that he deemed it a "proverbial *lingua franca* for clinical IQ test interpretation across applied psychological disciplines (p. 51)—as the "variability hypothesis."

For many of the commonly used intelligence tests (e.g., KABC-II, WISC-V, WAIS-IV), significant scatter is statistically defined as a difference of 23 points or greater (at least 1.5 standard deviation) between the highest and lowest index scores. For instance, an examinee with a Verbal Comprehension Index score of 105 and a Perceptual Reasoning Index score of 82 would be considered to have significant variability in their score profile. Although the procedures for accounting for this scatter when making clinical interpretations vary across cognitive tests, they all emphasize that significant factor score variability renders the FSIQ invalid for interpretation (Reschly et al., 2002).

These procedures for handling cases with significant scatter are outlined in various assessment manuals, guidebooks, and handbooks (Farmer et al., 2021). To wit, in their discussion of an approach to interpreting Weschler scores in the *Handbook of Psychological Assessment*, Groth-Marnat and Wright (2016) stated:

> Examiners can interpret the more global measures (Full Scale IQ, Global Ability Index) with greater meaning, usefulness, and certainty if there is not a high degree of difference among the index scores…With increasing differences, the purity of the global measures becomes contaminated so that interpretations of them become less meaningful. (p. 163)

Likewise, the interpretive manuals for many cognitive tests also provide users with interpretative procedures to follow when there is significant scatter. For instance, the *Technical and Interpretive Manuals* for the latest versions of the Wechsler scales (Wechsler, 2008b, 2014b) encourage users to interpret scores in a stepwise fashion beginning with the FSIQ, at which point users are urged to examine the index scores for significant differences among scores. If meaningful variability is observed, users are encouraged to relinquish clinical interpretation of the FSIQ and instead focus their interpretive weight on the profile of index scores.

As McGill (2016) points out, despite the prevalence of interpretive recommendations such as these to appear in technical and interpretive manuals for cognitive tests, no validity evidence is provided in these manuals. This, then, is in direct conflict with the validity standards delineated in the *Standards for Educational and Psychological Testing* (AERA et al., 2014).

Impact of Scatter on Structural Validity

In one of the earliest investigations of the variability hypothesis, Fiorello et al. (2002) used regression communality analysis to examine the structure of the WISC-III FSIQ in a sample of typical children with flat (n = 707) and scattered (n = 166) cognitive profiles. Results revealed

that the FSIQ communality for the flat group was largely comprised of shared variance (i.e., *g*; 89%), whereas the FSIQ communality for the scatter group was primarily comprised of unique variance (61%). These findings led Fiorello et al. to conclude that the FSIQ is not an accurate representation of global intellectual ability for those with significant scatter, in which case psychologists should place greater emphasis on index scores.

In a later study replicating these methods, Fiorello et al. (2007) examined the constitution of the WISC-IV FSIQ for children with a diagnosis of a learning disability (n = 128), attention-deficit/hyperactivity disorder ([ADHD]; n = 71), and a traumatic brain injury (n = 29). A commonality analysis revealed little shared variance among factor predictors of FSIQ for the learning disability group (1.7%), the ADHD group (2.4%), and the traumatic brain injury group (2.7%). Given these findings, Fiorello et al. (2007) concluded that diagnostic interpretation for children with disabilities should be at the level of index scores rather than the FSIQ.

Several scholars acted in response to the study by Fiorello et al. (2007), criticizing their methodology and conclusions (Dana & Dawes, 2007; Daniel, 2007; Schneider, 2008; Watkins et al., 2007). Among these critiques was the choice not to use a more standardized technique (e.g., factor analysis) to investigate the structure of the WISC-IV, as well as the choice to exclude participants with FSIQs below 80 or above 120 (Dana & Dawes, 2007). In another response, Daniel (2007) conducted a similar study examining the effect of index score scatter on the structural validity of the FSIQ. EFA results indicated that the FSIQ remained a valid estimation of global cognitive ability for groups with both flat and scattered cognitive profiles.

In a more recent study, McGill (2016) examined the structural validity of the KABC-II for participants in the normative sample who demonstrated significant variability in their factor scores (i.e., difference of 23 points or greater between the highest and lowest index scores; N =

1,209). An EFA with multiple factor extraction criteria supported the five-factor structure proposed by the test publisher. Using the Schmid-Leiman transformation to partition the variance between higher- and first-order dimensions (Schmid & Leiman, 1957), results demonstrated that the hierarchical *g* factor accounted for greater portions of total and common variance than any single domain specific factor. Given these findings, McGill concluded that, even in the presence of significant scatter, the hierarchical structuring of cognitive dimensions was not altered. Findings also revealed that discrediting the global factor in the presence of significant scatter does not prevent it from influencing first-order factor scores and thus the FSIQ should not automatically be disregarded as to comply with the variability hypothesis.

In a similar study aiming to extend McGill's (2016) work, Ayoub (2020) examined the effects of scatter on the structural validity of the global intelligence score. Invariance testing with a scatter group drawn from the KABC-II normative sample (as in McGill's study) and a mixed comparison group drawn from the KABC-II Normative Update (Kaufman et al., 2018) standardization sample indicated that the five-factor structure upheld for both the scattered and mixed comparison groups. Given findings revealing that the same constructs were measured across the two groups, Ayoub concluded that the global intelligence score maintains its validity despite profile scatter. This finding is consistent with other studies investigating the impact of scatter on the structural validity of the FSIQ (Daniel, 2007; McGill, 2016).

Impact of Scatter on Incremental Validity

Scholars have also investigated the impact of significant factor score variability on the incremental validity of cognitive tests. In a pioneering study, Watkins et al. (2007) examined the predictive validity of the WISC-III FSIQ in clinical and mixed/normative populations with and without index score variability. The study included three samples—412 children from the WISC-

III/WIAT linking sample, 112 children from the WISC-IV/WIAT-II linking sample, and 460 children in special education programs—and each sample was divided into two groups based on their cognitive profile (flat or scatter). Results indicated that the FSIQ was a significant predictor of academic achievement, but neither index score variability nor the interaction between the FSIQ and index score variability made a significant incremental contribution to the prediction of academic achievement scores. As such, Watkins et al. concluded that, even in the presence of significant scatter, the FSIQ is a robust predictor of academic achievement for children with and without disabilities. Further, these results were subsequently replicated on the DAS (Kotz et al., 2008) and in a study predicting long-term achievement outcomes with the WISC-III (Freberg et al., 2008).

As part of a comprehensive investigation of the incremental validity of the WJ-III-COG, McGill (2013) examined the effect of significant factor score variability. Hierarchical multiple regression analyses revealed that the FSIQ accounted for 39% to 60% of achievement variance on the WJ-III-ACH in the presence of factor score variability, with large effect size estimates. Conversely, factor scores accounted for 3% to 33% of the achievement variance when controlling for the effects of the FSIQ score. These results suggest that the FSIQ remained a stable predictor of academic achievement in the presence of significant factor score variability. In contrast to prior studies (e.g., Freberg et al., 2008; Kotz et al., 2008; Watkins et al., 2007), McGill also found that the incremental validity of the factor scores increased as scatter increased, particularly in predicting reading. Nonetheless, he advised against the practice of disregarding the FSIQ for predicting achievement outcomes when there is significant scatter.

In a later study, McGill (2016) examined the incremental validity of the KABC-II global score using participants in the KABC-II/KTEA-II normative sample that presented with

significant factor score variability (N = 1,209). Hierarchical multiple regression showed that the FSIQ accounted for large predictive effects of academic achievement, whereas the index scores accounted for trivial proportions of incremental predictive variance beyond the FSIQ. As such, McGill agreed with other scholars (e.g., Freberg et al., 2008; Kotz et al., 2008; Watkins et al., 2007) in concluding that scatter does not render the global cognitive score an invalid predictor of achievement.

Summary of The Literature

Overall, previous studies have consistently found that the FSIQ maintains its structural and incremental validity in the presence of significant factor score variability (e.g., Ayoub, 2020; Daniel, 2007; Kotz et al., 2008; McGill, 2016; Watkins et al., 2007). This has shown to be true for mixed/normative samples (e.g., Ayoub, 2020; Kotz et al., 2008; McGill, 2016) as well as clinical samples (Freberg et al., 2008; Watkins et al., 2007). Despite these findings, many psychologists continue to abide by the variability hypothesis, foregoing interpretation of the FSIQ when an examinee is observed to have significant scatter. Namely, in a recent survey of practicing school psychologists (N = 1, 317), 62% of respondents indicated that they do not interpret the FSIQ in the presence of significant differences between index scores (Kranzler et al., 2020). This practice also continues to be encouraged in many technical manuals and guidebooks to cognitive assessment (Farmer et al., 2021), despite the absence of supporting empirical or theoretical evidence.

To reduce the gap between empirical findings and clinical practice, it is necessary to address the limitations in this area of research. First and foremost is the dearth of investigations designed to examine the potential impact of scatter on the structural and incremental validity of cognitive measures. To date, only one study (McGill, 2016) was designed specifically for this

purpose. McGill's study, like many of the studies that investigated the impact of scatter on either structural or incremental validity, used an archived standardization sample. Additional research is needed to determine if previous findings generalize to clinical populations. Whereas two studies used a clinical sample (Freberg et al., 2008; Watkins et al., 2007), they only examined incremental validity. Moreover, the majority of studies used a child- or adolescent-aged population. To our knowledge, no previously published research in this area had been conducted with a clinical college-aged or adult sample.

Current Study Purpose and Research Questions

To address the aforementioned gaps in the literature, the present study aimed to investigate the effect of significant factor score variability on the structural and incremental validity of the WAIS-IV in a clinically referred sample of college students. Modeled after McGill's (2016) study, the current study served to determine whether significant variability in the index scores on the WAIS-IV renders the FSIQ an invalid measure of overall ability and thus erroneous for clinical interpretation. Accordingly, the following research questions were addressed:

1. Does significant factor score variability impact the structural validity of the WAIS-IV FSIQ in a clinically referred sample?
2. Is the global ability score (FSIQ) a more valid predictor of achievement in those without significant factor score variability?

Based on previous research, the following hypotheses were developed:

1. It was hypothesized that the WAIS-IV FSIQ would maintain its structural validity in the presence of significant index score variability. That is, it was predicted that the FSIQ would remain a valid estimation of global cognitive ability despite significant scatter.

2. In line with previous findings, it was hypothesized that significant index score variability would not undermine the incremental validity of the FSIQ. That is, it was predicted that scatter would not render the FSIQ an invalid predictor of achievement.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement on Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Ayoub, J. L. (2020). Does cognitive scatter attenuate the validity of the global intelligence score? (Publication No. 28091905) [Doctoral dissertation, Hofstra University]. ProQuest Dissertations & Theses Global.

Benson, N. F., Floyd, R. G., Kranzler, J. H., Eckert, T. L., Fefer, S. A., & Morgan, G. B. (2019). Test use and assessment practices of school psychologists in the United States: Findings from the 2017 National Survey. *Journal of School Psychology*, *72*, 29–48. https://doi.org/10.1016/j.jsp.2018.12.004

Benson, N. F., Maki, K. E., Floyd, R. G., Eckert, T. L., Kranzler, J. H., & Fefer, S. A. (2020). A national survey of school psychologists' practices in identifying specific learning disabilities. *School Psychology*, *35*(2), 146–157. https://doi.org/10.1037/spq0000344

Binder, L. M., Iverson, G. L., & Brooks, B. L. (2009). To err is human: "Abnormal" neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology*, *24*(1), 31–46. https://doi.org/10.1093/arclin/acn001

Bray, M. A., Kehle, T. J., & Hintze, J. M. (1998). Profile analysis with the Wechsler Scales: Why does it persist? *School Psychology International*, *19*, 209–220. https://doi.org/10.1177/0143034398193002

Burns, M. K., Petersen-Brown, S., Haegele, K., Rodriguez, M., Schmitt, B., Cooper, M., Clayton, K., Hutcheson, S., Conner, C., Hosp, J., & VanDerHeyden, A. M. (2016). Meta-analysis of academic interventions derived from neuropsychological data. *School Psychology Quarterly*, *31*(1), 28–42. https://doi.org/10.1037/spq0000117

Canivez, G. L. (2013a). Incremental criterion validity of WAIS-IV factor index scores: Relationships with WIAT-II and WIAT-III subtest and composite scores. *Psychological Assessment*, *25*, 484–495. https://doi.org/10.1037/a0032092

Canivez, G. L. (2013b). Psychometric versus actuarial interpretation of intelligence and related aptitude batteries. In D. H. Saklofske, C. R. Reynolds, & V. L. Schwean (Eds.). *The Oxford handbook of child psychological assessment* (pp. 84–112). Oxford University Press.

Canivez, G. L., & Kush, J. C. (2013). WAIS-IV and WISC-IV structural validity. *Journal of Psychoeducational Assessment*, *31*(2), 157–169. https://doi.org/10.1177/0734282913478036

Canivez, G. L., & McGill, R. J. (2016). Factor structure of the Differential Ability Scales-Second Edition: Exploratory and hierarchical factor analyses with the core subtests. *Psychological Assessment*, *28*(11), 1475–1488. https://doi.org/10.1037/pas0000279

Canivez, G. L., & Watkins, M. W. (2010a). Exploratory and higher-order factor analyses of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) adolescent subsample. *School Psychology Quarterly*, *25*, 223–235. https://doi.org/10.1037/a0022046

Canivez, G. L., & Watkins, M. W. (2010b). Investigation of the factor structure of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV): Exploratory and higher order factor analyses. *Psychological Assessment*, *22*, 827–836. https://doi.org/10.1037/a0020429

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2016). Factor structure of the Wechsler Intelligence Scale for Children-Fifth Edition: Exploratory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment*, *28*(8), 975–986. https://doi.org/10.1037/pas0000238

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2017). Structural validity of the Wechsler Intelligence Scale for Children-Fifth Edition: Confirmatory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment*, *29*(4), 458–472. https://doi.org/10.1037/pas0000358

Cohen, J. (1952). A factor-analytically based rationale for the Wechsler-Bellevue. *Journal of Consulting Psychology*, *16*, 272–277.

Cormier, D. C., McGrew, K. S., Bulut, O., & Funamoto, A. (2016). Revisiting the relations between the WJ-IV measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during the school-age years. *Journal of Psychoeducational Assessment*, *35*(8), 731–754. https://doi.org/10.1177/0734282916659208

Courville, T., Coalson, D. L., Kaufman, A. S., & Raiford, S. E. (2016). Does WISC-V scatter matter? In A. S. Kaufman, S. E. Raiford, & D. L. Coalson (Eds.). *Intelligent testing with the WISC-V* (pp. 209–226). Wiley.

Dana, J., & Dawes, R. (2007). Comment on Fiorello et al. "Interpreting intelligence test results for children with disabilities: Is global intelligence relevant?" *Applied Neuropsychology*, *14*(1), 21–25. https://doi.org/10.1080/09084280701280379

Daniel M. H. (2007). "Scatter" and the construct validity of FSIQ: Comment on Fiorello et al. (2007). *Applied Neuropsychology*, *14*(4), 291–295. https://doi.org/10.1080/09084280701719401

Decker, S. L., Hale, J. B., & Flanagan, D. P. (2013). Professional practice issues in the assessment of cognitive functioning for educational applications. *Psychology in the Schools*, *50*, 300–313. https://doi.org/10.1002/pits.21675

Dombrowksi, S. C., Canivez, G. L., Watkins, M. W., & Beaujean, A. A. (2015). Exploratory bifactor analysis of the Wechsler Intelligence Scale for Children-Fifth Edition with the 16 primary and secondary subtests. *Intelligence*, *53*, 194–201. https://doi.org/10.1016/j.intell.2015.10.009

Dombrowski, S. C., Golay, P., McGill, R. J., & Canivez, G. L. (2018). Investigating the theoretical structure of the DAS-II core battery at school age using Bayesian structural equation modeling. *Psychology in the Schools*, *55*, 190–207. https://doi.org/10.1002/pits.22096

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2017). Exploratory and hierarchical factor analysis of the WJ-IV cognitive at school age. *Psychological Assessment*, *29*(4), 394–407. https://doi.org/10.1037/pas0000350

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2018). An alternative conceptualization of the theoretical structure of the WJ IV cognitive at school age: A confirmatory factor analytic investigation. *Archives of Scientific Psychology*, *6*(1), 1–13. https://doi.org/10.1037/arc0000039

Dombrowski, S. C., Watkins, M. W., & Brogan, M. J. (2009). An exploratory investigation of the factor structure of the Reynolds Intellectual Assessment Scales (RIAS). *Journal of Psychoeducational Assessment*, *27*(6), 494–507. https://doi.org/10.1177/0734282909333179

Dombrowski, S. C., Watkins, M. W., McGill, R. J., Canivez, G. L., Holingue, C., Pritchard, A. E., & Jacobson, L. A. (2021). Measurement invariance of the Wechsler Intelligence Scale for Children, Fifth Edition 10-subtest primary battery: Can index scores be compared

across age, sex, and diagnostic groups? *Journal of Psychoeducational Assessment*, *39*(1), 89–99. https://doi.org/10.1177/0734282920954583

Farmer, R. L., McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2021). Why questionable assessment practices remain popular in school psychology: Instructional materials as pedagogic vehicles. *Canadian Journal of School Psychology*, *36*(2), 98–114. https://doi.org/10.1177/0829573520978111

Feifer, S. G., Nader, R. G., Flanagan, D. P., Fitzer, K. R., & Hicks, K. (2014). Identifying specific reading subtypes for effective educational remediation. *Learning Disabilities: A Multidisciplinary Journal*, *20*(1), 18–30.

Fiorello, C. A., Hale, J. B., Holdnack, J. A., Kavanagh, J. A., Terrell, J., & Long, L. (2007). Interpreting intelligence test results for children with disabilities: Is global intelligence relevant? *Applied Neuropsychology*, *21*, 2–12. https://doi.org/10.1080/09084280701280338

Fiorello, C. A., Hale, J. B., McGrath, M., Ryan, K., & Quinn, S. (2002). IQ interpretation for children with flat and variable test profiles. *Learning and Individual Differences*, *13*, 115–125. https://doi.org/10.1016/S1041-6080(02)00075-4

Flanagan, D. P., Costa, M., Palma, K., Leahy, M. A., Alfonso, V. C., & Ortiz, S. O. (2018). Cross-battery assessment, the cross-battery assessment software system, and the assessment–intervention connection. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 731–776). The Guilford Press.

Flanagan, D. P., Fiorello, C. A., & Ortiz, S. O. (2010). Enhancing practice through application of Cattell-Horn-Carroll theory and research: A "third-method" approach to specific learning disability identification. *Psychology in the Schools*, *47*, 739–760. https://doi.org/10.1002/pits.20501

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Wiley.

Freberg, M. E., Vandiver, B. J., Watkins, M. W., & Canivez, G. L. (2008). Significant factor score variability and the validity of the WISC-III Full Scale IQ in predicting later academic achievement. *Applied neuropsychology*, *15*(2), 131–139. https://doi.org/10.1080/09084280802084010

Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research*, *48*, 639–662. https://doi.org/10.1080/00273171.2013.804398

Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. *The Journal of Special Education*, *40*, 103–114. https://doi.org/10.1177/00224669060400020101

Grégoire, J., Coalson, D. L., & Jianjun, Z. (2011). Analysis of WAIS-IV index score scatter using significant deviation from the mean index score. *Assessment*, *18*(2), 168–177. https://doi.org/10.1177/1073191110386343

Groth-Marnat, G., & Wright, A. J. (2016). *Handbook of psychological assessment* (6th ed.). Wiley.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook*. Guilford.

Hale, J. B., Fiorello, C. A., Dumont, R., Willis, J. O., Rackley, C., & Elliott, C. (2008). Differential Ability Scales–Second Edition (Neuro)psychological predictors of math performance for typical children and children with math disabilities. *Psychology in the Schools*, *45*(9), 838–858. https://doi.org/10.1002/pits.20330

Harris, A. J., & Shakow, D. (1937). The clinical significance of numerical measures of scatter on the Stanford-Binet. *Psychological Bulletin*, *34*, 134–150. https://doi.org/10.1037/h0058420.

Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: conceptual, methodological, and statistical issues. *Psychological Assessment*, *15*(4), 446–455. https://doi.org/10.1037/1040-3590.15.4.446

Kaufman, A. S. (1979). *Intelligent testing with the WISC-R*. Wiley.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. Wiley.

Kaufman, A. S., & Kaufman, N. L. (2004a). *Kaufman Assessment Battery for Children* (2nd ed.). American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004b). *Kaufman Test of Educational Achievement* (2nd ed.). American Guidance Service.

Kaufman, A. S., Kaufman, N. L., Drozdik, L. W., & Morrison, J. (2018). *Kaufman Assessment*

Battery for Children-Second Edition Normative Update manual supplement. Pearson.

Kaufman, A. S., Lichtenberger, E. O., Fletcher-Janzen, E., & Kaufman, N. L. (2005). *Essentials of KABC-II assessment*. Wiley.

Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (Eds.). (2016). *Intelligent testing with the WISC-V*. Wiley.

Kavale, K. A., & Forness, S. R. (1984). A meta-analysis of the validity of Wechsler scale profiles and recategorizations: Patterns or parodies? *Learning Disability Quarterly*, *7*, 136–156. https://doi.org/10.2307/ 1510314

Kotz, K. M., Watkins, M. W., & McDermott, P. A. (2008). Validity of the General Conceptual Ability score from the Differential Ability Scales as a function of significant and rare interfactor variability. *School Psychology Review*, *37*(2), 261–278.

Kranzler, J. H., & Floyd, R. G. (2013). *Assessing intelligence in children and adolescents: A practical guide*. Guilford Press.

Kranzler, J. H., Maki, K. E., Benson, N. F., Eckert, T. L., Floyd, R. G., & Fefer, S. A. (2020). How do school psychologists interpret intelligence tests for the identification of specific learning disabilities? *Contemporary School Psychology*. https://doi.org/10.1007/s40688-020-00274-0

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). Oxford University Press.

Lichtenberger, E. O., & Kaufman, A. S. (2009). *Essentials of WAIS-IV assessment*. Wiley.

Lichtenberger, E. O., Sotelo-Dynega, M., & Kaufman, A. S. (2009). The Kaufman Assessment Battery for Children-Second Edition. In J. A. Naglieri & S. Goldstein (Eds.), *Practitioner's guide to assessing intelligence and achievement* (pp. 61–94). Wiley.

Lilienfeld, S. O., Ammirati, R., & David, M. (2012). Distinguishing between science from pseudoscience in school psychology: Science and scientific thinking as safeguards against human error. *Journal of School Psychology*, *50*, 7–36. https://doi.org/10.1016/j.jsp.2011.09.006

Lockwood, A. B., & Farmer, R. L. (2020). The cognitive assessment course: Two decades later. *Psychology in the Schools*, *57*(2), 265–283. https://doi.org/10.1002/pits.22298

Macmann, G. M., & Barnett, D. W. (1997). Myth of the master detective: Reliability of interpretations for Kaufman's "intelligent testing" approach to the WISC-III. *School Psychology Quarterly*, *12*, 197–234. https://doi.org/10.1037/h0088959

McDermott, P. A., Fantuzzo, J. W., & Glutting, J. J. (1990). Just say no to subtest analysis: A critique on Wechsler theory and practice. *Journal of Psychoeducational Assessment*, *8*, 290–302. https://doi.org/10. 1177/073428299000800307

McGill, R. J. (2013). Beyond g: Assessing the incremental validity of the Cattell-Horn-Carroll (CHC) broad ability factors on the Woodcock-Johnson III Tests of Cognitive Abilities (Publication No. 3621595) [Doctoral dissertation, Chapman University]. ProQuest Dissertations & Theses Global.

McGill, R. J. (2015). Interpretation of KABC-II scores: An evaluation of the incremental validity of Cattell-Horn-Carroll (CHC) factor scores in predicting achievement. *Psychological Assessment*, *27*(4), 1417–1426. https://doi.org/10.1037/pas0000127

McGill, R. J. (2016). Invalidating the full scale IQ score in the presence of significant factor score variability: Clinical acumen or clinical illusion? *Archives of Assessment Psychology*, *6*(1), 49–79.

McGill, R. J. (2018). Confronting the base rate problem: More ups and downs for cognitive scatter analysis. *Contemporary School Psychology*, *22*, 384–393. https://doi.org/10.1007/s40688-017-0168-4

McGill, R. J., & Busse, R. T. (2015). Incremental validity of the WJ III COG: Limited predictive effects beyond the GIA-E. *School Psychology Quarterly*, *30*(3), 353–365. https://doi.org/10.1037/spq0000094

McGill, R. J., & Busse, R. T. (2017). When theory trumps science: A critique of the PSW model for SLD identification. *Contemporary School Psychology*, *21*(1), 10–18. https://doi.org/10.1007/s40688-016-0094-x

McGill, R. J., & Dombrowski, S. C. (2018). Factor structure of the CHC model for the KABC-II: Exploratory factor analyses with the 16 core and supplementary subtests. *Contemporary School Psychology*, *22*, 279–293. https://doi.org/10.1007/s40688-017-0152-z

McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2018). Cognitive profile analysis in school psychology: History, issues, and continued concerns. *Journal of School Psychology*, *71*, 108–121. https://doi.org/10.1016/j.jsp.2018.10.007

McGill, R. J., Styck, K. M., Palomares, R. S., & Hass, M. R. (2016). Critical issues in specific learning disability identification: What we need to know about the PSW model. *Learning Disability Quarterly*, *39*(3), 159–170. https://doi.org/10.1177/0731948715618504

Miciak, J., Fletcher, J. M., Stuebing, K. K., Vaughn, S., & Tolar, T. D. (2014). Patterns of cognitive strengths and weaknesses: Identification rates, agreement, and validity for learning disabilities identification. *School Psychology Quarterly*, *29*, 21–37. https://doi.org/10.1037/spq0000037

Miciak, J., Taylor, W. P., Stuebing, K. K., & Fletcher, J. M. (2018). Simulation of LD identification accuracy using a pattern of processing strengths and weaknesses method with multiple measures. *Journal of Psychoeducational Assessment*, *36*, 21–33. https://doi.org/10.1177/0734282916683287

Miller, J. L., Saklofske, D. H., Weiss, L. G., Drozdick, L., Llorente, A. M., Holdnack, J. A., & Prifitera, A. (2016). Issues related to the WISC-V assessment of cognitive functioning in clinical and special groups. In L. G. Weiss, D. H. Saklofske, J. A. Holdnack, & A. Prifitera (Eds.), *WISC-V assessment and interpretation: Scientist-practitioner perspectives* (1st ed., pp. 287–343). Academic Press.

Nelson, J. M., & Canivez, G. L. (2012). Examination of the structural, convergent, and incremental validity of the Reynolds Intellectual Assessment Scales (RIAS) with a clinical sample. *Psychological Assessment*, *24*(1), 129–140. https://doi.org/10.1037/a0024878

Ortiz, S. O., & Flanagan, D. P. (2009). Kaufman on theory, measurement, interpretation, and fairness: A legacy in training, practice, and research. In J. C. Kaufman (Ed.). *Intelligent testing: Integrating psychological theory and clinical practice* (pp. 99–112). Cambridge University Press.

Pfeiffer, S. I., Reddy, L. A., Kletzel, J. E., Schmelzer, E. R., & Boyer, L. M. (2000). The practitioner's view of IQ testing and profile analysis. *School Psychology Quarterly*, *15*, 376–385. https://doi.org/10.1037/h0088795

Phipps, L., & Beaujean, A. A. (2016). Review of the pattern of strengths and weaknesses approach in specific learning disability identification. *Research and Practice in the Schools*, *4*(1), 18–28.

Rabin, A. I, & Guertin, W. H. (1951). Research with the Wechsler-Bellevue test: 1945-1950. *Psychological Bulletin*, *48*, 211–248.

Rapaport, D., Gil, M., & Schafer, R. (1945). *Diagnostic psychological testing: The theory, statistical evaluation, and diagnostic application of a battery of tests* (Vol. 1). Yearbook Medical.

Reschly, D. J., Myers, T. J., & Hartel, C. (Eds.). (2002). *Mental retardation: Determining eligibility for social security benefits*. National Academy Press.

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations* (5th ed.). Sattler Publisher.

Sattler, J. M. (2018). *Assessment of children: Cognitive foundations and applications* (6th ed.). Sattler Publisher.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*, 53–61. https://doi.org/10.1007/BF02289209

Schneider W. J. (2008). Playing statistical Ouija board with commonality analysis: Good questions, wrong assumptions. *Applied Neuropsychology*, *15*(1), 44–53. https://doi.org/10.1080/09084280801917566

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). Guilford Press.

Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 73–163). The Guilford Press.

Schultz, E. K., Simpson, C. G., Lynch, S. (2012). Specific learning disability identification: What constitutes a pattern of strengths and weaknesses? *Learning Disabilities: A Multidisciplinary Journal*, *18*(2), 87–97.

Sechrest, L. (1963). Incremental validity: A recommendation. *Educational and Psychological Measurement*, *23*(1), 153–158. https://doi.org/10.1177/001316446302300113

Smith, C. B., & Watkins, M. W. (2004). Diagnostic utility of the Bannatyne WISC-III pattern. *Learning Disabilities Research and Practice*, *19*, 49–56. https://doi.org/10.1111/j.1540-5826.2004.00089.x

Taylor, W. P., Miciak, J., Fletcher, J. M., & Francis, D. J. (2017). Cognitive discrepancy models for specific learning disabilities identification: Simulations of psychometric limitations. *Psychological Assessment*, *29*, 446–457. https://doi.org/10.1037/pas0000356

Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. *School Psychology Quarterly*, *15*(4), 465–479. https://doi.org/10.1037/h0088802

Watkins, M. W. (2003). IQ subtest analysis: Clinical acumen or clinical illusion? *The Scientific Review of Mental Health Practice*, *2*, 118–141.

Watkins, M. W. (2005). Diagnostic validity of Wechsler subtest scatter. *Learning Disabilities: A Contemporary Journal*, *3*, 20–29.

Watkins, M. W. (2009). Errors in diagnostic decision making and clinical judgment. In T. B. Gutkin & C. R. Reynolds (Eds.), *Handbook of school psychology* (4th ed., pp. 210–229). Wiley.

Watkins, M. W. (2017). The reliability of multidimensional neuropsychological measures: From alpha to omega. *The Clinical Neuropsychologist*, *31*, 1113–1126. https://doi.org/10.1080/13854046.2017.1317364.

Watkins, M. W., & Canivez, G. L. (2004). Temporal stability of WISC-III subtest composite: Strengths and weaknesses. Psychological Assessment, 16, 133–138. https://doi.org/10.1037/1040-3590.16.2.133

Watkins, M. W., & Glutting, J. J. (2000). Incremental validity of WISC-III profile elevation, scatter, and shape information for predicting reading and math achievement. *Psychological Assessment*, *12*, 402–408. https://doi.org/10.1037//1040-3590.12.4.402

Watkins, M. W., Glutting, J. J., & Lei, P.-W. (2007). Validity of the Full-Scale IQ when there is significant variability among WISC-III and WISC-IV factor scores. *Applied Neuropsychology*, *14*(1), 13–20. https://doi.org/10.1080/09084280701280353

Watkins, M. W., Glutting, J. J., & Youngstrom, E. A. (2005). Issues in subtest profile analysis. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 251–268). Guilford Press.

Watkins, M. W., Kush, J. C., & Schaefer, B. A. (2002). Diagnostic utility of the WISC-III learning disability index. Journal of Learning Disabilities, 35, 98–103. https://doi.org/10.1177/002221940203500201

Wechsler, D. (1939). *Wechsler-Bellevue Intelligence Scale*. Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children* (3rd ed.). Psychological Corporation.

Wechsler, D. (1992). *Wechsler Individual Achievement Test*. Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children* (4th ed.). Psychological Corporation.

Wechsler, D. (2008a). *Wechsler Adult Intelligence Scale* (4th ed.). Pearson.

Wechsler, D. (2008b). *Wechsler Adult Intelligence Scale–Fifth Edition technical and interpretive manual*. Pearson.

Wechsler, D. (2014a). *Wechsler Intelligence Scale for Children* (5th ed.). Pearson.

Wechsler, D. (2014b). *Wechsler Intelligence Scale for Children–Fifth Edition technical and interpretive manual*. Pearson.

Youngstrom, E. A. (2014). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. *Journal of Pediatric Psychology*, *39*, 204–221.

Youngstrom, E. A., & Van Meter, A. (2016). Empirically supported assessment of children and adolescents. *Clinical Psychology: Science and Practice*, *23*, 327–347. https://doi.org/10.1111/cpsp.12172

Youngstrom, E. A., Choukas-Bradley, S., Calhoun, C. D., & Jensen-Doss, A. (2015). Clinical guide to the evidence-based assessment approach to diagnosis and treatment. *Cognitive and Behavioral Practice*, *22*, 20–35. https://doi.org/10.1016/j.cbpra.2013.12.005

CHAPTER 3

METHOD

Sample and Participant Selection

The sample was comprised of 1,211 individuals who were evaluated at a university-based clinic specializing in the assessment of learning and attention disorders. The participants underwent a comprehensive psychoeducational evaluation between 2009 and 2020 and attended a public college or university (2 or 4 year) in one state in the Southeastern region of the United States. Participants were selected from this date range (2009–2020) based on the availability of scores, in that the testing clinic began administering the Fourth Edition of the Weschler Adult Intelligence Scale (WAIS-IV) in 2009 and standardized test administration procedures ceased in March 2020 due to the coronavirus disease pandemic (and remained that way when data were extracted in February 2021).

The original dataset contained 2,587 total individuals who were evaluated at the clinic during this 12-year period. Of these, 1,211 individuals were administered all of the 10 core subtests on the WAIS-IV. This was an inclusionary criterion because the 10 core subtests are necessary for deriving the four index scores and the global composite score (i.e., Full-Scale Intelligence Quotient [FSIQ]). The remaining 1,211 participants constituted the sample for this study.

Participants ranged in age from 16 to 64 years (M = 21.76; SD = 6.01), with a gender distribution that was nearly equal (male = 51.6%; female = 48.4%). Most participants were in college (75.4%), with some who were in high school or about to start college (14.2%) and others

who were transitioning to graduate school or in graduate school (7.8%). Moreover, the race/ethnicity composition was as follows: white = 82.8%, black = 8.3%, Hispanic/Latino = 3.6%, Asian/Pacific Islander = 3.1%, and other = 2.2%. The majority of participants had one (53.7%) or two (30.2%) current (i.e., diagnosed as a result of the evaluation) learning or psychological diagnosis(es), with only 108 (8.9%) having no current diagnosis. One participant was missing diagnostic information. The breakdown by diagnostic category is presented in Table 3.1.

Participants sought the evaluations to determine if they qualified for disability services through their postsecondary institution. Evaluation time was approximately 8–10 hours over the course of two days. Evaluations included qualitative (e.g., background information, clinical interview, record review, behavioral observation) and quantitative (standardized direct and indirect measures) information. Assessment instruments were chosen based on their psychometric properties and clinical usefulness with young adult, college, and adult populations. Participants were typically administered a standard testing battery developed by the testing clinic, along with additional assessments selected to meet the presenting concerns of each participant. Additionally, diagnostic decisions were made by an evaluation team that included at least one licensed psychologist. A diagnosis of a learning disability required an educational history consistent with learning disabilities, substantially limited (i.e., standard score < 90) academic achievement skills (e.g., reading decoding, fluency, and/or comprehension), and a cognitive/linguistic processing deficit meaningfully associated with the identified academic limitation. All other diagnoses were made based on criteria from the *Diagnostic and Statistical Manual of Mental Disorders - Fourth Edition, Text Revision* (DSM-IV-TR; American

Psychiatric Association, 2000) or - *Fifth Edition* (DSM-5; American Psychiatric Association, 2013).

Measures

Weschler Adult Intelligence Scale, Fourth Edition

The WAIS-IV (Weschler, 2008a) is an individually administered assessment of cognitive ability for adults between the ages of 16 and 90. The WAIS-IV is composed of 10 core subtests which produce four composite scores—the Verbal Comprehension Index (VCI), the Perceptual Reasoning Index (PRI), the Working Memory Index (WMI), and the Processing Speed Index (PSI). The index scores combine to yield the FSIQ, which is considered the "most valid measure of overall cognitive ability" (Weschler, 2008a, p. 5). The index scores and FSIQ are reported as standard scores with a mean of 100 and a standard deviation of 15.

Normative and psychometric data establishing the adequacy of the WAIS-IV as a reliable and valid measure of cognitive abilities can be found in the *WAIS-IV Technical and Interpretive Manual* (Weschler, 2008b). The WAIS-IV was standardized on a nationally representative sample (N = 2,200) in accordance with the 2005 United States Census. The manual reports excellent average internal consistency reliability coefficients for the four index scores (VCI = .96, PRI = .95, WMI = .94, PSI = .90) and the FSIQ (r_{xx} = .98).

The purported factor structure of the WAIS-IV (outlined graphically in Figure 3.1) has been confirmed with the standardization sample (Canivez & Watkins, 2010a, 2010b; Wechsler, 2008b) and in independent samples (e.g., Nelson et al., 2013). Intercorrelations of subtest and composite scores provide evidence of construct validity, as subtests contributing to a specific index scale have higher correlations with each other than with subtests comprising other scales (Weschler, 2008b). Further evidence for the construct validity of the WAIS-IV is established in

the technical manual using a series of confirmatory factor analyses (CFA), as well as by other scholars using exploratory factor analysis (EFA; e.g., Canivez & Watkins, 2010a, 2010b). Moreover, concurrent validity for the composite scores has been established through comparisons between the WAIS-IV and other cognitive measures. The technical manual reports a correlation of .94 between the WAIS-IV FSIQ and the WAIS-III FSIQ.

Woodcock Johnson III Tests of Achievement

The WJ-III-ACH (Woodcock et al., 2001) is a comprehensive academic assessment battery designed to measure five academic domains: reading, mathematics, written language, oral language, and academic knowledge. The WJ-III-ACH is comprised of 22 subtests that combine to provide 17 composites and a total achievement composite score. All scores are reported as standard scores with a mean of 100 and a standard deviation of 15.

Normative and psychometric data for the WJ-III-ACH can be found in the WJ-III technical manuals (McGrew et al., 2007; McGrew & Woodcock, 2001). The WJ-III Normative Update (WJ-III-NU) was standardized on a nationally representative sample (N = 8,782) in accordance with the 2000 United States Census. Median internal consistency estimates ranged from .90 to .93 for the composite achievement scores included in the current study. The technical manual also provides evidence for test-retest reliability; correlations were high for all composite scores, with an average coefficient of .96 across a one-year assessment interval.

Evidence for the concurrent validity of the WJ-III-ACH is established in the technical manual using correlations between the WJ-III-ACH and other academic achievement measures (e.g., Kaufman Test of Educational Achievement [KTEA]; Wechsler Individual Achievement Test [WIAT]). Correlations ranged from .66 to .82 for the WJ-III-ACH reading, mathematics, and writing composite scores with comparable scores on the KTEA and WIAT. Unfortunately,

construct validity data are lacking; the CFA conducted by the test authors was not comprehensive and no independent investigations have examined the factor structure of the WJ-III-ACH. As Kamphaus (2009) points out, however, the absence of a comprehensive construct validity study for academic achievement measures is not uncommon.

For the purposes of the current study, the Academic Fluency, Basic Reading Skills, Basic Writing Skills, and Math Calculation Skills composite scores were used. The subtests comprising each of the composites are listed in Table 3.2. The rationale for selecting these four composite scores was threefold. First, the four composites are comprised of the same subtests in both the WJ-III-ACH and WJ-IV-ACH, and all of these subtests are comparable (in terms of the actual tasks and what they measure) between the two test editions. Second, it was also based on the availability of scores in our sample, in that we had more scores for these composites compared to other composites. This was important because it reduced the number of cases with missing variables. Lastly, and perhaps most importantly, the Academic Fluency, Basic Reading Skills, Basic Writing Skills, and Math Calculation Skills composites were used as criterion variables because they provide a comprehensive and reliable estimate of academic achievement. Indeed, scholars (e.g., Villarreal, 2015; Ysseldyke & Nelson, 2012) recommend using WJ ACH composite scores over subtest scores for interpretation because the composite scores have higher reliabilities (.92–.97) and meet minimum expectations for scores used to make clinical decisions. Additionally, using composite scores rendered a more parsimonious presentation of results than would using numerous subtest scores (McGill, 2013). Although the WJ-ACH provides users with several additional composite scores (e.g., Reading Comprehension, Phoneme-Grapheme Knowledge, Broad Achievement), they were not included in this study due to reasons aforementioned (i.e., subtests differ between test versions, missing variables).

Woodcock Johnson IV Tests of Achievement

The WJ-IV-ACH (Schrank et al., 2014) is a revised and updated version of the WJ-III-ACH. As with the third edition, the WJ-IV-ACH is a comprehensive academic assessment battery designed to measure reading, mathematics, written language, oral language, and academic knowledge. The assessment is comprised of 20 subtests that combine to provide 19 composites and a broad achievement score. All scores are reported as standard scores with a mean of 100 and a standard deviation of 15.

Normative and psychometric data for the WJ-IV-ACH can be found in the technical manual (McGrew et al., 2014). The WJ-IV-ACH was standardized on a nationally representative sample (N = 7,416) in accordance with the 2010 United States Census. Regarding reliability, mean internal consistency estimates for the 20–29 age group ranged from .94 to .96 for the composite achievement scores included in the current study.

Evidence for the construct validity of the WJ-IV-ACH is established in the technical manual via intercorrelations among scores and using EFA and CFA methods. Notably, intercorrelations are consistently higher between composites from the same achievement domain and lower between composites from different domains, providing evidence of construct validity. The test authors also provide evidence for the concurrent validity of the WJ-IV-ACH using correlations with other academic achievement measures (e.g., KTEA-II; WIAT-III). Reading, mathematics, and written language composite scores on the WJ-IV-ACH showed strong correlations between comparable composite scores on the KTEA-II and WIAT-III, with all correlations exceeding .80.

For the purposes of the current study, the Academic Fluency, Basic Reading Skills, Basic Writing Skills, and Math Calculation Skills composites were used. The subtests comprising each

of the composites are listed in Table 3.2. The rationale for selecting these four composite scores is explained in the previous section.

Procedure

Following approval by the Institutional Review Board, archival data from the completed psychoeducational evaluations were retrieved from an internal database used to store deidentified assessment information for all individuals evaluated at the clinic. The database exists for research purposes and is only accessible by psychologists and administration at the clinic. All participants gave written informed consent to use their evaluation data for research purposes at the time of their evaluation. Either licensed psychologists or masters-level clinicians or doctoral students under the supervision of licensed psychologists administered all assessments. Data were extracted for participants who were administered the WAIS-IV.

Data Analysis

Data analyses for the current study occurred in several steps. The preliminary procedures and specific analyses for the two research questions are discussed in detail below.

Preliminary Analyses

Dependent Variables. For the purpose of data analysis, the WJ-III-ACH and WJ-IV-ACH data were combined for the Academic Fluency, Basic Reading Skills, Basic Writing Skills, and Math Calculation Skills composites. No participants were administered both versions of the WJ-ACH, so the WJ-III and WJ-IV scores were combined across all participants to create a single variable for each of the four achievement composites in this study. For example, the WJ-III-ACH Academic Fluency scores and the WJ-IV-ACH Academic Fluency scores were combined into a single WJ-ACH Academic Fluency variable.

Group Assignment. As recommended in assessment guidebooks for the WAIS-IV (e.g., Groth-Marnat & Wright, 2016; Lichtenberger et al., 2012; Weschler, 2008a), six pairwise comparisons for the four WAIS-IV index scores (e.g., VCI–PRI, VCI–WMI) were created for all cases in order to identify those with "significant" index score variability (i.e., a 23-point or higher discrepancy between their highest and lowest index standard scores). Out of 1,211 total cases, 616 (51%) individuals presented with significant variability in their WAIS-IV index scores. These participants comprised the Scattered Group. The remaining 595 (49%) participants without significant index score variability comprised the Flat Group.

Power Analysis. An a priori power analysis was conducted using G*Power (Faul et al., 2009), with an alpha coefficient of .05, a power value of .80, and a medium effect size ($f^2 = .15$). The estimate indicated that a minimum sample size of 85 was necessary to detect a significant model using a hierarchical multiple regression analysis with one independent variable (WAIS-IV FSIQ score) in the first block and four independent variables (WAIS-IV index scores) in the second block. Accordingly, based on the power analyses, as well as previous empirical investigations of the incremental validity of the FSIQ, it was expected that both groups would yield medium to large effect sizes with group sample sizes of 616 (Scattered) and 595 (Flat).

Descriptive Statistics. Prior to investigating the research questions, descriptive statistics were computed for the independent (WAIS-IV variables; see Figure 3.1) and dependent variables (WJ-ACH variables; see Table 3.2) in the study. Specifically, means and standard deviations were calculated for all WAIS-IV and WJ-ACH variables for the two groups. This allowed us to compare the cognitive and academic achievement scores of our group samples to the normative population and to identify general patterns in the data. Descriptive statistics by gender, age, year in school, race/ethnicity, and disability type for cognitive ability and academic achievement were

also evaluated for both groups. Moreover, data were assessed for normality (e.g., skewness, kurtosis), outliers, linearity, and multicollinearity.

Research Question 1: Impact of Scatter on Structural Validity

Exploratory Factor Analyses. Modeled after McGill's (2016) study, the current study utilized principal axis EFA to examine the effects of significant scatter on the structural validity of the WAIS-IV FSIQ. First, the intercorrelation matrices for the two groups (Scattered and Flat) were evaluated using Bartlett's test of sphericity (Bartlett, 1954) and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (Kaiser, 1974) to ensure that the matrices were suitable for factor analysis. Communality estimates were also evaluated to assess how well each variable was explained by the factors.

Second, as recommended by Gorsuch (1983), multiple criteria for determining the number of factors to retain were examined for both groups. These procedures included the visual scree test (Cattell, 1966), eigenvalues > 1 (Guttman, 1954), Horn's parallel analysis (HPA; Horn, 1965), and minimum average partials (MAP; Velicer, 1976). The visual scree test was used to visually investigate the optimal number of factors to retain. Given the subjectivity of the scree test, HPA and MAP procedures were also included to minimize the threat of overfactoring, as recommended by Frazier and Youngstrom (2007).

Third, the factors were extracted using principal axis factoring followed by a promax rotation ($k = 4$). Factors with at least two saliently loaded subtests (≥ .30) were considered viable (Child, 2006). Factor loadings were also examined for compliance with theoretical and simple structure.

Schmid-Leiman Hierarchical Analyses. To apportion subtest variance to the higher- and first-order factors, the four-factor solution was subjected to higher-order EFA and

transformed with the Schmid-Leiman (SL) orthogonalization procedure (Schmid & Leiman, 1957). Carroll (1995) argued that cognitive test subtests are influenced by both first-order factors (i.e., index scores) and the higher-order general factor (i.e., FSIQ), and thus recommended the use of the SL transformation of EFA loadings to apportion subtest variance to the higher- and first-order dimensions. This procedure has also been used in numerous Wechsler scale EFA investigations (e.g., Canivez & Watkins, 2010a, 2010b; Canivez et al., 2019, 2020; Dombrowski et al., 2022; Watkins & Canivez, 2022). Accordingly, in the present study, the oblique EFA factor solutions for the Flat and Scattered Groups were transformed with the SL orthogonalization procedure using the *MacOrtho* program (Watkins, 2020).

In the derived hierarchical SL factor model, the general and first-order factors were all directly related to the subtests and uncorrelated with each other (i.e., orthogonal; Watkins & Canivez, 2022). This enabled us to estimate the direct and unique influence of each of the five factors (FSIQ and four index) on each subtest score, thereby enhancing the interpretability of the higher- and first-order factors (Brunner et al., 2012). Additionally, by first extracting the variability associated with the *g* factor and then with the first-order factors, we were able to examine direct estimates of the proportion of variance accounted for by the general factor separately from the four first-order factors (Carroll, 1993). These results were examined both within and between the Flat and Scattered Groups.

Omega Estimates. Next, omega-hierarchical (ω_h) and omega-subscale (ω_s) were estimated as model-based reliability estimates of the latent factors (Reise, 2012). Many scholars (e.g., Chen et al., 2012; Hayes & Coutts, 2020; Watkins, 2017) recommend using omega coefficients as a measure of reliability over Chronbach's alpha, especially with multidimensional instruments like the WAIS-IV. In this context, the ω_h coefficient reflects the general intelligence

factor (i.e., FSIQ) reliability estimate with variability from the group factors (i.e., WAIS-IV index scores) removed, and the ω_s coefficient reflects the reliability estimate of a group factor with variability from all other group and general factors removed (Brunner et al., 2012; Reise, 2012; Zinbarg et al., 2006).

In the present study, omega estimates for the Flat and Scattered Groups were calculated from the orthogonalized SL EFA solution and produced using the *Omega* program (Watkins, 2013). Although subjective as there is no consensus on what constitutes adequate reliability (Watkins & Canivez, 2022), omega coefficients of .75 or higher were considered preferred, and those exceeding .50 were considered useful in determining whether a factor score provides unique, reliable variance (Gignac & Watkins, 2013; Reise et al., 2013; Rodriguez et al., 2016).

Research Question 2: Impact of Scatter on Incremental Validity

Hierarchical Multiple Regression Analyses. In line with previous research (e.g., Freberg et al., 2008; Kotz et al., 2008; McGill, 2013, 2016), the present study used hierarchical multiple regression analyses to determine how well the WAIS-IV FSIQ predicts academic achievement in the presence of significant scatter. Separate regression analyses were conducted for the Flat and Scattered Groups, as well as for each of the four WJ-ACH composite scores (for a total of eight models). The achievement variables—Academic Fluency, Basic Reading Skills, Basic Writing Skills, and Basic Mathematics Skills—served as the dependent variables in this study. As previously stated, the data for both groups were examined to ensure that assumptions for multiple regression analysis were met.

For each regression model, the WAIS-IV FSIQ was singularly entered into the first block and the four WAIS-IV index scores (VCI, PRI, WMI, and PSI) were entered both jointly and individually into the second block. The change in predicted achievement test score variance (R^2)

produced by the four WAIS-IV index scores in the second block provided an estimate of the incremental prediction beyond the FSIQ from the first block. Unique incremental contribution of index score prediction of achievement score variance (after partitioning out FSIQ variance) was estimated by the squared part correlations from the second block. Further, the resulting R^2 coefficients were interpreted as an effect size, with small = .01, medium = .09 and large = .25 (Cohen, 1988).

Table 3.1. *Diagnostic Categories of the Clinical Sample.*

Disability	*n*[a]
ADHD	572
Learning Disorder	506
Anxiety Disorder	287
Depressive Disorder	139
Autism Spectrum Disorder	61
Cognitive / Neurocognitive Disorder	27
Communication Disorder	22
Obsessive Compulsive Disorder	17
Post-Traumatic Stress Disorder	5
Tic / Tourette's Disorder	5
Bipolar Disorder	4
Other	40
No Diagnosis	*108*

Note. ADHD = attention-deficit/hyperactivity disorder.

[a] The total number of diagnoses is greater than the number of participants given that numerous participants had more than one diagnosis.

Table 3.2. *Subtests Comprising the WJ-ACH Composites.*

	WJ-III-ACH	WJ-IV-ACH
Composite	Subtests	Subtests
Basic Reading Skills	Letter-Word Identification	Letter-Word Identification
	Word Attack	Word Attack
Math Calculation Skills	Calculation	Calculation
	Math Fluency	Math Facts Fluency
Academic Fluency	Reading Fluency	Sentence Reading Fluency
	Math Fluency	Math Facts Fluency
	Writing Fluency	Sentence Writing Fluency

Note. WJ-III-ACH = Woodcock Johnson III Tests of Achievement (Woodcock et al., 2001); WJ-IV-ACH = Woodcock Johnson IV Tests of Achievement (Schrank et al., 2014).

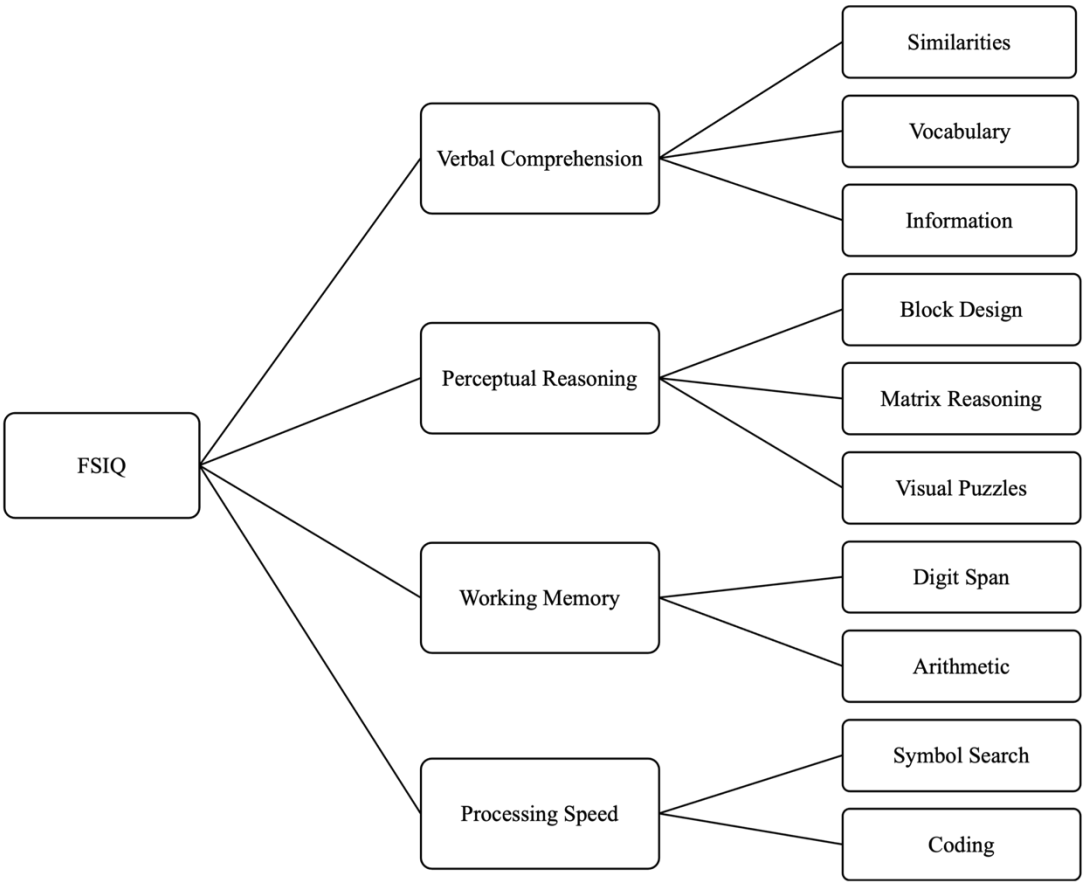

Figure 3.1. Indirect hierarchical model for the Weschler Intelligence Scales, Fourth Edition (adapted from Figure 5.1 [Weschler, 2008].

REFERENCES

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.).

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Bartlett, M. S. (1954). A further note on the multiplying factors for various chi square approximations in factor analysis. *Journal of the Royal Statistical Society Series A (General)*, *16*, 296–298.

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*, 796–846. https://doi.org/10.1111/j.1467-6494.2011.00749.x

Canivez, G. L., McGill, R. J., Dombrowski, S. C., Watkins, M. W., Pritchard, A. E., & Jacobson, L. A. (2020). Construct validity of the WISC-V in clinical cases: Exploratory and confirmatory factor analyses of the 10 primary subtests. *Assessment*, *27*(2), 274–296. https://doi.org/10.1177/1073191118811609

Canivez, G. L., & Watkins, M. W. (2010a). Exploratory and higher-order factor analyses of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) adolescent subsample. *School Psychology Quarterly*, *25*, 223–235. https://doi.org/10.1037/a0022046

Canivez, G. L., & Watkins, M. W. (2010b). Investigation of the factor structure of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV): Exploratory and higher order factor analyses. *Psychological Assessment*, *22*, 827–836. https://doi.org/10.1037/a0020429

Canivez, G. L., Watkins, M. W., & McGill, R. J. (2019). Construct validity of the Wechsler Intelligence Scale for Children–Fifth UK Edition: Exploratory and confirmatory factor analyses of the 16 primary and secondary subtests. *British Journal of Educational Psychology*, *89*, 195–224. https://doi.org/10.1111/bjep.12230

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

Carroll, J. B. (1995). On methodology in the study of cognitive abilities. *Multivariate Behavioral Research*, *30*, 429-452. https://doi.org/10.1207/s15327906mbr3003_6

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J.-P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality*, *80*(1), 219–251. https://doi.org/10.1111/j.1467-6494.2011.00739.x

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Dombrowski, S. C., McGill, R. J., Watkins, M. W., Canivez, G. L., Pritchard, A. E., & Jacobson, L. A. (2022). Will the real theoretical structure of the WISC-V please stand up? Implications for clinical interpretation. *Contemporary School Psychology*, *26*, 492–503. https://doi.org/10.1007/s40688-021-00365-6

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (2009). Statistical power analysis using G*Power 3.1: Tests for correlation and regression analysis. *Behavioral Research Methods, 41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Frazier, T. W., & Youngstrom, E. A. (2007). Historical increase in the number of factors measured by commercial tests of cognitive ability: Are we overfactoring? *Intelligence*, *35*, 169–182. https://doi.org/10.1016/j.intell.2006.07.002

Freberg, M. E., Vandiver, B. J., Watkins, M. W., & Canivez, G. L. (2008). Significant factor score variability and the validity of the WISC-III Full Scale IQ in predicting later academic achievement. *Applied neuropsychology*, *15*(2), 131–139. https://doi.org/10.1080/09084280802084010

Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research*, *48*, 639–662. https://doi.org/10.1080/00273171.2013.804398

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Erlbaum.

Groth-Marnat, G., & Wright, A. J. (2016). *Handbook of psychological assessment* (6th ed.). Wiley.

Guttman, L. (1954). Some necessary and sufficient conditions for common factor analysis. *Psychometrika*, *19*, 149–161. https://doi.org/10.1007/BF02289162

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures*, *14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185. https://doi.org/10.1007/BF02289447

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 31–36. https://doi.org/10.1007/BF02291575

Kamphaus, R. W. (2009). Assessment of intelligence and achievement. In T. B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology* (4th ed.; pp. 230–246). Wiley.

Kotz, K. M., Watkins, M. W., & McDermott, P. A. (2008). Validity of the General Conceptual Ability score from the Differential Ability Scales as a function of significant and rare interfactor variability. *School Psychology Review*, *37*(2), 261–278.

Lichtenberger, E. O., Kaufman, A. S., & Kaufman, N. L. (2012). *Essentials of WAIS-IV assessment* (2nd ed.). Wiley.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19–40. https://doi.org/10.1037/1082-989X.7.1.19

McGill, R. J. (2013). Beyond g: Assessing the incremental validity of the Cattell-Horn-Carroll (CHC) broad ability factors on the Woodcock-Johnson III Tests of Cognitive Abilities (Publication No. 3621595) [Doctoral dissertation, Chapman University]. ProQuest Dissertations & Theses Global.

McGill, R. J. (2016). Invalidating the full scale IQ score in the presence of significant factor score variability: Clinical acumen or clinical illusion? *Archives of Assessment Psychology*, *6*(1), 49–79.

McGill, R. J. (2018). Confronting the base rate problem: More ups and downs for cognitive scatter analysis. *Contemporary School Psychology*, *22*, 384–393. https://doi.org/10.1007/s40688-017-0168-4

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical manual: Woodcock-Johnson IV*. Riverside.

McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). *Technical manual: Woodcock-Johnson III Normative Update*. Riverside.

McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock-Johnson III technical manual*. Riverside.

Nelson, J. M., Canivez, G. L., & Watkins, M. W. (2013). Structural and incremental validity of the Wechsler Adult Intelligence Scale–Fourth Edition with a clinical sample. *Psychological Assessment*, *25*(2), 618–630. https://doi.org/10.1037/a0032086

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*(2), 129–140. https://doi.org/10.1080/00223891.2012.725437

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150. https://doi.org/10.1037/met0000045

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*, 53–61. https://doi.org/10.1007/BF02289209

Schrank, F. A., Mather, N., & McGrew, K. S. (2014). *Woodcock-Johnson IV Tests of Achievement*. Riverside.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *31*, 321–327. https://doi.org/10.1007/BF02293557

Villarreal, V. (2015). Test review: Woodcock-Johnson IV Tests of Achievement. *Journal of Psychoeducational Assessment*, *33*, 391–398.

Watkins, M. W. (2013). *Omega* [Computer software]. Ed & Psych Associates.

Watkins, M. W. (2017). The reliability of multidimensional neuropsychological measures: From alpha to omega. *The Clinical Neuropsychologist*, *31*, 1113–1126. https://doi.org/10.1080/13854046.2017.1317364.

Watkins, M. W. (2020). *MacOrtho* [Computer software]. EdPsych Associates.

Watkins, M. W., & Canivez, G. L. (2022). Assessing the psychometric utility of IQ scores: A tutorial using the Wechsler Intelligence Scale for Children–Fifth Edition. *School Psychology Review*, *51*(5), 619–633. https://doi.org/10.1080/2372966X.2020.1816804

Wechsler, D. (2008a). *Wechsler Adult Intelligence Scale* (4th ed.). Pearson.

Wechsler, D. (2008b). *Wechsler Adult Intelligence Scale–Fourth Edition technical and interpretive manual*. Pearson.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Riverside.

Ysseldyke, J., & Nelson, P. (2012). Assessment in special and inclusive education. In J. Banks (Ed.), *Encyclopedia of diversity in education* (pp. 165–168). Sage.

Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ω_h. *Applied Psychological Measurement*, *30*, 121–144. https://doi.org/10.1177/0146621605278814

CHAPTER 4

RESULTS

The present study investigated the impact of significant factor score variability on the structural and incremental validity of the Weschler Adult Intelligence Scales, Fourth Edition (WAIS-IV; Weschler, 2008a) in a clinically referred sample of college students. Since users are encouraged to forego interpretation of the full-scale intelligence quotient (FSIQ) if significant scatter is observed among index scores (e.g., Groth-Marnat & Wright, 2016; Hale & Fiorello, 2004; Lezak et al., 2012), there is a critical need to provide psychologists with information regarding the psychometric validity of this interpretive heuristic. Accordingly, the purpose of the present study was to (1) determine whether hierarchical structure remains plausible for clinically referred individuals presenting with significant levels of score variability, and (2) evaluate the incremental validity of the WAIS-IV FSIQ in accounting for meaningful portions of Woodcock-Johnson Tests of Achievement (WJ-ACH) achievement variance when significant scatter is observed. The results of data analyses investigating these two topics of study are presented in this chapter. Before examining the results of the research questions, we briefly review the results of the preliminary analyses.

Preliminary Analyses

Pairwise comparisons revealed that 616 (51%) out of the 1,211 total participants presented with significant variability in their WAIS-IV index scores. These 616 individuals comprised the Scattered Group, while the remaining 595 participants comprised the Flat Group.

Descriptive statistics for the WAIS-IV index scores by group (Flat and Scattered) are reported in Table 4.1. WAIS-IV index score means for both groups were slightly below average, albeit within 1 standard deviation, in comparison to the standardization sample. Mean scores were also lower than those typically obtained by college students and young adults (Beaujean et al., 2010; Lassiter et al., 2001; Scheiber et al., 2017). These descriptive statistics were expected given that lower cognitive scores are generally observed in clinically referred samples (Canivez et al., 2020; Lichtenberger et al., 2012; Watkins, 2010). Additionally, previous research has indicated that individuals with attention-deficit/hyperactivity disorder (ADHD) and learning disabilities—the two most prevalent diagnoses among the participants in the present study—on average, score lower on working memory and processing speed measures than those without these disorders (e.g., Frazier et al., 2004; Miller et al., 2016; Swanson & Hsieh, 2009). As illustrated in Table 4.1, the results of the current study are consistent with this finding; mean scores for the Flat and Scattered Groups were the lowest on the Working Memory Index and the Processing Speed Index.

WAIS-IV subtest intercorrelation matrices and descriptive statistics for the Flat Group and the Scattered Group are presented in Table 4.2 and Table 4.3, respectively. Regarding the Flat Group, score distributions were relatively normal for all index and subtest scores, with the exception of the Coding subtest, which had a positive kurtosis of 1.16. Regarding the Scattered Group, WAIS-IV index and subtest score distributions were relatively normal, with 0.62 the largest univariate skew and -0.72 the largest univariate kurtosis. As such, the data for both groups were appropriate for data analysis.

Research Question 1: Impact of Scatter on Structural Validity

Research question one in the present study investigated whether significant factor score variability impacts the structural validity of the WAIS-IV FSIQ using hierarchical exploratory factor analysis (EFA) procedures. Following examination of the factor structures for the Flat and Scattered Groups, orthogonalization procedures were used to examine the proportions of WAIS-IV subtest variance attributed to the higher- and first-order factors, thus enabling us to better understand the underlying structure of the WAIS-IV in the presence of scatter (Gignac, 2007). The results of these analyses are detailed below, first with the Flat Group and then followed by the Scattered Group.

Structural Validity of the Flat Group

Exploratory Factor Analyses. EFA for the Flat Group was performed using the principal axis factor extraction method. To ensure that the dataset was suitable for factor analysis, Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy were evaluated. Bartlett's test of sphericity was significant (χ^2 (45) = 2673.15, $p <$ .001), indicating the presence of correlations among variables (Bartlett, 1954). The KMO coefficient of .94 was well above the minimum standard for conducting factor analysis (Kaiser, 1974). Communality estimates ranged from .404 (Digit Span) to .894 (Vocabulary), further confirming that each variable shared some common variance with the other variables. Given these values, it was determined that the dataset for the Flat Group was suitable for factor analysis.

Multiple criteria were examined to determine the number of factors to retain for the Flat Group (Gorsuch, 1983). Although factor selection criteria (e.g., eigenvalues, visual scree test) suggested retention of one or two factors, the WAIS-IV *Technical and Interpretive Manual*

(Wechsler, 2008b) claimed four factors. Prior research with the WAIS-IV also indicated that four factors would suffice (e.g., Weiss et al., 2013). Given that it is better to overfactor than underfactor (Wood et al., 1996), four factors were extracted for the Flat Group to correspond with the theoretical structure defined in the WAIS-IV manual.

Table 4.4 displays the results from extracting four WAIS-IV factors followed by promax ($k = 4$) rotation. The *g* loadings (factor structure coefficients from the first unrotated factor) ranged from .613 (Matrix Reasoning) to .811 (Vocabulary) and were all in the fair to good range based on Kaufman's (1994) criteria (≥ .70 = good, .50-.69 = fair, < .50 = poor). Additionally, most WAIS-IV subtests were saliently (≥ .30) and properly associated with their theoretical factor. However, the Information and Arithmetic subtests loaded saliently on more than one factor (i.e., cross-loaded), and the Matrix Reasoning subtest had a high factor pattern coefficient (.734) on a theoretically inconsistent factor. As such, the factor pattern was inconsistent with the WAIS-IV theoretical structure. Nonetheless, the high factor intercorrelations (.63 to .77) displayed in Table 4.4 suggests the presence of a higher-order (general intelligence) factor structure requiring further explication (Carroll, 1993; Gorsuch, 1983).

Schmid-Leiman Hierarchical Analyses. To apportion variance associated with the higher-order general factor, the four first-order oblique EFA factor solution (Table 4.4) was transformed with the Schmid-Leiman (SL) orthogonalization procedure (Schmid & Leiman, 1957). Results for the higher-order factor analysis of four first-order WAIS-IV factors are presented in Table 4.5. All but two subtests (Matrix Reasoning and Digit Span) were properly associated (i.e., higher residual variance) with their theoretically proposed factor after removing *g* variance. The hierarchical *g* factor accounted for 52.7% of the total variance and 90.6% of the

common variance. The general factor also accounted for between 43.3% (Digit Span) and 62.7% (Vocabulary) of individual subtest variability.

At the first-order level, the Verbal Comprehension factor accounted for an additional 3.2% of the total variance and 5.5% of the common variance, the Processing Speed factor accounted for an additional 1% of the total variance and 1.8% of the common variance, the Perceptual Reasoning factor accounted for an additional 0.9% of the total variance and 1.6% of the common variance, and the Working Memory factor accounted for an additional 0.3% of the total variance and 0.5% of the common variance. The general and index factors combined to measure 58.1% of the variance in WAIS-IV scores resulting in 41.9% unique variance (combination of specific and error variance). In comparison to the first-order index factors, the higher order *g* factor accounted for substantially greater portions of WAIS-IV total and common variance.

Omega Estimates. Based on the SL results in Table 4.5, omega-hierarchical (ω_h) and omega-subscale (ω_s) coefficients were estimated. The ω_h coefficient for general intelligence (.908) was high and sufficient for scale interpretation. The ω_s coefficients for the four WAIS-IV group factors were considerably lower, ranging from .017 to .109. This suggests that the factor index scores likely possess too little unique variance for clinical interpretation (Reise, 2012; Reise et al., 2013).

Structural Validity of the Scattered Group

Exploratory Factor Analyses. As with the Flat Group, EFA for the Scattered Group was performed using the principal axis factor extraction method. Bartlett's test of sphericity was significant (χ^2 (45) = 2362.15, $p < .001$), indicating that the correlation matrix was not random. The KMO coefficient of .76 was above the minimum standard for factor analysis (Kaiser, 1974).

Communality estimates ranged from .361 (Matrix Reasoning) to .547 (Visual Puzzles), further confirming that each variable shared some common variance with the other variables. Given these values, it was determined that the dataset for the Scattered Group was suitable for factor analysis (Kaiser, 1974).

Multiple criteria were examined to determine the number of factors to retain for the Scattered Group (Gorsuch, 1983). In line with the factor model presented in the WAIS-IV *Technical and Interpretive Manual* (Wechsler, 2008b), factor selection criteria (e.g., eigenvalues, visual scree test) recommended retention of four factors. As such, four factors were extracted for the Scattered Group to correspond with the theoretical structure defined in the WAIS-IV manual.

Table 4.6 displays the results from extracting four WAIS-IV factors followed by promax ($k = 4$) rotation. The *g* loadings (factor structure coefficients from the first unrotated factor) ranged from .348 (Coding) to .711 (Visual Puzzles), with most subtests in the fair to good range based on Kaufman's (1994) criteria (≥ .70 = good, .50-.69 = fair, < .50 = poor). Additionally, all WAIS-IV subtests were saliently (≥ .30) and properly associated with their theoretical factor, thereby achieving desired simple structure. The factor intercorrelations, which ranged from .13 to .50 (Table 4.6), suggest the presence of a higher-order (general intelligence) factor structure requiring further explication (Carroll, 1993; Gorsuch, 1983).

Schmid-Leiman Hierarchical Analyses. To apportion variance associated with the higher-order general factor, the four first-order oblique EFA factor solution (Table 4.6) was transformed with the SL orthogonalization procedure (Schmid & Leiman, 1957). Results for the higher-order factor analysis of four first-order WAIS-IV factors are presented in Table 4.7. All subtests were properly associated (i.e., higher residual variance) with their theoretically proposed factor after removing *g* variance. The hierarchical *g* factor accounted for 29.1% of the total

variance and 46.7% of the common variance. The general factor also accounted for between 8.5% (Coding) and 47.5% (Arithmetic) of individual subtest variability.

At the first-order level, the Verbal Comprehension factor accounted for an additional 10.8% of the total variance and 17.4% of the common variance, the Processing Speed factor accounted for an additional 11.4% of the total variance and 18.2% of the common variance, the Perceptual Reasoning factor accounted for an additional 6% of the total variance and 9.6% of the common variance, and the Working Memory factor accounted for an additional 5% of the total variance and 8% of the common variance. The general and group factors combined to measure 62.3% of the variance in WAIS-IV scores resulting in 37.7% unique variance (combination of specific and error variance). In comparison to the first-order group factors, the higher order *g* factor accounted for substantially greater portions of WAIS-IV total and common variance.

Omega Estimates. Based on the SL results in Table 4.7, omega-hierarchical (ω_h) and omega-subscale (ω_s) coefficients were estimated. The ω_h coefficient for general intelligence (.699) was high and sufficient for scale interpretation. The ω_s coefficients for the four WAIS-IV group factors were lower, ranging from .267 to .679. While the coefficient obtained for Processing Speed (.679) was sufficient for interpretation, the three remaining factor index scores likely possess too little unique variance for clinical interpretation (Reise, 2012; Reise et al., 2013).

Summary of Results: Impact of Scatter on Structural Validity

Research question one in the present study used hierarchical EFA procedures to investigate the impact of significant factor score variability on the structural validity of the WAIS-IV in a clinically referred sample. EFA results revealed that the desired simple structure was achieved for the Scattered Group. EFA transformation with the SL orthogonalization

procedure showed that all WAIS-IV subtests remained properly associated with their theoretical factor after removing *g* variance. The higher-order general factor also accounted for substantially greater portions of WAIS-IV total and common variance as compared to the four first-order group factors. Additionally, omega estimates for the Scattered Group revealed that the omega coefficient for FSIQ was high and sufficient for scale interpretation. The omega coefficient for Processing Speed was also high, whereas the coefficients for the Verbal Comprehension, Perceptual Reasoning, and Working Memory factors were considerably lower. A detailed discussion of these results as well as the implications of the results of the present study is presented in the subsequent chapter.

Research Question 2: Impact of Scatter on Incremental Validity

Research question two in the present study investigated whether significant scatter impacts the incremental validity of the WAIS-IV FSIQ using hierarchical multiple regression analysis. The Academic Fluency, Basic Reading Skills, Basic Writing Skills, and Math Calculation Skills composites from the WJ-ACH were used as measures of academic achievement in this study. As previously explained, each of the four variables include scores from both the Woodcock-Johnson III Tests of Achievement (Woodcock et al., 2001) and the Woodcock-Johnson IV Tests of Achievement (Schrank et al., 2014).

Descriptive statistics for the WJ-ACH achievement variables by group (Flat and Scattered) are reported in Table 4.8. For both groups, the mean (92.96 to 98.59) and standard deviation (11.45 to 14.99) ranges generally reflect values that would be expected for normally distributed standard score variables (McGrew et al., 2007, 2014). Skewness values ranging from -0.68 to -0.01 provided some evidence of relatively normal score distributions. Additionally, visual inspection of the histogram distributions and probability plots suggested normal

distributions for all WJ-ACH variables. Inspection of the residual plots of the data also indicated that the regression models utilized in this study met the assumptions for homoscedasticity of the residuals. As such, the data for both the Flat and Scattered Groups met assumptions for multiple regression analysis.

Hierarchical Multiple Regression Analyses

Table 4.9 presents results from hierarchical multiple regression analyses for the Flat Group and the Scattered Group. The regression models illustrate the proportions of WJ-ACH achievement variance accounted for by the WAIS-IV FSIQ in the first block, followed by the unique incremental variance accounted for by the four WAIS-IV index scores (Verbal Comprehension Index [VCI], Perceptual Reasoning Index [PRI], Working Memory Index [WMI], Processing Speed Index [PSI]) in the second block. The change in predicted achievement variance (R^2) from the second block provided the estimate of the incremental prediction provided by the WAIS-IV factors both jointly and individually. Cohen's (1988) guidelines for interpreting R^2 as an effect size (small = .01, medium = .09 and large = .25) were used.

Flat Group. The WAIS-IV FSIQ accounted for large and statistically significant ($p <$ .001) effects across the WJ-ACH Academic Fluency (45.2%), Basic Reading (32.3%), Basic Writing (26.1%), and Math Calculation (46.8%) scores. Incremental predictive contributions (after controlling for the effects of the FSIQ) of the four WAIS-IV index scores combined were also statistically significant across all of the achievement variables for the Flat Group. Statistically significant ($p < .001$) portions of WJ-ACH Academic Fluency (9.4%), Basic Reading (7.2%), Basic Writing (12.6%), and Math Calculation (5.2%) score variance was incrementally accounted for by the WAIS-IV factor index scores that represented small to

medium effect sizes. Additionally, the unique contributions of WAIS-IV factor index scores in predicting each of the four WJ-ACH composite scores were as follows: Academic Fluency ranged from 0.0% (PRI) to 1.1% (PSI), Basic Reading Skills ranged from 0.5% (PRI) to 2.0% (WMI), Basic Writing Skills ranged from 0.3% (PRI) to 2.3% (WMI), and Math Calculation Skills ranged from 0.2% (VCI and PRI) to 1.2% (WMI).

Scattered Group. The WAIS-IV FSIQ accounted for large and statistically significant ($p < .001$) effects across the WJ-ACH Academic Fluency (34.8%), Basic Reading (32.0%), Basic Writing (28.5%), and Math Calculation (36.8%) scores. Incremental predictive contributions (after controlling for the effects of the FSIQ) of the four WAIS-IV index scores combined were also statistically significant across all of the achievement variables for the Scattered Group. Statistically significant ($p < .001$) portions of WJ-ACH Academic Fluency (22.8%), Basic Reading (13.8%), Basic Writing (20.7%), and Math Calculation (16.2%) score variance was incrementally accounted for by the WAIS-IV factor index scores that represented medium effect sizes. Additionally, the unique contributions of WAIS-IV factor index scores in predicting each of the WJ-ACH composite scores were as follows: Academic Fluency ranged from 0.0% (PRI) to 1.0% (PSI), Basic Reading Skills ranged from 0.1% (PRI) to 0.9% (WMI), Basic Writing Skills ranged from 0.3% (PRI) to 2.3% (WMI), and Math Calculation Skills ranged from 0.5% (VCI and PRI) to 2.4% (WMI).

Summary of Results: Impact of Scatter on Incremental Validity

Research question two in the present study used hierarchical multiple regression analysis to investigate the impact of significant factor score variability on the incremental validity of the WAIS-IV in a clinically referred sample. Results revealed that the WAIS-IV FSIQ accounted for large and statistically significant effects across the WJ-ACH Math Calculation, Academic

Fluency, Basic Reading, and Basic Writing scores for the Scattered Group. Although the incremental predictive contributions of the WAIS-IV index scores across the four WJ-ACH achievement variables were statistically significant, effect size estimates for these effects were moderate. The unique contributions of the individual WAIS-IV factor scores were consistently small for the Scattered Group. A detailed discussion of these results as well as the implications of the results of the present study is presented in the subsequent chapter.

Table 4.1. *Descriptive Statistics for the WAIS-IV Index Scores.*

	Flat Group (N = 595)			
Index	*M*	*SD*	*SK*	*K*
Verbal Comprehension Index	100.43	11.94	-0.12	-0.21
Perceptual Reasoning Index	98.45	12.14	-0.31	0.12
Working Memory Index	95.26	12.17	0.00	0.00
Processing Speed Index	96.11	11.19	-0.03	0.86
Full Scale IQ	97.41	12.72	-0.11	0.19
	Scattered Group (N = 616)			
Index	*M*	*SD*	*SK*	*K*
Verbal Comprehension Index	111.66	14.77	0.01	0.01
Perceptual Reasoning Index	101.66	14.35	-0.12	-0.34
Working Memory Index	94.92	14.96	0.35	-0.14
Processing Speed Index	91.29	13.72	0.53	0.46
Full Scale IQ	101.24	12.40	0.01	0.08

Note. WAIS-IV = Weschler Adult Intelligence Scale, Fourth Edition.

Table 4.2. *WAIS-IV Subtest Intercorrelations and Descriptive Statistics for the Flat Group (N = 595).*

Subtest	BD	SI	DS	MR	VO	AR	SS	VP	IN	CD
Block Design (BD)	–									
Similarities (SI)	.47	–								
Digit Span (DS)	.47	.37	–							
Matrix Reasoning (MR)	.40	.38	.44	–						
Vocabulary (VO)	.56	.60	.49	.44	–					
Arithmetic (AR)	.56	.49	.45	.51	.55	–				
Symbol Search (SS)	.52	.44	.45	.39	.50	.47	–			
Visual Puzzles (VP)	.63	.45	.43	.40	.54	.56	.50	–		
Information (IN)	.54	.44	.44	.44	.62	.56	.42	.53	–	
Coding (CD)	.47	.44	.43	.43	.44	.51	.51	.44	.41	–
Statistic	BD	SI	DS	MR	VO	AR	SS	VP	IN	CD
Mean	9.07	10.39	9.01	10.46	10.33	9.34	9.43	9.82	9.65	9.19
Standard Deviation	2.80	2.64	2.46	2.45	2.69	2.67	2.39	2.56	2.50	2.39
Skewness	0.12	-0.09	0.27	-0.53	-0.19	0.01	0.10	0.18	0.10	0.21
Kurtosis	-0.69	-0.03	0.41	0.61	-0.48	-0.62	0.37	-0.28	-0.63	1.16

Note. All coefficients were statistically significant ($p < .01$, two tailed). WAIS-IV = Weschler Adult Intelligence Scale, Fourth Edition.

Table 4.3. *WAIS-IV Subtest Intercorrelations and Descriptive Statistics for the Scattered Group (N = 616).*

Subtest	BD	SI	DS	MR	VO	AR	SS	VP	IN	CD
Block Design (BD)	–									
Similarities (SI)	.25*	–								
Digit Span (DS)	.25*	.23*	–							
Matrix Reasoning (MR)	.45*	.30*	.33*	–						
Vocabulary (VO)	.20*	.62*	.24*	.25*	–					
Arithmetic (AR)	.37*	.38*	.62*	.41*	.38*	–				
Symbol Search (SS)	.34*	.14*	.13*	.14*	.04	.19*	–			
Visual Puzzles (VP)	.68*	.29*	.27*	.52*	.24*	.40*	.32*	–		
Information (IN)	.29*	.51*	.20*	.34*	.64*	.41*	.13*	.35*	–	
Coding (CD)	.20*	.13*	.15*	.12*	.08	.18*	.66*	.20*	.07	–
Statistic	BD	SI	DS	MR	VO	AR	SS	VP	IN	CD
Mean	9.94	12.18	8.96	10.76	12.24	9.28	8.69	10.29	11.82	8.09
Standard Deviation	3.26	2.98	2.88	2.74	3.04	3.03	2.94	2.88	2.90	2.68
Skewness	0.26	-0.10	0.62	-0.50	-0.15	0.20	0.47	0.02	-0.07	0.48
Kurtosis	-0.48	-0.01	0.45	0.13	0.01	-0.72	0.28	-0.70	-0.20	0.51

Note. WAIS-IV = Weschler Adult Intelligence Scale, Fourth Edition.
* $p < .01$.

Table 4.4. *WAIS-IV Exploratory Factor Analysis for the Flat Group (N = 595): Four Oblique Factor Solution with Promax Rotation.*

		Factor Pattern Coefficients				Factor Structure Coefficients			
Subtest	g^a	F1	F2	F3	F4	F1	F2	F3	F4
Similarities	.649	**.396**	.045	.065	.240	.628	.569	.552	.569
Vocabulary	.811	**.981**	-.018	-.037	.007	.945	.700	.654	.588
Information	.712	**.252**	.386	.304	-.162	.659	.691	.664	.496
Block Design	.754	.009	**.754**	-.070	.124	.608	.798	.601	.631
Matrix Reasoning	.613	-.021	**-.103**	.734	.086	.480	.506	.702	.520
Visual Puzzles	.731	-.022	**.784**	-.042	.067	.583	.784	.589	.596
Digit Span	.624	.086	.089	**.298**	.235	.515	.554	.595	.567
Arithmetic	.750	.016	.320	**.445**	.048	.606	.708	.736	.609
Symbol Search	.676	.066	.180	-.057	**.587**	.533	.615	.545	.720
Coding	.653	-.043	-.014	.258	**.554**	.482	.556	.611	.700
Eigenvalue		0.74	5.32	0.67	0.64				
Variance (%)		7.36	53.15	6.74	6.40				
Factor Correlations		F1	F2	F3	F4				
	F1	–							
	F2	.76	–						
	F3	.71	.77	–					
	F4	.63	.73	.71	–				

Note. Factor loadings are denoted in bold. WAIS-IV = Weschler Adult Intelligence Scale, Fourth Edition.
[a] General structure coefficients are based on the first unrotated factor coefficients (*g* loadings).

Table 4.5. *Sources of Variance in the WAIS-IV for the Flat Group (N = 595) According to an Orthogonalized Higher-Order Factor Model.*

	General		F1: VC		F2: PR		F3: WM		F4: PS			
Subtest	*b*	Var	*b*	Var	*b*	Var	*b*	Var	*b*	Var	h^2	u^2
Similarities	.663	.440	.206	.042							.482	.518
Vocabulary	.792	.627	.510	.260							.887	.113
Information	.727	.529	.131	.017							.546	.454
Block Design	.779	.607			.209	.044					.651	.349
Matrix Reasoning	.660	.436			-.028	.001					.436	.564
Visual Puzzles	.756	.572			.217	.047					.619	.381
Digit Span	.658	.433					.092	.008			.441	.559
Arithmetic	.788	.621					.138	.019			.640	.360
Symbol Search	.713	.508							.235	.055	.564	.436
Coding	.703	.494							.222	.049	.543	.457
Total Variance (%)	52.7		3.2		0.9		0.3		1.0		58.1	41.9
Common Variance (%)	90.6		5.5		1.6		0.5		1.8		100	
$\omega_{h/s}$	.908		.109		.025		.017		.067			

Note. WAIS-IV = Weschler Adult Intelligence Scale, Fourth Edition; PR = Perceptual Reasoning; VC = Verbal Comprehension; WM = Working Memory; PS = Processing Speed; *b* = standardized loading of subtest on factor; Var = variance explained in the subtest; h^2 = communality; u^2 = uniqueness; ω_h = omega-hierarchical; ω_s = omega-subscale.

Table 4.6. *WAIS-IV Exploratory Factor Analysis for the Scattered Group (N = 616): Four Oblique Factor Solution with Promax Rotation.*

		Factor Pattern Coefficients				Factor Structure Coefficients			
Subtest	g^a	F1	F2	F3	F4	F1	F2	F3	F4
Similarities	.585	**.679**	.022	.016	.049	.703	.348	.366	.150
Vocabulary	.622	**.946**	-.119	-.020	-.013	.882	.285	.376	.065
Information	.622	**.698**	.138	-.043	-.020	.736	.418	.360	.113
Block Design	.636	-.032	**.783**	-.024	.051	.309	.775	.363	.325
Matrix Reasoning	.582	.071	**.513**	.179	-.080	.374	.605	.453	.153
Visual Puzzles	.711	-.007	**.911**	-.056	-.010	.368	.877	.395	.307
Digit Span	.533	-.100	-.042	**.860**	.008	.299	.348	.793	.162
Arithmetic	.699	.134	.087	**.685**	.020	.507	.497	.798	.214
Symbol Search	.426	-.005	.090	-.051	**.852**	.122	.370	.171	.873
Coding	.348	.017	-.090	.067	**.781**	.113	.234	.195	.765
Eigenvalue		3.81	1.65	1.02	1.20				
Variance (%)		38.12	16.53	10.21	11.98				
Factor Correlations		F1	F2	F3	F4				
	F1	–							
	F2	.44	–						
	F3	.48	.50	–					
	F4	.13	.36	.21	–				

Note. Salient pattern coefficients (≥ .30) are denoted in bold. WAIS-IV = Weschler Adult Intelligence Scale, Fourth Edition.

[a] General structure coefficients are based on the first unrotated factor coefficients (*g* loadings).

Table 4.7. *Sources of Variance in the WAIS-IV for the Scattered Group (N = 616) According to an Orthogonalized Higher-Order Factor Model.*

	General		F1: VC		F2: PR		F3: WM		F4: PS			
Subtest	*b*	Var	*b*	Var	*b*	Var	*b*	Var	*b*	Var	h^2	u^2
Similarities	.485	.235	.521	.271							.507	.493
Vocabulary	.491	.241	.752	.526							.767	.233
Information	.519	.269	.535	.286							.556	.444
Block Design	.611	.373			.465	.216					.590	.410
Matrix Reasoning	.565	.319			.304	.092					.412	.588
Visual Puzzles	.682	.465			.540	.292					.757	.243
Digit Span	.565	.319					.552	.305			.624	.376
Arithmetic	.689	.475					.440	.194			.668	.332
Symbol Search	.359	.129							.786	.618	.747	.253
Coding	.291	.085							.720	.518	.603	.397
Total Variance (%)	29.1		10.8		6.0		5.0		11.4		62.3	37.7
Common Variance (%)	46.7		17.4		9.6		8.0		18.2		100	
$\omega_{h/s}$	.699		.482		.267		.301		.679			

Note. WAIS-IV = Weschler Adult Intelligence Scale, Fourth Edition; VC = Verbal Comprehension; PR = Perceptual Reasoning; WM = Working Memory; PS = Processing Speed; *b* = standardized loading of subtest on factor; Var = variance explained in the subtest; h^2 = communality; u^2 = uniqueness; ω_h = omega-hierarchical; ω_s = omega-subscale.

Table 4.8. *Descriptive Statistics for the Achievement Variables.*

	Flat Group (N = 595)				
Variable	*n*	*M*	*SD*	*SK*	*K*
Academic Fluency	572	94.22	13.15	-0.25	1.10
Basic Reading Skills	585	92.96	11.45	-0.65	1.26
Basic Writing Skills	271	97.79	11.45	-0.08	0.31
Math Calculation Skills	586	94.55	14.72	-0.32	0.57
	Scattered Group (N = 616)				
Variable	*n*	*M*	*SD*	*SK*	*K*
Academic Fluency	577	93.88	13.91	-0.01	-0.17
Basic Reading Skills	612	95.73	12.67	-0.68	0.71
Basic Writing Skills	285	98.59	14.33	-0.11	0.12
Math Calculation Skills	609	93.82	14.99	-0.23	-0.18

Note. Achievement variables derived from the Woodcock-Johnson III Tests of Achievement (Woodcock et al., 2001) and the Woodcock-Johnson IV Tests of Achievement (Schrank et al., 2014).

Table 4.9. *Incremental Contribution of Observed WAIS-IV Factor Scores in Predicting WJ-ACH Scores beyond the WAIS-IV Full-Scale IQ for the Flat Group and the Scattered Group.*

Flat Group (N = 595)	Academic Fluency (n = 572)			Basic Reading Skills (n = 585)			Basic Writing Skills (n = 271)			Math Calculation Skills (n = 586)		
	R^2	ΔR^2	Incr (%)[a]	R^2	ΔR^2	Incr (%)[a]	R^2	ΔR^2	Incr (%)[a]	R^2	ΔR^2	Incr (%)[a]
FSIQ	.45*	–	45.2	.32*	–	32.3	.26*	–	26.1	.47*	–	46.8
Index Scores[b]	.55	.09*	9.4	.40	.07*	7.2	.39	.13*	12.6	.52	.05*	5.2
VCI	.46	.00*	0.4	.34	.02*	1.6	.28	.01*	1.4	.47	.00	0.2
PRI	.45	.00	0.0	.33	.01*	0.5	.26	.00	0.3	.47	.00	0.2
WMI	.46	.01*	0.9	.34	.02*	2.0	.28	.02*	2.3	.48	.01*	1.2
PSI	.46	.01*	1.1	.33	.01*	0.8	.27	.01	0.8	.47	.00	0.3
Scattered Group (N = 616)	Academic Fluency (n = 577)			Basic Reading Skills (n = 612)			Basic Writing Skills (n = 285)			Math Calculation Skills (n = 609)		
	R^2	ΔR^2	Incr (%)[a]	R^2	ΔR^2	Incr (%)[a]	R^2	ΔR^2	Incr (%)[a]	R^2	ΔR^2	Incr (%)[a]
FSIQ	.35*	–	34.8	.32*	–	32.0	.29*	–	28.5	.37*	–	36.8
Index Scores[b]	.58	.23*	22.8	.46	.14*	13.8	.49	.21*	20.7	.53	.16*	16.2
VCI	.35	.00	0.3	.33	.01*	0.7	.30	.01*	1.3	.37	.01*	0.5
PRI	.35	.00	0.0	.32	.00	0.1	.29	.00	0.3	.37	.01*	0.5
WMI	.35	.01*	0.5	.33	.01*	0.9	.31	.02*	2.3	.39	.02*	2.4
PSI	.36	.01*	1.0	.32	.00	0.2	.29	.01*	0.9	.38	.01*	1.1

Note. WAIS-IV = Weschler Adult Intelligence Scale, Fourth Edition; WJ-ACH = Woodcock-Johnson Tests of Achievement; FSIQ = Full-Scale IQ; VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed Index.

[a] For increments (Incr), all unique contributions are squared part correlations equivalent to changes in R^2 if this variable was entered last in block entry regression procedure.

[b] Degrees of freedom = 4; partialing out FSIQ.

*$p < .05$

REFERENCES

Bartlett, M. S. (1954). A further note on the multiplying factors for various chi square approximations in factor analysis. *Journal of the Royal Statistical Society Series A (General)*, *16*, 296–298.

Beaujean, A. A., Firmin, M. W., Michonski, J. D., Berry, T., & Johnson, C. (2010). A multitrait–multimethod examination of the Reynolds Intellectual Assessment Scales in a college sample. *Assessment*, *17*(3), 347–360. https://doi.org/10.1177/1073191109356865

Canivez, G. L., McGill, R. J., Dombrowski, S. C., Watkins, M. W., Pritchard, A. E., & Jacobson, L. A. (2020). Construct validity of the WISC-V in clinical cases: Exploratory and confirmatory factor analyses of the 10 primary subtests. *Assessment*, *27*(2), 274–296. https://doi.org/10.1177/1073191118811609

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Frazier, T. W., Demaree, H. A., & Youngstrom, E. A. (2004). Meta-analysis of intellectual and neuropsychological test performance in attention-deficit/hyperactivity disorder. *Neuropsychology*, *18*(3), 543–555. https://doi.org/10.1037/0894-4105.18.3.543

Gignac, G. E. (2007). Multi-factor modeling in individual differences research: Some recommendations and suggestions. *Personality and Individual Differences*, *42*(1), 37–48. https://doi.org/10.1016/j.paid.2006.06.019

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Erlbaum.

Groth-Marnat, G., & Wright, A. J. (2016). *Handbook of psychological assessment* (6th ed.). Wiley.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook*. Guilford.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 31–36. https://doi.org/10.1007/BF02291575

Lassiter, K. S., Bell, N. L., Hutchinson, M. B., & Matthews, T. D. (2001). College student performance on the General Ability Measure for Adults and the Wechsler Intelligence Scale for Adults–Third Edition. *Psychology in the Schools*, *38*(1), 1–10. https://doi.org/10.1002/1520-6807(200101)38:1<1::AID-PITS1>3.0.CO;2-M

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). Oxford University Press.

Lichtenberger, E. O., Kaufman, A. S., & Kaufman, N. L. (2012). *Essentials of WAIS-IV assessment* (2nd ed.). Wiley.

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical manual: Woodcock-Johnson IV*. Riverside.

McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). *Technical manual: Woodcock-Johnson III Normative Update*. Riverside.

Miller, J. L., Saklofske, D. H., Weiss, L. G., Drozdick, L., Llorente, A. M., Holdnack, J. A., & Prifitera, A. (2016). Issues related to the WISC-V assessment of cognitive functioning in clinical and special groups. In L. G. Weiss, D. H. Saklofske, J. A. Holdnack, & A. Prifitera (Eds.), *WISC-V assessment and interpretation: Scientist-practitioner perspectives* (1st ed., pp. 287–343). Academic Press.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*(2), 129–140. https://doi.org/10.1080/00223891.2012.725437

Scheiber, C., Chen, H., Kaufman, A. S., & Weiss, L. G. (2017). How much does WAIS-IV perceptual reasoning decline across the 20 to 90-year lifespan when processing speed is controlled? *Applied Neuropsychology*: Adult, *24*(2), 116–131. https://doi.org/10.1080/23279095.2015.1107564

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*, 53–61. https://doi.org/10.1007/BF02289209

Schrank, F. A., Mather, N., & McGrew, K. S. (2014). *Woodcock-Johnson IV Tests of Achievement*. Riverside.

Swanson, H. L., & Hsieh, C. (2009). Reading disabilities in adults: A selective meta-analysis of the literature. *Review of Educational Research*, *79*(4), 1362–1390. https://doi.org/10.3102/0034654309350931

Watkins, M. W. (2010). Structure of the Wechsler Intelligence Scale for Children–Fourth Edition among a national sample of referred students. *Psychological Assessment*, *22*, 782–787. https://doi.org/10.1037/a0020043

Wechsler, D. (2008a). *Wechsler Adult Intelligence Scale* (4th ed.). Pearson.

Wechsler, D. (2008b). *Wechsler Adult Intelligence Scale–Fourth Edition technical and interpretive manual*. Pearson.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013). WISC-IV and clinical validation of the four- and five-factor interpretive approaches. *Journal of Psychoeducational Assessment*, *31*, 114–131. https://doi.org/10.1177/0734282913478032

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and over-extraction on principal axis factor analysis with varimax rotation. *Psychological Methods*, *1*, 254–265. https://doi.org/10. 1037/1082-989X.1.4.354

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Riverside.

CHAPTER 5

DISCUSSION

Within the intelligence testing literature, it is frequently suggested that significant variability among index scores, or scatter, renders the full-scale intelligence quotient (FSIQ) an invalid measure of overall ability (e.g., Groth-Marnat & Wright, 2016; Hale & Fiorello, 2004; Lezak et al., 2012). When this occurs, psychologists are encouraged to forego interpretation of the FSIQ and to focus their interpretation on the profile of index scores for diagnostic decision making (McGill, 2016). The existing empirical research, however, challenges the validity and utility of these methods.

Notably, previous studies have consistently found that the FSIQ upholds its structural and incremental validity in the presence of significant factor score variability (e.g., Ayoub, 2020; Daniel, 2007; Kotz et al., 2008; McGill, 2013, 2016; Watkins et al., 2007). Despite the overwhelming empirical evidence, many psychologists continue to forego interpretation of the FSIQ when an examinee is observed to have significant scatter (Kranzler et al., 2020; Pfeiffer et al., 2000). This practice also continues to be encouraged in many training programs (Decker et al., 2013; Lockwood & Farmer, 2020) and in technical manuals and interpretive guidebooks on cognitive assessment (e.g., Flanagan & Alfonso, 2017; Groth-Marnat & Wright, 2016; Miller et al., 2016; Weschler, 2008b) despite the lack of supporting empirical or theoretical evidence.

One plausible explanation for the continued prevalence of these interpretive techniques in clinical practice is the limited amount of relevant literature. Although previous studies have examined the impact of scatter on structural validity (e.g., Ayoub, 2020; Daniel, 2007), and

others have examined the incremental validity of scores in the presence of significant factor variability (e.g., Freberg et al., 2008; Kotz et al., 2008; McGill, 2013; Watkins et al., 2007), only one study to date (McGill, 2016) was designed specifically to investigate the potential impact of scatter on the structural validity *and* the incremental validity of cognitive measures. McGill's (2016) study, like many of the studies that examined the impact of scatter on psychometric validity (e.g., Ayoub, 2020; Freberg et al., 2008; McGill, 2013), used an archived standardization sample. Further research is needed to determine if previous findings generalize to clinical populations, as only two studies have examined impacts of scatter with a clinical sample (Freberg et al., 2008; Watkins et al., 2007). The existing research in this area also primarily focuses on child and adolescent populations, indicating a need for research with postsecondary and adult populations. Additionally, it is important to investigate the impact of scatter on the structural and incremental validity of other commonly used intelligence measures, such as the Weschler scales.

To address these gaps in the literature, the current study investigated the impact of significant factor score variability on the structural and incremental validity of the Weschler Adult Intelligence Scales, Fourth Edition (WAIS-IV; Weschler, 2008a) in a clinically referred sample of college students. Modeled after McGill's (2016) study, the current study sought to determine whether significant variability in the index scores on the WAIS-IV renders the FSIQ an invalid measure of overall ability and thus erroneous for clinical interpretation. Put simply, the present study aimed to answer the question that commonly arises for psychologists, "does scatter matter?"

Impact of Scatter on Structural Validity

Research question one in the present study examined whether significant factor score variability would impact the structural validity of the WAIS-IV FSIQ with a clinically referred sample. It was hypothesized that the FSIQ would maintain its hierarchical structure in the presence of significant index score variability. That is, it was predicted that the FSIQ would remain a valid estimation of global cognitive ability despite significant scatter. To investigate this research question, exploratory factor analysis (EFA) was used to examine the factor structure of the Flat Group and the Scattered Group. For both groups, the resulting EFA factor structure was then transformed with the Schmid-Leiman (SL) orthogonalization procedure in order to examine the proportions of WAIS-IV subtest variance attributed to the higher- and first-order factors.

Structural Validity of the Flat Group

The present study found that, when considering multiple EFA factor extraction criteria for the Flat Group, all criteria suggested extracting one or two factors. This is consistent with results from other studies that examined the WAIS-IV factor structure via EFA (e.g., Canivez & Watkins, 2010a, 2010b; Frazier & Youngstrom, 2007) and inconsistent from the confirmatory factor analysis (CFA) results reported in the WAIS-IV *Technical and Interpretative Manual* (Weschler, 2008b). Although factor extraction criteria suggested retention of one or two factors, four factors were extracted in accordance with the theoretical structure of the WAIS-IV. In doing so, however, the factor pattern did not align with the WAIS-IV theoretical structure, as three subtests (Information, Arithmetic, Matrix Reasoning) were not properly associated with their theoretically proposed first-order factor.

Results from the present study demonstrated highly correlated first-order factors for the Flat Group, suggesting the presence of a higher-order general factor. Given Carroll's (1993) assertion that cognitive measures are composed of reliable variance that is attributable to both the higher-order general factor and the first-order factors, the present study employed additional procedures (e.g., Carroll, 1995; Schmid & Leiman, 1957) to apportion subtest variance appropriately to higher- and first-order factors.

The application of the SL transformation to the Flat Group's EFA factor structure revealed that the higher-order *g* factor accounted for far more total and common variance than any of the four individual first-order factors. In fact, the general factor (FSIQ) accounted for more total and common variance than that explained by the four first-order factors combined. None of the first-order factors accounted for meaningful common variance. The Verbal Comprehension factor accounted for the most additional variance beyond the general factor, explaining only 3.2% and 5.5% of the total and common variance, respectively. These findings are consistent with those found by Canivez and Watkins (2010a, 2010b) and Nelson et al. (2013), although a direct comparison cannot be made because their samples were not limited to individuals without significant scatter. Likewise, the present results are similar to those obtained from investigations of other intelligence tests, including studies of the Wechsler Intelligence Scale for Children (Bodin et al., 2009; Watkins, 2006, 2010; Watkins et al., 2006), Reynolds Intellectual Assessment Scales (Dombrowski et al., 2009; Nelson & Canivez, 2012; Nelson et al., 2007), Stanford-Binet Intelligence Scales (Canivez, 2008; DiStefano & Dombrowski, 2006), and Woodcock-Johnson Tests of Cognitive Abilities (Dombrowski et al., 2017; Strickland et al., 2015).

Moreover, model-based reliabilities of the WAIS-IV factors for the Flat Group were estimated with omega coefficients (Reise, 2012; Reise et al., 2013; Rodriguez et al., 2016). Results revealed a high omega coefficient for the general factor, indicating that the WAIS-IV FSIQ score is sufficient for interpretation. On the other hand, coefficients for the four group factors were all low, suggesting that the WAIS-IV index scores possess little unique variance apart from *g.* These findings reveal that, for individuals with a "flat" cognitive profile, the index factor scores appear to have too little unique variance beyond the influence of general intelligence to support confident clinical interpretation (Gignac & Watkins, 2013; Reise et al., 2013; Rodriguez et al., 2016). On the whole, in the absence of significant scatter, the FSIQ score provides the most psychometrically sound information regarding individual cognitive performance.

Structural Validity of the Scattered Group

To examine the impact of significant factor score variability on the structural validity of the WAIS-IV FSIQ, the aforementioned hierarchical EFA procedures were replicated with the Scattered Group in the present study. EFA results using multiple factor extraction criteria supported the extraction of four factors in accordance with the theoretical structure of the WAIS-IV. All of the WAIS-IV core subtests were properly associated with their theoretically proposed first-order group factors. This finding is consistent with CFA results reported in the WAIS-IV *Technical and Interpretative Manual* (Weschler, 2008b), as well as results from a study that used CFA to examine the factor structure of the WAIS-IV with a clinically referred sample (Nelson et al., 2013).

Given the moderate correlations among first-order factors for the Scattered Group, the SL orthogonalization procedure (Schmid & Leiman, 1957) was employed to apportion subtest

variance appropriately to higher- and first-order factors, as recommended by Carroll (1995). Following SL transformation to the Scattered Group's EFA factor structure, the higher-order *g* factor accounted for substantially greater total and common variance than any of the four individual first-order factors. In fact, the general factor (FSIQ) accounted for more than twice the amount of the total variance of any single index factor. Whereas the combination of *g* and uniqueness outweighed the contributions made by the four first-order index factors, meaningful common variance was accounted for by the Verbal Comprehension (17.4%) and Processing Speed (18.2%) factors. This finding suggests that additional consideration of Verbal Comprehension and Processing Speed may provide useful information as it relates to individual performance beyond the FSIQ when significant levels of scatter is observed. Similar results were found in an investigation of the WAIS-IV factor structure with a clinically referred sample (Nelson et al., 2013), as well as results from studies examining the factor structure of other intelligence tests (e.g., Canivez et al., 2019; Strickland et al., 2015; Styck & Watkins, 2016, 2017).

Moreover, as done with the Flat Group, omega coefficients were used to estimate model-based reliabilities of the WAIS-IV factors for the Scattered Group (Reise, 2012; Reise et al., 2013; Rodriguez et al., 2016). Results revealed a high coefficient for the general factor, indicating that the WAIS-IV FSIQ score is sufficient for interpretation. On the contrary, coefficients were low and suggested minimal unique variance for all but the Processing Speed group factor. This is consistent with results from similar investigations of the structural validity of cognitive tests (e.g., Nelson et al., 2013; Strickland et al., 2015). Taken together, the present findings reveal that, for individuals with a "scattered" cognitive profile, the index factor scores (aside from Processing Speed) appear to have too little unique variance beyond the influence of

general intelligence to support confident clinical interpretation (Gignac & Watkins, 2013; Reise et al., 2013; Rodriguez et al., 2016). Even in the presence of significant scatter, the FSIQ score provides the most psychometrically sound information regarding individual cognitive performance in comparison to index scores.

Summary: Impact of Scatter on Structural Validity

Results of the present study established that, even in the presence of significant scatter, the hierarchical structure of the WAIS-IV was not altered with a clinically referred sample, findings consistent with those obtained by McGill (2016) in a similar study using the Kaufman Assessment Battery for Children standardization sample (Kaufman & Kaufman, 2004). Whereas general factor attenuation was evident, especially regarding its effects on the Processing Speed factor, the total and common variance accounted for by the FSIQ was consistently significantly greater. These findings suggest that invalidation of the general index score in the presence of significant scatter does not preclude its influence on first-order factor scores, and thus that the WAIS-IV FSIQ should not automatically be disregarded in these conditions.

Comparison between the Flat and Scattered Groups in the present study further supports the four-factor theoretical model suggested by the WAIS-IV test publisher. Namely, forcing the theoretical model to the Flat Group sample resulted in theoretically inconsistent subtest migration and inadmissible factors, yet desired simple structure was obtained for the Scattered Group. This finding is not surprising given that significant scatter is prevalent in the WAIS-IV normative sample (Grégoire et al., 2011).

The present results are consistent with our hypothesis that the WAIS-IV FSIQ would maintain its structural validity in the presence of significant index score variability. In other words, findings suggest that the FSIQ remains a valid estimation of global cognitive ability

despite significant scatter. Moreover, in addition to structural validity, examining incremental validity is particularly informative for determining the appropriate interpretation of WAIS-IV scores in the presence of significant factor score variability (McGill, 2016).

Impact of Scatter on Incremental Validity

To further investigate the influence of significant factor score variability on the interpretability of cognitive test scores, research question two in the present study examined whether significant scatter would impact the incremental validity of the WAIS-IV FSIQ with a clinically referred sample. In line with previous findings (e.g., Freberg et al., 2008; Kotz et al., 2008; McGill, 2013, 2016; Watkins et al., 2007), it was hypothesized that the FSIQ would maintain its incremental validity in the presence of significant index score variability. That is, it was predicted that the FSIQ would remain a valid predictor of academic achievement despite significant scatter. To investigate this research question, hierarchical multiple regression analysis was used to examine the incremental contributions of the WAIS-IV FSIQ and index scores on academic achievement. The Academic Fluency, Basic Reading Skills, Basic Writing Skills, and Math Calculation Skills composites from the Woodcock-Johnson Tests of Achievement (WJ-ACH; Schrank et al., 2014; Woodcock et al., 2001) were used as achievement variables in the current study.

Incremental Validity of the Flat Group

Hierarchical multiple regression results for the Flat Group indicated that the WAIS-IV FSIQ accounted for large and statistically significant portions of the four WJ-ACH scores. This finding is consistent with results from previous investigations of the incremental validity of the WAIS-IV (e.g., Canivez, 2013; Nelson et al., 2013). Of the specific WJ-ACH academic composite scores for the Flat Group, the Math Calculation Skills composite was best predicted

by the FSIQ, followed by Academic Fluency, Basic Reading Skills, and lastly, Basic Writing Skills.

Incremental prediction of the four WAIS-IV index scores combined (after controlling for the effects of the FSIQ) was statistically significant for each of the four academic achievement composite scores. However, effect sizes were small for all academic composites, with the exception of a moderate effect in predicting Basic Writing Skills beyond the FSIQ score. Additionally, results indicated that the unique incremental contributions of the four individual WAIS-IV index scores in predicting academic achievement were all trivial. That is, none of the WAIS-IV index scores demonstrated meaningful predictive achievement variance (for any of the four WJ-ACH scores) beyond that accounted for by the FSIQ score. This finding is consistent with previous research on the incremental validity of the WAIS-IV (e.g., Canivez, 2013; Nelson et al., 2013), as well as with investigations of other intelligence tests (e.g., Glutting et al., 2006; McGill, 2015, 2016; McGill & Busse, 2015).

In summary, results from the present study revealed that, for individuals with a "flat" cognitive profile, the WAIS-IV FSIQ consistently accounted for the greatest portions of achievement variance across the four WJ-ACH composites, over and above that accounted for by the WAIS-IV index scores. These findings suggest that the FSIQ score should be given the greatest interpretive weight in the absence of significant scatter.

Incremental Validity of the Scattered Group

To investigate the impact of significant factor score variability on the incremental validity of the WAIS-IV FSIQ, the hierarchical multiple regression analyses were replicated with the Scattered Group in the present study. This enabled us to examine the extent to which WAIS-IV factor scores provided meaningful improvements in prediction of WJ-ACH scores beyond that

already accounted for by the FSIQ. Results revealed that the FSIQ accounted for large and statistically significant portions of the four WJ-ACH composite scores. Of the specific WJ-ACH scores for the Scattered Group, the Math Calculation Skills composite was best predicted by the FSIQ, followed by Academic Fluency, Basic Reading Skills, and lastly, Basic Writing Skills.

Incremental prediction of the four index scores combined (after controlling for the effects of the FSIQ) was statistically significant for all academic achievement composite scores, with effect sizes falling in the moderate range for each of the four WJ-ACH scores. Additionally, results indicated that the unique incremental contributions of the four individual WAIS-IV index scores in predicting academic achievement were all trivial. That is, none of the WAIS-IV index scores demonstrated meaningful predictive achievement variance (for any of the four WJ-ACH scores) beyond that accounted for by the FSIQ score.

The current study was the first to investigate the impact of significant index score variability on the predictive validity of the WAIS-IV FSIQ with a clinically referred sample. Consistent with results from studies using other intelligence tests (e.g., Freberg et al., 2008; Kotz et al., 2008; McGill, 2013, 2016; Watkins et al., 2007), present findings indicated that the WAIS-IV FSIQ accounted for the greatest portions of achievement variance across the four WJ-ACH composites, over and above that accounted for by the WAIS-IV index scores. Whereas the results suggest that the four index scores combined had a moderate effect in predicting achievement scores beyond the FSIQ score, the FSIQ individually accounted for the largest amount of variance for all achievement indicators on the WJ-ACH. Taken together, current findings suggest that significant factor score variability does not undermine the incremental validity of the WAIS-IV FSIQ for clinically referred individuals. These results are consistent with existing literature indicating that the global composite remains a valid predictor of academic

achievement despite significant scatter (e.g., Freberg et al., 2008; Kotz et al., 2008; McGill, 2013, 2016; Watkins et al., 2007).

Summary: Impact of Scatter on Incremental Validity

The present findings revealed that, even in the presence of significant scatter, the WAIS-IV FSIQ is a valid predictor of academic achievement on the WJ-ACH for clinically referred individuals. Notably, results indicated that the FSIQ accounted for the largest amount of variance across achievement indicators on the WJ-ACH. The FSIQ also consistently accounted for greater portions of achievement variance than that accounted for by the WAIS-IV index factor scores. This was true for both the Flat and Scattered Groups in the current study. These findings suggest that the FSIQ should be afforded the greatest interpretive weight because it was consistently a stronger predictor of academic achievement than the four factor indices, both individually and combined (McGill, 2015; Nelson et al., 2013).

Comparison between the Flat and Scattered Groups in the present study further support the argument that examiners should not forego interpreting the FSIQ in the presence of significant factor score variability. There were no significant differences in results between the two groups, therefore suggesting that significant scatter does not weaken the predictive validity of the FSIQ. This is consistent with findings from previous incremental validity research that compared low and high scatter groups (e.g., Freberg et al., 2008; Ryan et al., 2002; Watkins et al., 2007).

The present results support the hypothesis that the WAIS-IV FSIQ maintains its incremental validity in the presence of significant factor score variability. These results do not support the common practice and recommendation in the WAIS-IV *Technical and Interpretative Manual* (Weschler, 2008b) to forego interpretation of the FSIQ in the presence of significant

index score variability. Instead, the present findings add to the empirical literature suggesting that the global ability composite remains a strong and valid predictor of academic achievement despite significant scatter.

Limitations and Future Research

The current study is not without limitations that should be considered when interpreting the results. First, several characteristics of the sample limit generalizability of these results. The sample consisted primarily of young adults with various learning or attention disorders who planned on attending, were attending, or had attended college. Although the age range was broad, most individuals were of typical college age, and therefore, age groups near the high end of the range were not well represented. Another characteristic that limits the generalizability of these findings is that ethnicity was not representative of the U.S. population at large. Although the current sample was a relatively large, clinically referred sample, additional research is needed to determine if these results generalize to other clinical populations such as those with intellectual disabilities and other related neurocognitive impairments. Future research should also examine the generalizability of the current results to other cognitive and academic achievement measurement instruments.

Second, a non-referred comparison group was not included in the current study. Inclusion of a comparison group would have enabled further examination of the factor structure of the WAIS-IV and the incremental validity of the index scores. That is, inclusion of both referred and non-referred groups would have allowed for examination of underlying factor structure, as well as investigation of incremental validity across groups (Bowden et al., 2008). Similarly, the sample in the current study was comprised of a range of disability types, with a quarter of individuals having multiple disabilities and a small portion without a disability. It is possible that

results may have differed across clinical groups (e.g., attention-deficit/hyperactivity disorder, specific learning disability), though previous investigations have found little evidence to support the diagnostic utility of cognitive score variability (McGill, 2018; Ryan et al., 2006; Watkins, 1999, 2005). Nonetheless, future research should investigate differential impacts of significant scatter on the validity of cognitive measures across disability groups (Styck & Watkins, 2016, 2017) and other variables (e.g., FSIQ score; McLean et al., 1989).

Third, four factors were extracted for both groups even though EFA extraction criteria did not fully support a four-factor model for the Flat Group. Forcing the theoretical model to the Flat Group resulted in cross-loading and theoretically inconsistent subtest migration, which presents a methodological and interpretive confound (McGill & Dombrowski, 2018; McGill & Spurgin, 2017). As such, EFA results and the resulting omega coefficients for the Flat Group should be interpreted with caution.

Lastly, future research should employ CFA techniques to investigate the tenability of other measurement models (e.g., correlated factors, bifactor) and the invariance of measurement models across groups with and without significant levels of scatter for the WAIS-IV and other related cognitive instruments (McGill, 2016). Such examinations would enable researchers to consider the convergence or divergence of CFA and EFA results and the potential implications for clinical interpretation. As indicated by Gorsuch (1983), greater confidence can be placed in the internal structure of a test when EFA and CFA results are consistent.

Implications for Clinical Practice

The present findings have implications for interpretation of cognitive assessment scores in the presence of significant scatter within clinical practice. Namely, results of the present study indicate that the global ability score on the WAIS-IV maintains its structural and incremental

validity in the presence of significant index score variability. These findings suggest that the FSIQ remains a valid estimation of global cognitive ability and a valid predictor of academic achievement despite significant scatter. Further, the present results reveal that invalidation of the global composite on the WAIS-IV in the presence of significant scatter does not remove its influence on first-order factor and subtest scores and thus should not automatically be disregarded in those circumstances to adhere to a popular 'rule of thumb' (McGill, 2016).

Despite the lack of empirical support for the variability hypothesis, the interpretive procedure continues to be promoted by assessment "experts" (e.g., Groth-Marnat & Wright, 2016; Miller et al., 2016; Sattler, 2018) and employed in training and clinical practice (Farmer et al., 2021; Kranzler et al., 2020; Lockwood & Farmer, 2020; Sotelo-Dynega & Dixon, 2014). These interpretive techniques, which act in direct contrast to evidence-based assessment practices set forth by professional standards (American Psychological Association [APA], 2006, 2017; National Association of School Psychologists [NASP], 2020), are referred to as *low-value practices* (LVPs; Farmer et al., 2021). LVPs are defined as those that: (a) have limited evidence for their clinical utility, (b) are not the most effective available practice, (c) have unacceptable risk of adverse effects, or (d) are diagnostically or therapeutically unnecessary (McKay et al., 2018).

Given what we know about LVPs, why do psychologists continue to engage in low-value assessment practices such as discounting the global composite score in the presence of significant scatter? Scholars have suggested that these procedures may endure because of the *ad antiquitatem fallacy*, the belief that a practice that has been used for a long time must be legitimate (Lilienfeld et al., 2012), or the *ad populum fallacy*, the belief that a technique that is widely used must be valid or effective (Lilienfeld et al., 2006). In other words, the variability

hypothesis and other LVPs may persist because psychologists simply abide by the procedures that have been recommended in assessment guidebooks for decades, taught in their graduate training programs, and used by other psychologists around them (Farmer et al., 2021; Kranzler et al., 2020; Lockwood & Farmer, 2020).

Although there is likely no single answer to our question, it is important that psychologists are cognizant of the "warning signs" that a particular assessment practice may be of low value (Farmer et al., 2022; Meichenbaum & Lilienfeld, 2018). First and foremost, it is imperative that all psychologists possess the requisite skills to differentiate between evidence-based practices and non-evidence-based practices (Lilienfeld et al., 2012). Current professional standards emphasize evidence-based practice (APA, 2006, 2017; NASP, 2020), highlighting the importance of using scientific findings to inform service delivery. Additionally, graduate training programs play a vital role in promoting evidence-based practice and shaping professional service delivery among future psychologists (Benson et al., 2020; Sotelo-Dynega & Dixon, 2014). Given this crucial role, professors who teach cognitive assessment courses should carefully consider whether instructional materials are responsive to empirical research and ensure that course content incorporates peer-reviewed literature to protect against long-standing LVPs such as the variability hypothesis.

Conclusion

Results from the current study coupled with previous research provide evidence that psychologists should reconsider the practice of automatically disregarding the FSIQ when an examinee presents with significant scatter. The present findings strongly suggest that the FSIQ is a valid estimation of global cognitive ability and a valid predictor of academic achievement regardless of significant scatter. Thus, until there is sufficient empirical evidence to suggest

otherwise, "psychologists must be willing to risk casting aside practices that have long promised clinical gold, but for decades have failed to deliver meaningful outcomes" (Dombrowski et al., 2022, p. 10). Accordingly, psychologists are encouraged to abide by the notion that *scatter does not matter*.

REFERENCES

American Psychological Association. (2006). Evidence-based practice in psychology. *American Psychologist*, *61*(4), 271–285. https://doi.org/10.1037/0003-066X.61.4.271

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct*. http://www.apa.org/ethics/code/principles.pdf

Ayoub, J. L. (2020). Does cognitive scatter attenuate the validity of the global intelligence score? (Publication No. 28091905) [Doctoral dissertation, Hofstra University]. ProQuest Dissertations & Theses Global.

Benson, N. F., Maki, K. E., Floyd, R. G., Eckert, T. L., Kranzler, J. H., & Fefer, S. A. (2020). A national survey of school psychologists' practices in identifying specific learning disabilities. *School Psychology*, *35*(2), 146–157. https://doi.org/10.1037/spq0000344

Bodin, D., Pardini, D. A., Burns, T. G., & Stevens, A. B. (2009). Higher order factor structure of the WISC-IV in a clinical neuropsychological sample. *Child Neuropsychology*, *15*, 417–424.

Bowden, S. C., Gregg, N., Bandalos, D., Davis, M., Coleman, C., Holdnack, J. A., & Weiss, L. G. (2008). Latent mean and covariance differences with measurement equivalence in college students with developmental difficulties versus the Wechsler Adult Intelligence Scale–III/Wechsler Memory Scale–III normative sample. *Educational and Psychological Measurement*, *68*, 621–642. https://doi.org/10.1177/0013164407310126

Canivez, G. L. (2008). Orthogonal higher-order factor structure of the Stanford-Binet Intelligence Scales—Fifth Edition for children and adolescents. *School Psychology Quarterly*, *23*, 533–541. https://doi.org/10.1037/a0012884

Canivez, G. L. (2013). Incremental criterion validity of WAIS-IV factor index scores: Relationships with WIAT-II and WIAT-III subtest and composite scores. *Psychological Assessment*, *25*, 484–495. https://doi.org/10.1037/a0032092

Canivez, G. L., & Watkins, M. W. (2010a). Exploratory and higher-order factor analyses of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) adolescent subsample. *School Psychology Quarterly*, *25*, 223–235. https://doi.org/10.1037/a0022046

Canivez, G. L., & Watkins, M. W. (2010b). Investigation of the factor structure of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV): Exploratory and higher order factor analyses. *Psychological Assessment*, *22*, 827–836. https://doi.org/10.1037/a0020429

Canivez, G. L., Watkins, M. W., & McGill, R. J. (2019). Construct validity of the Wechsler Intelligence Scale for Children–Fifth UK Edition: Exploratory and confirmatory factor analyses of the 16 primary and secondary subtests. *British Journal of Educational Psychology*, *89*, 195–224. https://doi.org/10.1111/bjep.12230

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

Carroll, J. B. (1995). On methodology in the study of cognitive abilities. *Multivariate Behavioral Research*, *30*, 429-452. https://doi.org/10.1207/s15327906mbr3003_6

Daniel M. H. (2007). "Scatter" and the construct validity of FSIQ: Comment on Fiorello et al. (2007). *Applied Neuropsychology*, *14*(4), 291–295. https://doi.org/10.1080/09084280701719401

Decker, S. L., Hale, J. B., & Flanagan, D. P. (2013). Professional practice issues in the assessment of cognitive functioning for educational applications. *Psychology in the Schools*, *50*, 300–313. https://doi.org/10.1002/pits.21675

DiStefano, C., & Dombrowski, S. C. (2006). Investigating the theoretical structure of the Stanford-Binet, Fifth Edition. *Journal of Psychoeducational Assessment*, *24*, 123–136. https://doi.org/10.1177/0734282905285244

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2017). Exploratory and hierarchical factor analysis of the WJ-IV cognitive at school age. *Psychological Assessment*, *29*(4), 394–407. https://doi.org/10.1037/pas0000350

Dombrowski, S. C., McGill, R. J., Watkins, M. W., Canivez, G. L., Pritchard, A. E., & Jacobson, L. A. (2022). Will the real theoretical structure of the WISC-V please stand up? Implications for clinical interpretation. *Contemporary School Psychology*, *26*, 492–503. https://doi.org/10.1007/s40688-021-00365-6

Dombrowski, S. C., Watkins, M. W., & Brogan, M. J. (2009). An exploratory investigation of factor structure of the Reynolds Intellectual Assessment Scales (RIAS). *Journal of Psychoeducational Assessment*, *27*, 279–286. https://doi.org/10.1177/0734282909333179

Farmer, R. L., McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2021). Why questionable assessment practices remain popular in school psychology: Instructional materials as pedagogic vehicles. *Canadian Journal of School Psychology*, *36*(2), 98–114. https://doi.org/10.1177/0829573520978111

Farmer, R. L., McGill, R. J., Lockwood, A. B., Dombrowski, S. C., Canivez, G. L., & Zaheer, I. (2022). Warning signs for hype in school-based assessment: Implications for training and pedagogy. *School Psychology Training and Pedagogy*, *39*(1), 11–24. https://doi.org/10.31234/osf.io/ypcv7

Fiorello, C. A., Hale, J. B., Holdnack, J. A., Kavanagh, J. A., Terrell, J., & Long, L. (2007). Interpreting intelligence test results for children with disabilities: Is global intelligence relevant? *Applied Neuropsychology*, *21*, 2–12. https://doi.org/10.1080/09084280701280338

Flanagan, D. P., & Alfonso, V. C. (2017). *Essentials of WISC-V assessment*. Wiley.

Frazier, T. W., & Youngstrom, E. A. (2007). Historical increase in the number of factors measured by commercial tests of cognitive ability: Are we overfactoring? *Intelligence*, *35*, 169–182. https://doi.org/10.1016/j.intell.2006.07.002

Freberg, M. E., Vandiver, B. J., Watkins, M. W., & Canivez, G. L. (2008). Significant factor score variability and the validity of the WISC-III Full Scale IQ in predicting later academic achievement. *Applied neuropsychology*, *15*(2), 131–139. https://doi.org/10.1080/09084280802084010

Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research*, *48*, 639–662. https://doi.org/10.1080/00273171.2013.804398

Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. *The Journal of Special Education*, *40*, 103–114. https://doi.org/10.1177/00224669060400020101

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Erlbaum.

Grégoire, J., Coalson, D. L., & Jianjun, Z. (2011). Analysis of WAIS-IV index score scatter using significant deviation from the mean index score. *Assessment*, *18*(2), 168–177. https://doi.org/10.1177/1073191110386343

Groth-Marnat, G., & Wright, A. J. (2016). *Handbook of psychological assessment* (6th ed.). Wiley.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook*. Guilford.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children* (2nd ed.). American Guidance Service.

Kotz, K. M., Watkins, M. W., & McDermott, P. A. (2008). Validity of the General Conceptual Ability score from the Differential Ability Scales as a function of significant and rare interfactor variability. *School Psychology Review*, *37*(2), 261–278.

Kranzler, J. H., Maki, K. E., Benson, N. F., Eckert, T. L., Floyd, R. G., & Fefer, S. A. (2020). How do school psychologists interpret intelligence tests for the identification of specific learning disabilities? *Contemporary School Psychology*. https://doi.org/10.1007/s40688-020-00274-0

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). Oxford University Press.

Lilienfeld, S. O., Ammirati, R., & David, M. (2012). Distinguishing between science from pseudoscience in school psychology: Science and scientific thinking as safeguards against human error. *Journal of School Psychology*, *50*, 7–36. https://doi.org/10.1016/j.jsp.2011.09.006

Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2006). Why questionable psychological tests remain popular. *Scientific Review of Alternative Medicine*, *10*, 6–15.

Lockwood, A. B., & Farmer, R. L. (2020). The cognitive assessment course: Two decades later. *Psychology in the Schools*, *57*(2), 265–283. https://doi.org/10.1002/pits.22298

McGill, R. J. (2013). Beyond g: Assessing the incremental validity of the Cattell-Horn-Carroll (CHC) broad ability factors on the Woodcock-Johnson III Tests of Cognitive Abilities (Publication No. 3621595) [Doctoral dissertation, Chapman University]. ProQuest Dissertations & Theses Global.

McGill, R. J. (2015). Interpretation of KABC-II scores: An evaluation of the incremental validity of Cattell-Horn-Carroll (CHC) factor scores in predicting achievement. *Psychological Assessment*, *27*(4), 1417–1426. https://doi.org/10.1037/pas0000127

McGill, R. J. (2016). Invalidating the full scale IQ score in the presence of significant factor score variability: Clinical acumen or clinical illusion? *Archives of Assessment Psychology*, *6*(1), 49–79.

McGill, R. J. (2018). Confronting the base rate problem: More ups and downs for cognitive scatter analysis. *Contemporary School Psychology*, *22*, 384–393.

https://doi.org/10.1007/s40688-017-0168-4

McGill, R. J., & Busse, R. T. (2015). Incremental validity of the WJ III COG: Limited predictive effects beyond the GIA-E. *School Psychology Quarterly*, *30*(3), 353–365. https://doi.org/10.1037/spq0000094

McGill, R. J., & Dombrowski, S. C. (2018). Factor structure of the CHC model for the KABC II: Exploratory factor analyses with the 16 core and supplemental subtests. *Contemporary School Psychology*, *22*, 279–293. https://doi.org/10.1007/s40688-017-0152-z

McGill, R. J., & Spurgin, A. R. (2017). Exploratory higher order analysis of the Luria interpretive model on the Kaufman Assessment Battery for Children-Second Edition (KABC-II) school-age battery. *Assessment*, *24*(4), 540–552. https://doi.org/10.1177/1073191115614081

McKay, V. R., Morshed, A. B., Brownson, R. C., Proctor, E. K., & Prusaczyk, B. (2018). Letting go: Conceptualizing intervention de-implementation in public health and social service settings. *American Journal of Community Psychology*, *62*(1–2), 189–202. https://doi.org/10.1002/ajcp.12258

McLean, J. E., Kaufman, A. S., & Reynolds, C. R. (1989). Subtest scatter on the WAIS-R. *Journal of Clinical Psychology*, *45*, 919–926.

Meichenbaum, D., & Lilienfeld, S. O. (2018). How to spot hype in the field of psychotherapy: A 19-item checklist. *Professional Psychology: Research and Practice*, *49*(1), 22–30. https://doi.org/10.1037/pro0000172

Miller, J. L., Saklofske, D. H., Weiss, L. G., Drozdick, L., Llorente, A. M., Holdnack, J. A., & Prifitera, A. (2016). Issues related to the WISC-V assessment of cognitive functioning in clinical and special groups. In L. G. Weiss, D. H. Saklofske, J. A. Holdnack, & A. Prifitera (Eds.), *WISC-V assessment and interpretation: Scientist-practitioner perspectives* (1st ed., pp. 287–343). Academic Press.

National Association of School Psychologists. (2020). *The professional standards of the National Association of School Psychologists*. https://www.nasponline.org/standards-and-certification/nasp-practice-model/about-the-nasp-practice-model

Nelson, J. M., & Canivez, G. L. (2012). Examination of the structural, convergent, and incremental validity of the Reynolds Intellectual Assessment Scales (RIAS) with a

clinical sample. *Psychological Assessment*, *24*(1), 129–140. https://doi.org/10.1037/a0024878

Nelson, J. M., Canivez, G. L., Lindstrom, W., & Hatt, C. (2007). Higher-order exploratory factor analysis of the Reynolds Intellectual Assessment Scales with a referred sample. *Journal of School Psychology*, *45*, 439–456. https://doi.org/10.1016/j.jsp.2007.03.003

Nelson, J. M., Canivez, G. L., & Watkins, M. W. (2013). Structural and incremental validity of the Wechsler Adult Intelligence Scale–Fourth Edition with a clinical sample. *Psychological Assessment*, *25*(2), 618–630. https://doi.org/10.1037/a0032086

Pfeiffer, S. I., Reddy, L. A., Kletzel, J. E., Schmelzer, E. R., & Boyer, L. M. (2000). The practitioner's view of IQ testing and profile analysis. *School Psychology Quarterly*, *15*, 376–385. https://doi.org/10.1037/ h0088795

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*(2), 129–140. https://doi.org/10.1080/00223891.2012.725437

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150. https://doi.org/10.1037/met0000045

Ryan, J. J., Kreiner, D. S., & Burton, D. B. (2002). Does high scatter affect the predictive validity of WAIS–III IQs? *Applied Neuropsychology*, *9*(3), 173–178.

Ryan, J. J., Tree, H. A., Morris, J., & Gontkovsky, S. T. (2006). Wechsler Adult Intelligence Scale-III inter-subtest scatter: A comparison of brain-damaged patients and normal controls. *Journal of Clinical Psychology*, *62*(10), 1319–1326. https://doi.org/10.1002/jclp.20307

Sattler, J. M. (2018). *Assessment of children: Cognitive foundations and applications* (6th ed.). Sattler Publisher.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*, 53–61. https://doi.org/10.1007/BF02289209

Schrank, F. A., Mather, N., & McGrew, K. S. (2014). *Woodcock-Johnson IV Tests of Achievement*. Riverside.

Sotelo-Dynega, M., & Dixon, S. G. (2014). Cognitive assessment practices: A survey of school psychologists. *Psychology in the Schools*, *51*(10), 1031–1045. https://doi.org/10.1002/pits.21802

Strickland, T., Watkins, M. W., & Caterino, L. C. (2015). Structure of the Woodcock–Johnson III cognitive tests in a referral sample of elementary school students. *Psychological Assessment*, *27*, 689–697. https://doi.org/10.1037/pas0000052

Styck, K. M., & Watkins, M. W. (2016). Structural validity of the WISC-IV for students with learning disabilities. *Journal of Learning Disabilities*, *49*(2), 216–224. https://doi.org/10.1177/0022219414539565

Styck, K. M., & Watkins, M. W. (2017). Structural validity of the WISC-IV for students with ADHD. *Journal of Attention Disorders*, *21*(11), 921–928. https://doi.org/10.1177/1087054714553052

Watkins, M. W. (1999). Diagnostic utility of WISC-III subtest variability among students with learning disabilities. *Canadian Journal of School Psychology*, *15*(1), 11–20.

Watkins, M. W. (2005). Diagnostic validity of Wechsler subtest scatter. *Learning Disabilities: A Contemporary Journal*, *3*, 20–29.

Watkins, M. W. (2006). Orthogonal higher order structure of the Wechsler Intelligence Scale for Children–Fourth Edition. *Psychological Assessment*, *18*(1), 123–125. https://doi.org/10.1037/1040-3590.18.1.123

Watkins, M. W. (2010). Structure of the Wechsler Intelligence Scale for Children–Fourth Edition among a national sample of referred students. *Psychological Assessment*, *22*, 782–787. https://doi.org/10.1037/a0020043

Watkins, M. W., Glutting, J. J., & Lei, P.-W. (2007). Validity of the Full-Scale IQ when there is significant variability among WISC-III and WISC-IV factor scores. *Applied Neuropsychology*, *14*(1), 13–20. https://doi.org/10.1080/09084280701280353

Watkins, M. W., Wilson, S. M., Kotz, K. M., Carbone, M. C., & Babula, T. (2006). Factor structure of the Wechsler Intelligence Scale for Children–Fourth Edition among referred students. *Educational and Psychological Measurement*, *66*, 975–983. https://doi.org/10.1177/0013164406288168

Wechsler, D. (2008a). *Wechsler Adult Intelligence Scale* (4th ed.). Pearson.

Wechsler, D. (2008b). *Wechsler Adult Intelligence Scale–Fourth Edition technical and interpretive manual*. Pearson.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Riverside.