

Illinois State University ISU ReD: Research and eData

Theses and Dissertations

9-12-2016

Optimizing Cross-Battery Assessment Procedures For Reading-Based Specific Learning Disorder Diagnosis: A Monte Carlo Study

Zachary Joseph Roman
Illinois State University, zjr159811@gmail.com

Follow this and additional works at: <https://ir.library.illinoisstate.edu/etd>

 Part of the [Educational Psychology Commons](#), and the [Quantitative Psychology Commons](#)

Recommended Citation

Roman, Zachary Joseph, "Optimizing Cross-Battery Assessment Procedures For Reading-Based Specific Learning Disorder Diagnosis: A Monte Carlo Study" (2016). *Theses and Dissertations*. 630.
<https://ir.library.illinoisstate.edu/etd/630>

This Thesis and Dissertation is brought to you for free and open access by ISU ReD: Research and eData. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of ISU ReD: Research and eData. For more information, please contact ISUREd@ilstu.edu.

OPTIMIZING CROSS-BATTERY ASSESSMENT PROCEDURES FOR
READING-BASED SPECIFIC LEARNING DISORDER DIAGNOSIS:
A MONTE CARLO STUDY

Zachary Joseph Roman

75 Pages

Specific Learning Disorder is a disorder in which deficits in academic skills are caused by specific cognitive deficits. Cross-Battery Assessment (Flanagan, Alfonso, & Ortiz, 2013), is a popular method of diagnosing specific learning disorders. A recent study using data simulation methods claimed that cross-battery assessment was insufficiently accurate for clinical use (Stuebing, Fletcher, Branum-Martin, & Francis, 2012). However, the study used the general population base rate for specific learning disorders, resulting in misleadingly low accuracy estimates. The current study attempted to accurately simulate the cross-battery assessment method with a referred population prevalence to provide a fair analysis of cross-battery assessment's diagnostic accuracy. Under the assumptions modeled, cross-battery assessment's positive predictive value for specific learning disorder detection was between 71% and 93%, with negative predictive value between 43% to 80%. In addition, each additional testing phase adds increased diagnostic accuracy with diminishing returns.

KEYWORDS: Cross-Battery Assessment, Diagnostic Accuracy, Monte Carlo, Signal Detection Theory, Simulation Study, SLD, Specific Learning Disorder, XBA

OPTIMIZING CROSS-BATTERY ASSESSMENT PROCEDURES FOR
READING-BASED SPECIFIC LEARNING DISORDER DIAGNOSIS:
A MONTE CARLO STUDY

ZACHARY JOSEPH ROMAN

A Thesis Submitted in Partial
Fulfillment of the Requirements
for the Degree of

MASTER OF SCIENCE

Department of Psychology

ILLINOIS STATE UNIVERSITY

2016

© 2016 Zachary Joseph Roman

OPTIMIZING CROSS-BATTERY ASSESSMENT PROCEDURES FOR
READING-BASED SPECIFIC LEARNING DISORDER DIAGNOSIS:
A MONTE CARLO STUDY

ZACHARY JOSEPH ROMAN

COMMITTEE MEMBERS:

W. Joel Schneider, Chair

Matthew Hesson-McInnis

ACKNOWLEDGMENTS

I would first and foremost like to thank my advisor professor W. Joel Schneider of Illinois State University for his efforts in completing this project, without him this document would not exist. Many hours were spent in his office, working out complex details, many of which I once thought inconceivable. As a professor and a mentor, Joel's contribution to this project and my education have motivated me to strive for the level of excellence he showed me. It did not go unnoticed and will not be forgotten. I also want to express my thanks to professors Matthew Hesson-McInnis and Karla J. Deopke of Illinois State University for their contributions as my thesis committee. I also want to express my sincere gratitude to Dawn Flanagan, Samuel Ortiz, and Vincent Alfonso for generously loaning me a copy of the XBASS software and providing a detailed explanation of its calculations. I would also like to thank Karla Stuebing of the University of Houston. It was her original work that started the conversation that this document contributes to. I also want to thank my peers and future colleagues, Tatsu Shigeta, Aaron Neitzel, and Daniel Schloessor for their unknowing contributions through "pub talk brainstorming" which spurred the creative leaps needed to move forward. Finally, I would like to thank Melanie Messick for her unwavering support and on demand availability for proof reading and editing, her APA style knowledge bars none. This accomplishment would not have been possible without each person's unique contributions. For this, I thank you.

Z. J. R.

CONTENTS

	Page
ACKNOWLEDGMENTS	i
CONTENTS	ii
TABLES	v
FIGURES	vi
CHAPTER	
I. THE PROBLEM AND ITS BACKGROUND	1
Statement of the Problem	1
II. REVIEW OF RELEVANT LITERATURE	2
Diagnosing Specific Learning Disorder	2
Defining Specific Learning Disorder	2
Differentiating Specific Learning Disorder from Intellectual Disability	4
Cattell-Horn-Carroll Theory	4
CHC Theory Hierarchy	5
Fluid reasoning (Gf)	5
Working memory (Gwm)	6
Long term storage and retrieval (Glr)	7
Visual processing (Gv)	7
Auditory processing (Ga)	7
Speed of processing (Gs)	7
Defining Cross-Battery Assessment	8
Cross-Battery Assessment Procedure	9
Level 1 identification	9
Level 2 identification	10
Level 3 identification	10
Level 4 identification	11
Level 5 identification	11

	Cognitive Profiles	12
	Normative cognitive profile	12
	Intellectual disability cognitive profile	13
	Specific Learning Disorder reading subtype cognitive profile	13
	Academic Assessment in Cross-Battery Assessment Specific Learning Disorder Identification	14
	Monte Carlo Studies and Simulated Data	15
	Diagnostic Decision Making	15
	Accuracy of tests	15
	Accuracy of decision making	16
	Influences on the accuracy of diagnostic decision making	17
	Prior Evaluation of Cross-Battery Assessment	21
	Simulation Studies	23
III.	METHODOLOGY	25
	Creating Data: Latent and Observed	25
	Classification of Cases	30
	Summary	32
IV.	RESULTS	34
	Diagnostic Accuracy of the Cross-Battery Assessment Method	34
	Accuracy of Cross-Battery Assessment Academic Deficit Identification	36
	Accuracy of Cross-Battery Assessment Cognitive Deficits Identification	37
	Accuracy of G Probability Above Average	38
	Accuracy of Cross-Battery Assessment Defined Unexpected Deficit	39
	Diagnostic Accuracy of Repeated Testing	40
	Accuracy Cognitive Deficits Identification with Repeated Testing	42
	Accuracy of G Probability Above Average with Repeated Testing	44
	Accuracy of Unexpected Deficit	46
V.	DISCUSSION, IMPLICATIONS, AND LIMITATIONS	48
	Accuracy of the Cross-Battery Assessment Method	48
	Improving the Diagnostic Accuracy of Cross-Battery Assessment with Repeated Testing	49
	Limitations	50
	REFERENCES	52

TABLES

Table		Page
1.	Forms of Specific Learning Disorder Impairment	3
2.	Primary CHC Abilities Contributing to Reading Achievement	9
3.	Raw (λ) and Fisher Transformed (z_λ) Subtest Loadings	26
4.	Academic Tests, Elements of Reading, and Relevant Cognitive Abilities	32
5.	Cross-Battery Assessment Specific Learning Disorder Criteria and Decisions Rules	33
6.	Overall Accuracy of Cross-Battery Assessment at Different Prevalence Levels	35
7.	Cross-Battery Assessment Diagnostic Accuracy of Academic Deficit Identification	37
8.	Cross-Battery Assessment Diagnostic Accuracy of Cognitive Deficit Identification	38
9.	Cross-Battery Assessment Diagnostic Accuracy of Unexpected Academic Deficits	39
10.	Including More Subtests Per Cognitive Cluster Increases Diagnostic Accuracy	41
11.	Accuracy of Identifying Specific Cognitive Deficits with Repeated Testing	42
12.	Accuracy of Identifying Average or g Scores with Repeated Testing	45
13.	Accuracy of Identifying Unexpected Low Achievement with Repeated Testing	46

FIGURES

Figure		Page
1.	CHC Theory Structural Model	6
2.	Intellectual Disability Cognitive Profile Example	13
3.	Specific Learning Disability Cognitive Profile Example	14
4.	Diagnosis Status and Cut Scores	17
5.	The Effect of Changing Cut Scores on Diagnostic Accuracy	18
6.	Effect of Prevalence on Diagnostic Accuracy (Sensitivity = .49, Specificity = .97)	20
7.	Example Distributions of Testing Populations and Cut Scores	21
8.	Comparing Referred to General Population	24
9.	Standardized Path Coefficients Used to Generate Latent Scores	28
10.	Distributions of Simulated Subtest Loadings on Latent Abilities	29
11.	Overall Accuracy of Cross-Battery Assessment as a Function of Prevalence	36
12.	Validity of Cognitive Ability Tests by Number of Tests in Composite	40

CHAPTER I

THE PROBLEM AND ITS BACKGROUND

Statement of the Problem

There are approximately 32,000 working school psychologists in the United States (Charvat, 2005), many of whom perform hundreds of psychoeducational evaluations every year. These assessments, while costly in time and money to school districts, are crucial for identifying children in need of assistance. Approximately 7% of children have been diagnosed with a Specific Learning Disorder (SLD), and recent research suggests that the disorder's true prevalence may be even higher (Stuebing et al., 2012). With over 324 million people living in the United States (U. S. Census Bureau, 2016, September 9), the number of children in need is large. Inaccurate identification of learning disorders diverts scarce resources from children who need them and puts unnecessary burdens on school districts that can little afford to waste time and money. Given these implications, it is important that the systems we have in place for identifying learning disorders are thoroughly evaluated and optimized as much as is feasible. A prior investigation of cross-battery assessment by Stuebing et al. (2012) has attempted to evaluate the diagnostic accuracy of the detection method by using simulated data. This investigation paints a dark picture for cross-battery assessment, showing it to be largely inaccurate. However, a closer look at the methods employed show the simulation did not properly mimic the cross-battery assessment method. Thus, an evaluation which accurately emulates cross-battery assessment is still needed.

CHAPTER II

REVIEW OF RELEVANT LITERATURE

Diagnosing Specific Learning Disorder

Diagnosis of specific learning disorder is typically done by a school psychologist, clinical neuropsychologist, or by a professional in a related discipline. There are three main approaches to specific learning disorder diagnosis: The Discrepancy Model, The Response-to-Intervention Model, The Patterns of Strengths and Weaknesses Model. The Discrepancy Model defines specific learning disorder as when a person's academic achievement is substantially lower than his or her general intelligence. This model was once dominant but is rarely used today because leading scholars deem it theoretically and psychometrically unsound (Fletcher et al., 1998; Speece & Shekita, 2002). The Response-to-Intervention Model defines specific learning disorder as poor academic achievement that persists despite appropriate instruction (Fuchs & Fuchs, 1998). The Patterns of Strengths and Weakness Model defines specific learning disorder as a specific cognitive deficit that causes low academic achievement. The cross-battery assessment model (Flanagan, Ortiz, & Alfonso, 2013) is the most popular method that operationalizes the Patterns of Strengths and Weaknesses approach to specific learning disorder assessment. Though widely used, cross-battery assessment has a number of underlying assumptions that can be questioned (Stuebing et al., 2012).

Defining Specific Learning Disorder

Specific Learning Disorder is a group of disorders in which specific cognitive deficits interfere with academic achievement (Kavale, Spaulding, & Beam, 2009). Specific learning disorder is not diagnosed when academic problems are primarily caused by poor instruction,

cultural differences, or low intelligence. Rather, learning to read, write, and calculate is particularly difficult because basic psychological functions such as working memory capacity, attention, memory, and language processing are deficient.

When diagnosing specific learning disorder with the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-V; American Psychiatric Association, 2013), the academic domain and specific area of deficit is specified. Table 1 provides the forms of specific learning disorder impairment as provided by the DSM-V.

Table 1

Forms of Specific Learning Disorder Impairment

SLD Subtype	Impairment
Reading	<ul style="list-style-type: none"> Word reading accuracy Reading rate or fluency Reading comprehension
Written expression	<ul style="list-style-type: none"> Spelling accuracy Grammar and punctuation accuracy Clarity or organization of written expression
Mathematics	<ul style="list-style-type: none"> Number sense Memorization of arithmetic facts Accurate or fluent calculation Accurate math reasoning

Impairments in reading, writing, and calculation are sometimes referred to as dyslexia, dysgraphia, and dyscalculia, respectively. The prevalence of each disorder is between 4 and 9 percent of the population (American Psychiatric Association, 2013).

Differentiating Specific Learning Disorder from Intellectual Disability

Intellectual disability is characterized by intellectual deficits that interfere with most or all areas of adaptive functioning, including social interactions and activities of daily living (American Psychiatric Association, 2013; Katz & Lazcano-Ponce, 2008). In specific learning disorder, cognitive abilities are in general at least average except for one or more narrow deficits. It is assumed that distinguishing clearly between intellectual disability and specific learning disorder is important for selecting appropriate interventions. In specific learning disorder, interventions tend to be more time-limited and narrowly targeted toward the specific deficits the person has. In contrast, support for people with intellectual disability are necessarily more comprehensive and longer term.

Cattell-Horn-Carroll Theory

The cross-battery assessment method of detecting specific learning disorder relies heavily on the Cattell-Horn-Carroll Theory of Cognitive Ability (CHC Theory) as its guiding framework of intelligence. CHC Theory has evolved over a long period of time. In the early 1900s, intelligence research was primarily focused on the general factor of intelligence (McGrew, 2009). Later research focused on other factors of intelligence beyond the general factor (Kaufman, 2009).

Cattell (1941, 1943) gave theoretical and empirical justification for splitting the general factor of intelligence into two factors, fluid intelligence (Gf) and crystallized intelligence (Gc). Later, he and John Horn found strong evidence for several other general factors of intelligence, including visual spatial ability (Gv), auditory processing (Ga), processing speed (Gs), and several other factors related to different aspects of memory (Horn & Cattell, 1966). At this time, the

Horn-Cattell model was just one of many viable models of intelligence. However, it became more recognized when Carroll (1993) reanalyzed every relevant dataset available before 1990 and found that the Horn-Cattell model was largely supported. Carroll's theory differed primarily on the inclusion of a general intelligence factor (g).

Because Carroll's theory and the Horn-Cattell model were so similar, both Carroll and Horn consented to McGrew's proposal that the two theories be yoked under a common framework to facilitate communication among scholars and practitioners (McGrew, 1997, 2005). Since this combination, CHC Theory has become the dominant model for multidimensional measures of intelligence (Alfonso, Flanagan, & Radwan, 2005; Sternberg, 2012).

CHC Theory Hierarchy

CHC Theory is a hierarchically structured model of cognitive ability (Newton & McGrew, 2010). CHC Theory uses the terms broad abilities and narrow abilities to classify constructs into their respective level of the hierarchy. There are about at least 8 broad abilities, as well as several others that are hypothesized to exist (McGrew, 2009). Associated with each broad ability, there are multiple narrow abilities (about 70 in all). Figure 1 shows the CHC hierarchy of broad and narrow abilities related to academic achievement. The following section will define CHC Theory's broad and narrow abilities, which have been found to be related to specific learning disorder and academic achievement at large. Definitions were drawn from Flanagan et al. (2013), Newton and McGrew (2010), McGrew (2009) and Alfonso et al. (2005).

Fluid reasoning (Gf). Gf is the ability to logically come to solutions to novel problems. Research has found support for at least three narrow abilities within Gf: Inductive Reasoning, General Sequential Reasoning, and Quantitative Reasoning. *Inductive Reasoning* is the ability to recognize characteristics of a greater whole, or pattern of a set of problems. *General Sequential Reasoning* is the ability to take rules or conditions and apply them to a problem solving solution.

This process can be considered hypothetical deductive reasoning. *Quantitative Reasoning* is the ability to inductively and or deductively produce logical solutions to mathematical problems.

Working memory (Gwm). Gwm is the ability to manipulate and process information stored in short-term memory. Two narrow abilities are of concern to cross-battery assessment, Memory Span and Attentional Control. Memory Span refers to the amount of information that is immediately accessible to consciousness. Short-term memories usually last no more than 30 seconds, unless they are consciously maintained by repeating the information over and over. Attentional Control is the ability to intentionally manipulate recently obtained information (e.g., sorting a list or multiplying large numbers in one’s head).

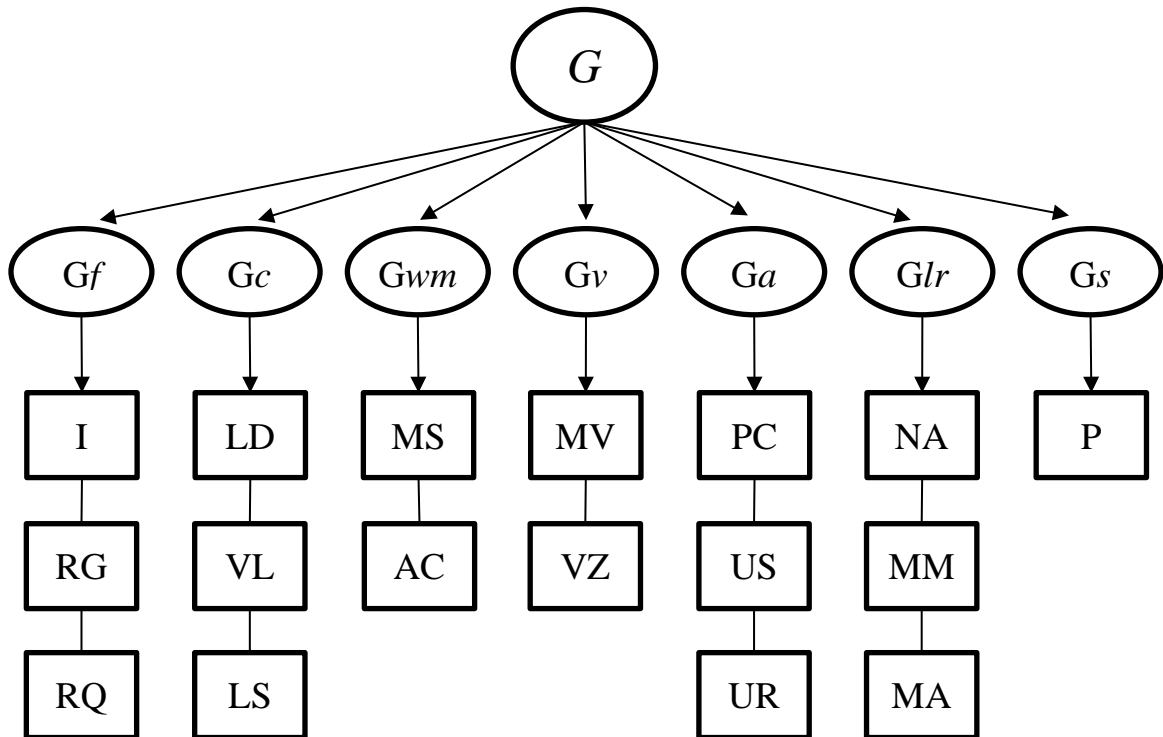


Figure 1. CHC Theory structural model.

Long term storage and retrieval (Glr). Similar to Gwm, Glr is the ability to store and retrieve information. However, Glr relates to the retrieval of information for times greater than approximately one minute after storage. Three principal narrow abilities of Glr are the target of cross-battery assessment: Naming Facility, Meaningful Memory, and Associative Memory. Naming Facility is the ability to quickly express accurate names for things and concepts when presented with a visual example, which is typically a picture. Meaningful Memory is the ability to retain information regarding the relationship of a set of items. Associative Memory is the ability to recall an item of a pair, when the other item is presented.

Visual processing (Gv). Gv is the ability to store, retrieve, and transform visual perceptions. Two primary narrow abilities under Gv are important to academic achievement and are subsequently assessed in the cross-battery assessment method of detecting specific learning disorder: Visual Memory, which is the ability to store mental images for later use; and Visualization, which is the ability to manipulate visual imagery in the mind's eye.

Auditory processing (Ga). Ga encompasses abilities that rely on perceiving and processing sound. Within Ga, three narrow abilities are of primary concern to the identification of specific learning disorder through the cross-battery assessment process of identifying specific learning disorder. Phonetic Coding is the ability to code, process and identify phonetic information (speech sounds). Sound Discrimination is the ability to recognize and discriminate between changes in phonemes or speech sounds with no distracting conditions. Auditory Stimulus Distortion is the ability to perceive sound correctly even when the sounds are distorted or when the listening environment is noisy.

Speed of processing (Gs). Gs is the ability to automatically solve basic, cognitive tasks when time restraints are present. The primary narrow ability of Gs of interest to the cross-battery assessment procedure or identifying specific learning disorder is Perceptual Speed, which is the ability to accurately search for, contrast, and recognize visual elements presented together or

separately in the visual field. Table 2 provides a summary of CHC abilities contributing to reading achievement.

Defining Cross-Battery Assessment

The cross-battery assessment approach is a systematic assessment method which relies on the CHC hierarchy to make diagnostic decisions of academic ability through analyzing a person's scores on relevant CHC abilities (Flanagan et al., 2013). Cross-battery assessment's putative advantage over other assessment methods is its ability to integrate findings from multiple test batteries into a coherent framework (Flanagan, Ortiz, & Alfonso, 2008). This advantage provides practitioners the means to investigate fully a student's cognitive ability profile, which a single battery alone typically cannot provide.

Table 2

Primary CHC Abilities Contributing to Reading Achievement

Broad Ability	Narrow Abilities
Crystallized Intelligence (Gc)	Language Development (LD) Lexical Knowledge (VL) Listening Ability (LS)
Long Term Memory(Glr)	Meaningful Memory (MM) Naming Facility (NA) Associative Memory (MA)
Fluid intelligence (Gf)	Inductive Reasoning (I)
Auditory Processing (Ga)	Phonetic Coding (PC)
Working Memory (Gwm)	Attentional Control (AC)

Cross-Battery Assessment Procedure

Cross-battery assessment specific learning disorder identification requires five levels of identification prior to diagnosis. Prior to the implementation of the procedure, a referral is first made, typical by a teacher or parent who finds evidence for a specific deficit. After a referral is made the procedure begins at Level 1.

Level 1 identification. The first level of identification is concerned with confirming the referral concern by observing an academic deficit. Investigation of academic performance is

conducted by reviewing class performance. If a deficit is present congruent with the referral concern, practitioners are instructed to proceed to Level 2 identification (Flanagan et al., 2013).

Level 2 identification. The second level of identification is concerned with identifying exclusionary factors, like second language status, or vision and hearing deficits. Level 2 identification takes cultural, and environmental factors into account such as inadequate instruction. If the academic deficit cannot be explained by exclusionary factors, the practitioner is instructed to proceed to Level 3 Identification (Flanagan et al., 2013).

Level 3 identification. At the third level of identification, cognitive ability and neuropsychological assessments now take place. The core battery is the primary battery which should measure the major broad abilities in CHC Theory related to the referral concern (Flanagan et al., 2013). Unfortunately, no battery measures all narrow abilities that might be needed for a particular case. Therefore, practitioners evaluate the core battery to decide which abilities are missing and need to be measured by tests in other batteries.

After the initial round of testing, practitioners input scores into the cross-battery assessment Software System (XBASS), which provides a detailed graphical and statistical output of the student's broad and narrow CHC Theory abilities (D. Flanagan, Ortiz, & Alfonso, 2015). If there is ambiguity as to the interpretation of a student's profile of abilities, the XBASS may prompt the practitioner to administer more tests before a reliable interpretation can be made. Failing to conduct follow-up tests is likely to result in non-cohesive test scores, making interpretations of composite tests unreliable (Flanagan et al., 2013). Those interested in the process that the XBASS uses to determine score cohesion and production of test composites should refer to Tables 3.2 - 3.5 in Flanagan et al. (2013) as well as the *Information tab* of the XBASS. If a cognitive deficit is observed practitioners are instructed to move to the Fourth level of Identification.

Level 4 identification. Level Four Identification is concerned with four questions:

1. Is the cognitive deficit contributing to the academic deficit?

If the answer to this question is true, the academic deficit will be accompanied by CHC theory cognitive deficits that have been shown to be associated with that academic weakness. It is common for people of average academic ability to have below average cognitive abilities that do not contribute to academic deficits, so both must be present.

2. Does the student display overall average ability?

This question is concerned with identifying that the student shows overall average ability. Recall that students with specific learning disorder show overall average cognitive ability and not a global deficit in functioning. This question is assessed by inspection of the overall estimated ability or CHC theories G value.

3. Is the learning deficit domain-specific?

This question further probes at how broad the deficit is. If the deficit is evident across multiple academic and cognitive domains, a different diagnosis may be required.

4. Is the underachievement unexpected?

This question is concerned with the predicted achievement at a given G value. For a specific learning disorder diagnosis under cross-battery assessment identification procedures, overall performance must be below what would be predicted from the G value. Operationalized, the prediction residual must be in the bottom 10% of prediction residuals. If the answer to all of these questions is *Yes*, the practitioner can now make a positive specific learning disorder diagnosis and move to the fifth and final level of identification.

Level 5 identification. Level Five identification is concerned with identifying the degree to which specific learning disorder is impairing the student. It is also concerned with concluding how much special resources and intervention techniques should be provided to the diagnosed individual.

Cognitive Profiles

A *cognitive profile* is a graphical representation on a standardized scale of a student's CHC abilities. In the XBASS, scores are presented in the index score metric ($\mu = 100, \sigma = 15$). Practitioners compare an individual profile to those that are characteristic of individuals with particular disorders.

Normative cognitive profile. A normative cognitive profile will have the majority of scores within one standard deviation (± 15) of the standardized population mean of 100. Most people have some variability in their profile, but the deviations are not typically extreme. Figure provides an example of a relatively flat normative cognitive profile in the average range.

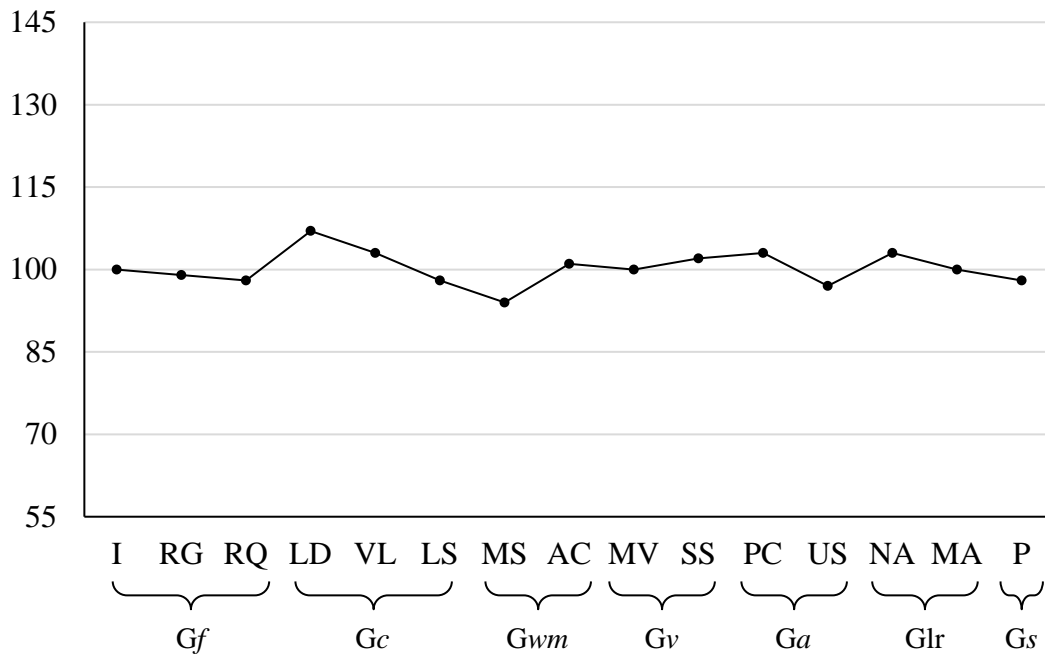


Figure 2. Normative cognitive profile.

Intellectual disability cognitive profile. A diagnosis of specific learning disorder is inappropriate if almost all abilities are very low. Figure 2 shows a cognitive profile of a person with intellectual disability. Note that intellectual disability is characterized by low scores across all CHC Theory Broad and Narrow abilities, though there may be isolated areas of relative strength in some individual.

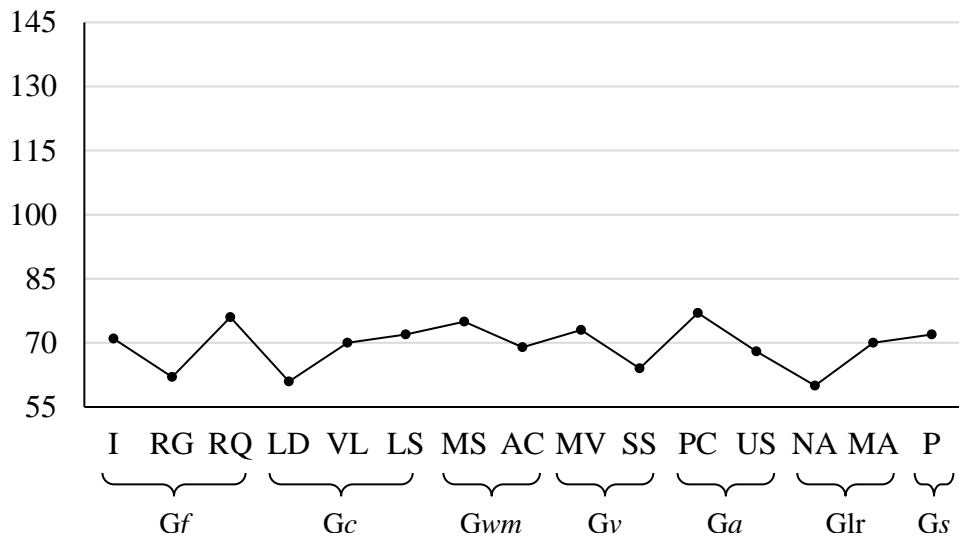


Figure 2. Intellectual disability cognitive profile example.

Specific Learning Disorder reading subtype cognitive profile. In Figure 3 most of the abilities are broadly average but there are specific deficits in Gwm, Ga, and Gs. If there are academic weaknesses in reading, a person with this profile may be diagnosed with a reading disability.

Academic Assessment in Cross-Battery Assessment Specific Learning Disorder

Identification

Identifying an academic deficit is necessary for a positive specific learning disorder diagnosis. This deficit must be in conjunction with a corresponding cognitive deficit. To assess academic deficiencies academic skills batteries are used. There are four primary academic abilities that contribute to reading: Reading fluency, reading decoding, and reading comprehension. The Wood-Cock Johnson Fourth Edition (WJIV; Woodcock, Johnson, Mather, 1990) a commonly used battery in cross-battery assessment utilizes four primary reading tests to assess the reading based academic abilities: letter word identification, reading fluency, passage comprehension, and work attack.

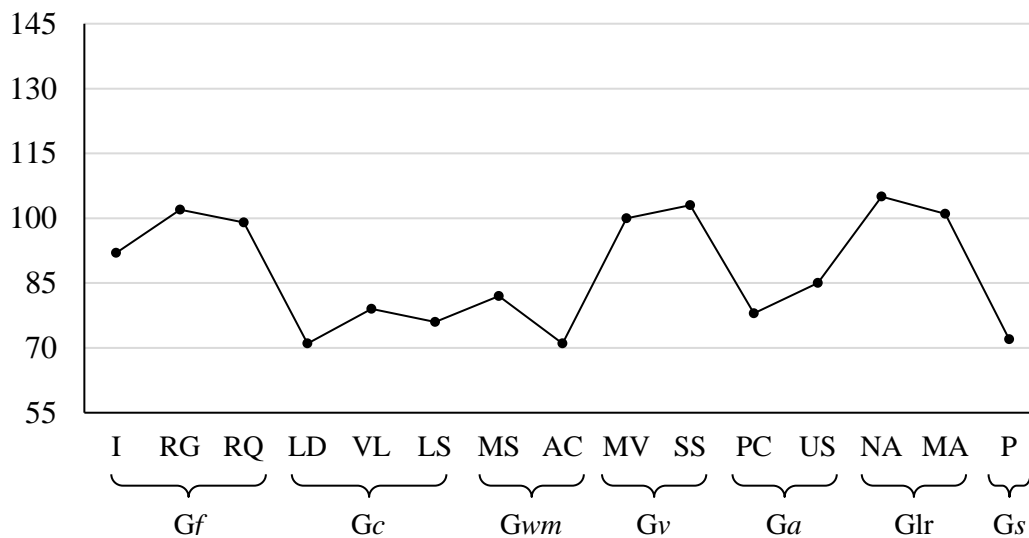


Figure 3. Specific learning disability cognitive profile example.

Monte Carlo Studies and Simulated Data

Diagnostic Decision Making

The accuracy of a diagnostic tool is commonly characterized using four test characteristics: Sensitivity, Specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV). Sensitivity and Specificity describe the accuracy of the test, while PPV and NPV describe the accuracy of decision making (Antczak-Bouckoms, Tulloch, Bouckoms, Keith, & Lavori, 1990). Whereas these two categories may seem to be the same, they are derived very differently, and describe different aspects of the diagnostic method.

To analyze diagnostic accuracy, the true score of every case must be known. The true “score” in this sense is the true diagnostic status (Meaning positive or negative diagnosis for Specific Learning Disorder). With real data, true scores are typically unobtainable. However, data simulation allows for the ability to create true scores (actual diagnosis) and obtained scores (test-identified diagnosis) for each case, allowing researchers to evaluate how well they align. To the degree that the data simulation process mimics the processes that generate real data, the simulated diagnostic accuracy results are informative.

Accuracy of tests. Sensitivity and specificity are both test characteristics that describe the accuracy of a test or diagnostic procedure. Sensitivity is the probability that a person who truly has a disorder will test positive for the disorder (Antczak-Bouckoms et al., 1990). Which is commonly referred to as the hit rate.

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} = \frac{\text{Correctly Identified SLD}}{\text{True SLD Cases}}$$

Specificity can be conceptualized as the probability of testing negative for a disorder that the given person truly does not have (Antczak-Bouckoms et al., 1990). Which is commonly referred to as the *true negative rate*.

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} = \frac{\text{Correctly Identified Non SLD}}{\text{True Non SLD Cases}}$$

Accuracy of decision making. PPV and NPV are both test characteristics that relate to the accuracy of diagnosis of a test. PPV can be conceptualized as; given a test outcome of positive, what is the probability a person truly has the disorder. (Antczak-Bouckoms et al., 1990; Pauker & Kassirer, 1980).

$$\text{PPV} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} = \frac{\text{Correctly Identified SLD Cases}}{\text{All Identified SLD Cases}}$$

NPV can be conceptualized as the probability of testing negative given you truly do not have the disorder (Antczak-Bouckoms et al., 1990; Pauker & Kassirer, 1980).

$$\text{NPV} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Negatives}} = \frac{\text{Correctly Identified Non SLD Cases}}{\text{All Identified SLD Cases}}$$

Figure 4 shows a distribution of an average diagnostic psychometric. The overlapping distributions exemplify the area in which decision making becomes more difficult. Someone with a very low or high score can easily be diagnosed accurately, however scores near the cut-off point are subject to falling into the false positive/negative range.

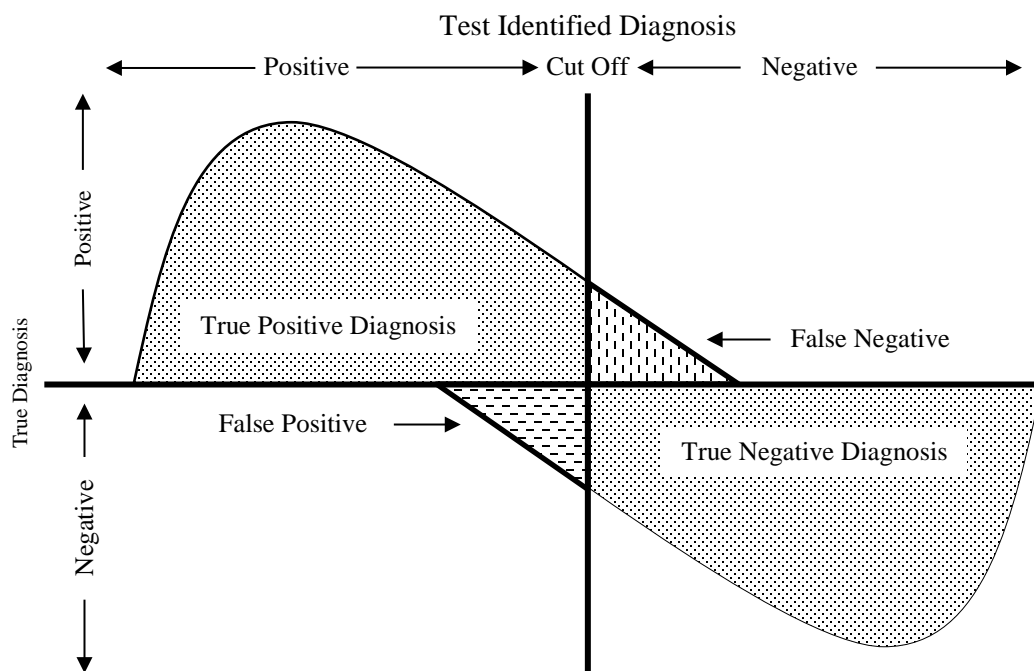


Figure 4. Diagnosis status and cut scores.

Influences on the accuracy of diagnostic decision making. The accuracy of diagnostic decision-making is influenced by multiple factors; including the threshold score (also referred to as the *cut point*) for making decisions, the prevalence of the disorder within the population being tested, and tests' ability to discriminate among the diagnostic groups (i.e., it's predictive validity).

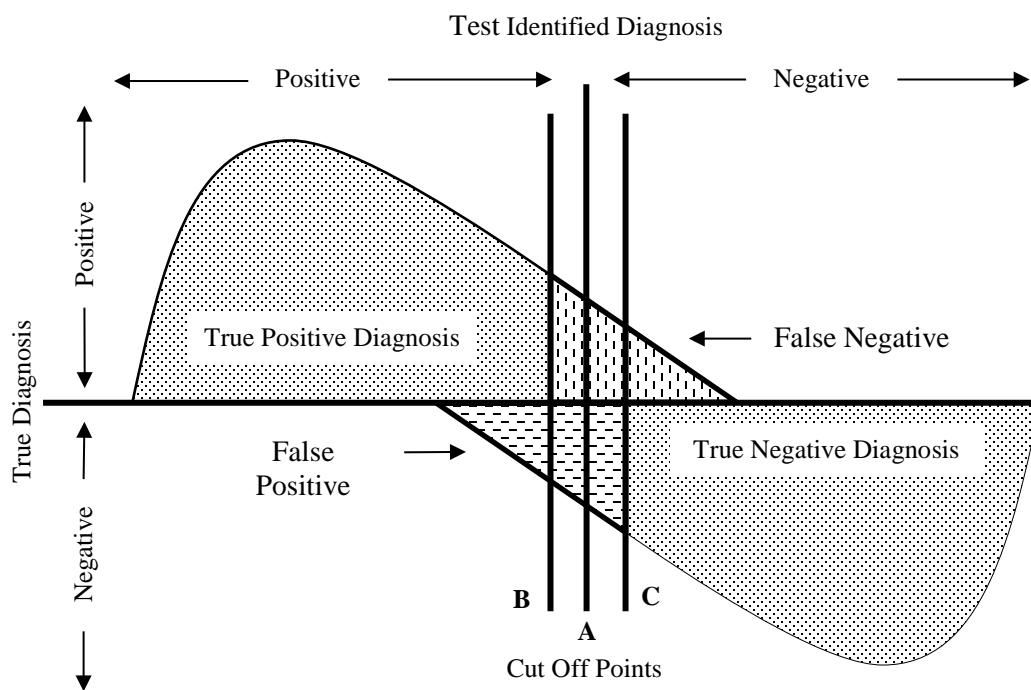


Figure 5. The effect of changing cut scores on diagnostic accuracy. Whereas cutoff B results in an increased false negative rate and decreased false positive rate and C results in a decreased false negative rate and an increased false positive rate.

Different cut scores minimize different kinds of diagnostic errors while increasing others. Figure 5 exemplifies this relationship. At cut-off point A the false positive and false negative rates are roughly equal. If one wishes to minimize false negatives (improving PPV and Sensitivity), the cut score can be moved to a higher value on the right. Unfortunately, doing so increases the false positive rate (decreasing Specificity and NPV). Lowering the cut score will have the opposite effect. False positives will be reduced (increasing NVP and Specificity) but false negatives will be increased (lowering PPV and Sensitivity).

Whereas the cut score is easily changed, the underlying validity of a test is not usually under a test developer's direct control. If a test has low reliability, it is possible to increase its reliability by making the test longer or by minimizing other kinds of measurement error (e.g., dropping confusing items or rewording them). If an unreliable test has underlying validity, increasing its reliability will indirectly raise its validity. However, increasing the reliability of a test with no validity will not improve the diagnostic accuracy of the test.

$$PPV = \frac{\text{Sensitivity} * \text{Prevalance}}{\text{Sensitivity} * \text{Prevalance} + (1 - \text{Specificity}) * (1 - \text{Prevalance})} \quad (1)$$

$$NPV = \frac{\text{Specificity} * (1 - \text{Prevalance})}{(1 - \text{Sensitivity}) * \text{Prevalance} + \text{Specificity} * (1 - \text{Prevalance})} \quad (2)$$

The prevalence of a disorder within a population also plays an important role in decision making accuracy. A low prevalence requires high specificity to provide acceptable levels of PPV. Inversely when prevalence is high, specificity must be very high to obtain acceptable levels of NPV. The relationship between prevalence, specificity, sensitivity, and its effects on PPV and NPV are in Equations 1 and 2 but they are more easily understood by examining Figure 6.

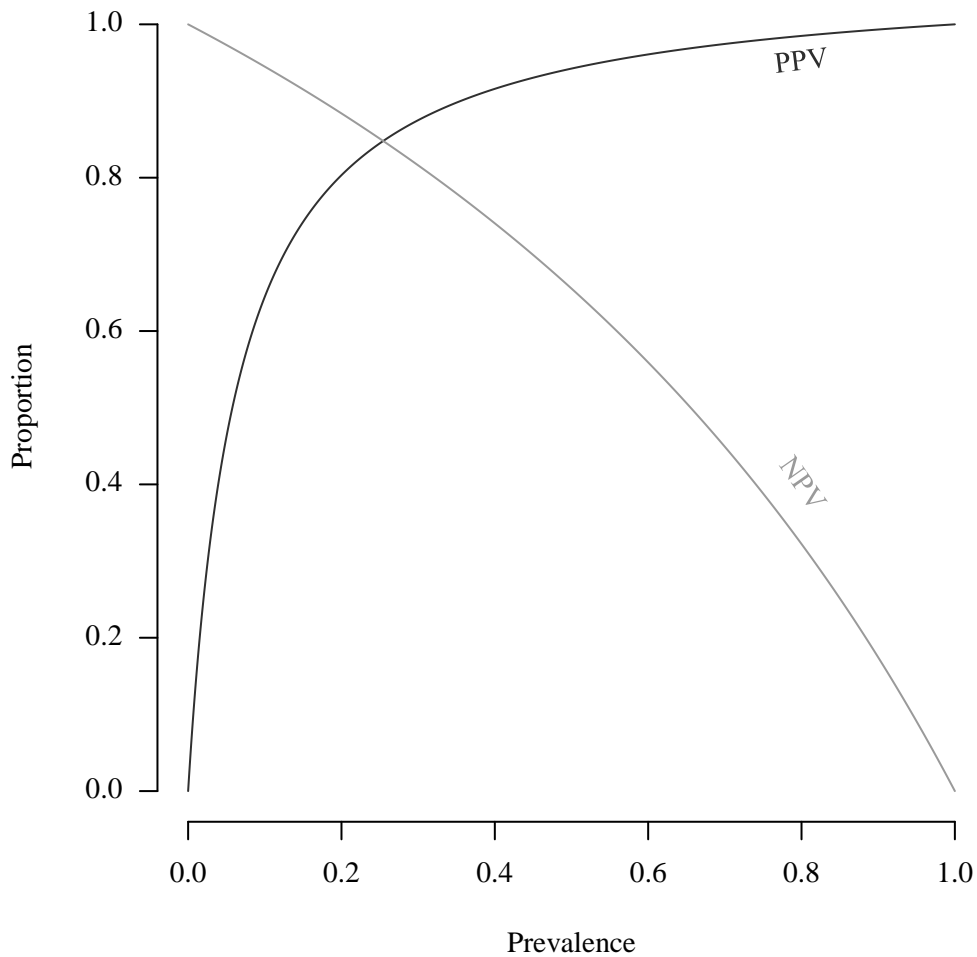


Figure 6. Effect of prevalence on diagnostic accuracy (Sensitivity = .49, Specificity = .97).

If the different diagnostic groups have distributions that are easily distinguished (i.e., their means are far apart or their within-group variability is low), the test's diagnostic accuracy statistics improve. As seen in the first panel of Figure 7, when the within-group distributions of a test overlap almost completely, the test does not permit accurate diagnostic decisions. In the second panel, the distributions have no overlap. If the cut score is anywhere between the distributions, the test will make perfectly accurate diagnostic decisions. The typical test is more

like the third panel in which the distributions are distinguishable but have some overlap, resulting in increased (but still imperfect) diagnostic accuracy. Unfortunately, the underlying distributions of the diagnostic groups are not typically under our control, and thus cannot be directly manipulated.

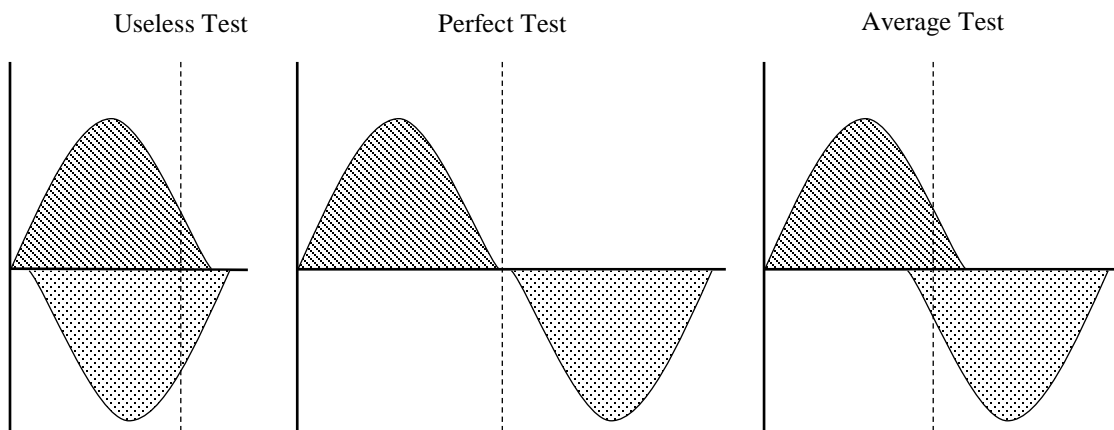


Figure 7. Example distributions of testing populations and cut scores.

Prior Evaluation of Cross-Battery Assessment

Stuebing et al. (2012) used data simulation methods to evaluate the diagnostic accuracy of the cross-battery assessment procedure for identifying specific learning disorder. Four cluster scores from the Woodcock-Johnson III (WJ III) were simulated: Gc, Ga, Gsm, and Glr. The correlations between the latent scores were estimated using the WJ III correlation matrix and reliability coefficients like so: $\rho_{XY} = \frac{r_{xy}}{\sqrt{r_{xx}r_{yy}}}$. For each cluster, a latent construct score and an accompanying “observed” cluster score were generated like so: $X = T_x\sqrt{r_{xx}} + e_x\sqrt{1 - r_{xx}}$.

Stuebing et al. (2012) then used cross-battery assessment procedure evaluation criteria to classify cases as having a specific learning disorder or not, based on observed scores. This process was conducted again at the latent level, allowing the latent specific learning disorder status to be compared with the “observed” specific learning disorder status.

Discrepancies between latent and observed decisions were used to calculate the diagnostic accuracy statistics for cross-battery assessment. Results varied, depending on the specific decision rules and the cognitive ability identified as the specific deficit. Specificity estimates were high, ranging from .94 to .97. Likewise, NPV was also high, ranging from .95 to .97. Thus, cross-battery assessment procedures were reasonably accurate in identifying those without specific learning disorder. However, the range of results for sensitivity (.46–.55) and PPV (.33–.53) suggest that cross-battery assessment is less accurate in identifying those with specific learning disorder. Practically speaking, if these results are correct, roughly half of children identified as having specific learning disorder by the cross-battery assessment method do not actually have specific learning disorder.

These numbers would, on the surface, appear to discredit cross-battery assessment as a viable method for specific learning disorder diagnosis. However, there are a number of important ways in which the procedures used by Stuebing et al. (2012) did not replicate the cross-battery assessment procedures (Schneider & Flanagan, 2016). Flanagan (Personal Communication, W. J. Schneider, April 15, 2016) is well aware of the potential for diagnostic errors, which is why cross-battery assessment has a number of features to minimize their occurrence. For example, before claiming that a particular low score is indeed a deficit, follow-up testing must confirm the presence of a deficit. That is, multiple rounds of assessment are necessary for diagnosis in many situations. In the study by Stuebing et al. (2012), there were no simulated follow-up measures. Thus, the study by Stuebing et al. (2012) provide diagnostic accuracy statistics for the first round

of cross-battery assessment testing. The accuracy of the cross-battery assessment procedure as articulated by its authors is still unknown.

There is a second reason to doubt that the low values of diagnostic decision making in the study by Stuebing et al. (2012) are as bad as they appear. In that study, the population of scores was simulated to mimic that of the general population. In the general population, the prevalence rates for reading disorders is between 5% and 15% (American Psychiatric Association, 2013). In the study by Stuebing et al. (2012), they considered the base rate to be between 3.7% to 7.8% (Stuebing et al., 2012).

As seen in Equation 1, PPV is directly proportional to the prevalence (base rate). Thus, in any context in which the base rate is lower, the PPV of the diagnostic procedure will be lower. Thus, the Stuebing et al. (2012) study found what the PPV would be if clinicians began testing people at random from the general population, where the prevalence is relatively low (5.4–7.8%). However, clinicians typically only assess children who are struggling in school to keep up. The referred population has a much high base rate for learning problems. In clinical settings, the PPV is likely to be much higher than that observed by Stuebing et al. (2012). How much higher? The base rate would vary considerably depending on the setting, but it seems reasonable that in many settings it is between 30 and 70%. With a higher base rate in a referred population, the PPV is higher, and the NPV is lower (See Figure 8). Given these concerns, it appears worthwhile to re-evaluate the accuracy of cross-battery assessment with improved methods.

Simulation Studies

Psychometric tests are used to assess constructs that are not directly observable. Although tests can provide reasonable estimates of those underlying constructs, there is always error and uncertainty associated with real measurements. Measurement error and bias come in many forms. Applied to the cross-battery assessment procedure of detecting specific learning disorder, the errors include referral bias, administration errors, and various kinds of cognitive

biases (e.g., confirmation bias; Kahneman, 2011) which interfere with accurate measurement and interpretation. With empirical data, it is virtually impossible to know how far off an individual data point is from its true value. For example, when assessing how accurate an empirical measure is, assessments of construct validity are implemented. Results obtained by the method are compared to others. However, every instrument has an error component, making it impossible to completely disentangle error variance from true variance. In contrast, with simulated data, it is possible to know exactly how much error there is in each measurement. Data simulation provides a metaphorical clean room to assess the overall validity of these diagnostic procedures.

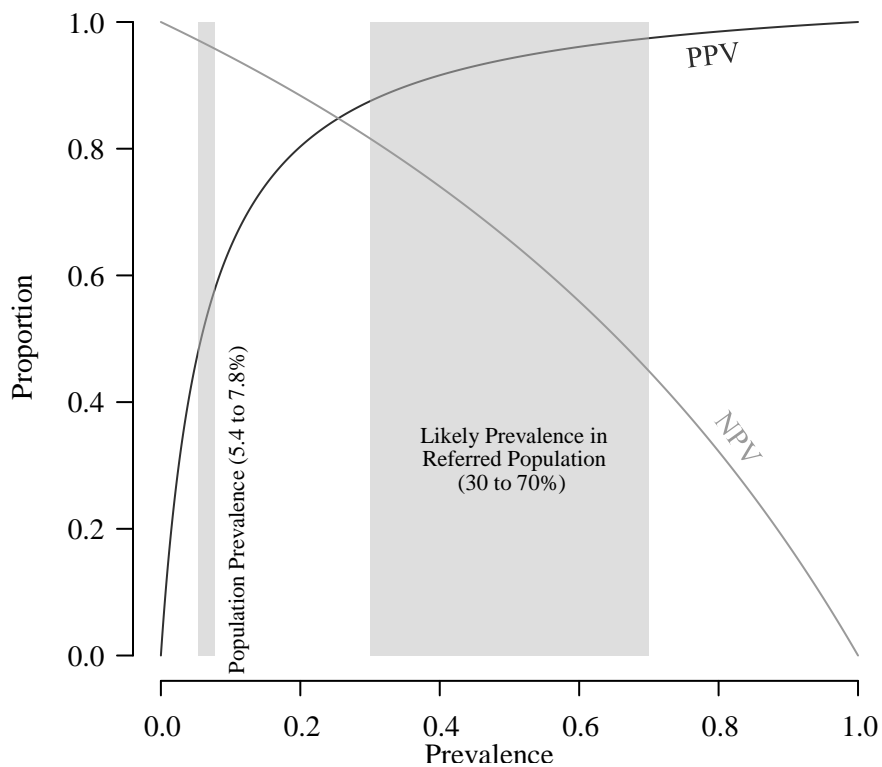


Figure 8. Comparing referred to general population.

CHAPTER III

METHODOLOGY

Stuebing and colleagues did the field a great service by starting the conversation about the accuracy of cross-battery assessment diagnostic methods. Although cross-battery assessment does need to be scrutinized, the methods by which cross-battery assessment is evaluated need to align with cross-battery assessment principles. The current study improved upon the methods Stuebing and colleagues used so that cross-battery assessment can be given a more thorough and accurate evaluation.

The current study produced not only one round of observed and latent scores, but as many iterations as needed per the cross-battery assessment method. Further, I used Monte Carlo data methods to estimate how many rounds of follow-up testing are needed to achieve adequate levels of accuracy in reading based specific learning disorder diagnosis. The term *adequate* is, of course, subjective, so a range has been provided. I will also explore how different base rates affect the accuracy of cross-battery assessment methods so that the accuracy of cross-battery assessment in referred populations can be estimated. I hypothesized that using a prevalence reflective of the referred population will show that cross-battery assessment is more accurate than a naïve reading of Stuebing et al. (2012) would suggest.

Creating Data: Latent and Observed

The first step for generating appropriate simulated data was to use multiple CHC based standardization samples to create a defensible structural model of the relations among CHC constructs and academic abilities. CHC models from the standardization samples found in the technical manuals of the KABC-II, WJ IV, and WISC-IV, were utilized (Kaufman,

Lichtenberger, Fletcher-Janzen, & Kaufman, 2005; McGrew & Woodcock, 2001; Woodcock, Johnson, & Mather, 1990). Furthermore, two large scale investigations of CHC relationships were included Keith and Reynolds (2010) and Reynolds, Keith, Fine, Fisher, and Low (2007). Table 3 provides the average subtest loadings by construct, as well as the average g loadings and standard deviations.

Table 3

Raw (λ) and Fisher Transformed (z_λ) Subtest Loadings

Ability	λ		z_λ	
	Mean	SD	Mean	SD
Ga	0.59	0.12	0.7	0.6
Gc	0.8	0.07	1.13	0.81
Gf	0.64	0.11	0.77	0.65
Glr	0.63	0.14	0.78	0.65
Gs	0.72	0.09	0.94	0.73
Gv	0.66	0.11	0.83	0.68
Gwm	0.69	0.1	0.87	0.7

Users of cross-battery assessment administer many different tests with a wide variety of loadings. To simulate what would happen in practice, subtest scores were simulated such that the loading on the latent variable was randomized for each individual. As standardized subtest loadings are correlations, the subtest loadings were simulated as normal variates and then the inverse Fisher's r to z transformation was applied. The means and standard deviations of the

normal variates came from the collected database of subtest loadings, to which the Fisher's r to z transformation was applied. Formula 3 shows the process, where F is the Fishers r to z transformation, F^{-1} is the inverse Fisher's r to z transformation, λ is the vector of subtest loadings, n is the sample size and `rnorm` is the call to create random normal data.

$$F^{-1}(\text{rnorm}(n, \text{mean}(F(\lambda)), \text{sd}(F(\lambda)))) \quad (3)$$

The second step for generating simulated data was to specify the latent variable model in terms of RAM notation and then use the `mvtnorm` package (Genz et al., 2015) in the R statistical programming environment to generate the data. If \mathbf{A} is the matrix of asymmetric paths (i.e. direct effects), \mathbf{S} is the matrix of symmetric paths (i.e., covariances), \mathbf{u} is the vector of exogenous variables (including disturbance and error terms), and \mathbf{v} is the vector of endogenous variables (including observed factor indicator variables), then \mathbf{u} can be generated with the Cholesky factorization of matrix \mathbf{S} via the `mvtnorm` package. The \mathbf{v} can be generated with Equation 3. 500,000 cases were simulated using the obtained structural model.

The second step for generating simulated data was to specify the latent variable model in terms of RAM notation and then use the `mvtnorm` package (Genz et al., 2015) in the R statistical programming environment to generate the data. If \mathbf{A} is the matrix of asymmetric paths (i.e. direct effects), \mathbf{S} is the matrix of symmetric paths (i.e., covariances), \mathbf{u} is the vector of exogenous variables (including disturbance and error terms), and \mathbf{v} is the vector of endogenous variables (including observed factor indicator variables), then \mathbf{u} can be generated with the Cholesky factorization of matrix \mathbf{S} via the `mvtnorm` package. The \mathbf{v} can be generated with Equation 4.

$$\mathbf{v} = \mathbf{u}((\mathbf{I} - \mathbf{A})^{-1})' \quad (4)$$

The standardized path coefficients in Figure 9 were used to generate latent construct scores. This model was developed using the WJ IV standardization sample. The paths first were chosen according to CHC theory and then non-significant paths were dropped (see Appendix for

details). This model was used mostly because it is based on real data and it seems plausible in light of CHC theory. No strong claims are being advanced here that this is the best possible model. If other paths had been included or some of these paths had been omitted, the results of this study would have been somewhat different, but the overall conclusions about the accuracy of cross-battery assessment would have been roughly the same.

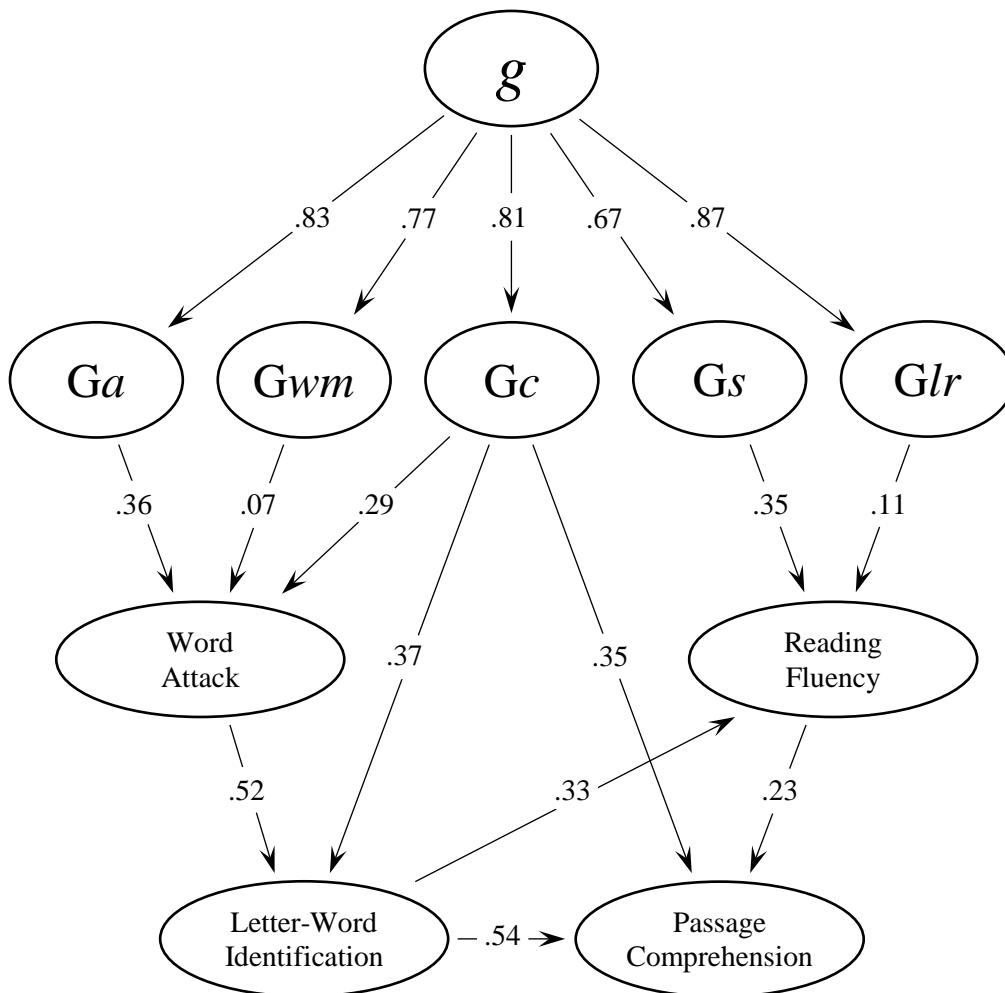


Figure 9. Standardized path coefficients used to generate latent scores.

Observed achievement scores ($N = 100,000$) were generated from the latent scores combined with error, the weights for which were taken from the WJ IV Technical Manual (McGrew, LaForte, & Schrank, 2015). Random subtest loadings were generated using parameters from Table 4 in a Fisher distribution (See Figure 10).

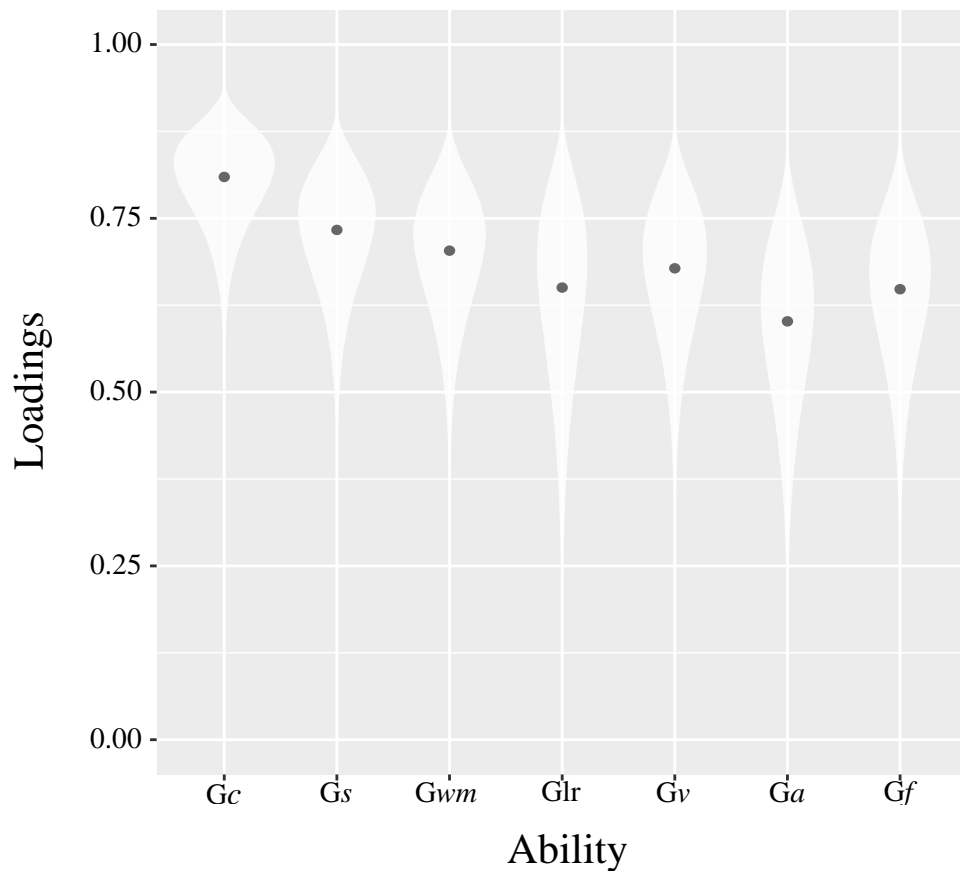


Figure 10. Distributions of simulated subtest loadings on latent abilities.

Classification of Cases

The classification of each observed case was determined to be specific learning disorder, or Not specific learning disorder. The classification method modeled mirrors that of the cross-battery assessment procedure to specific learning disorder identification as outlined by Flanagan et al. (2013) levels of identification. First, a reading based weakness must be observed to qualify for observed specific learning disorder identification; the deficit must be observed in reading fluency, decoding, comprehension, or a combination of these. At this step at least a one standard deviation deficit must have been observed in a reading based academic area. Next, this deficit must be observed in conjunction with a deficit in the CHC construct(s) related to the academic weakness. For example, if reading fluency is observed to be a deficit, an accompanying deficit in either G_c , G_f , or G_{wm} must be present to support a diagnosis of specific learning disorder. Table 4 provides the elements of reading and the associated cognitive abilities. If these criteria are not met, the case was observed as not having specific learning disorder.

Next, the academic deficit must be unexpected. This deficit is operationalized by having an estimated general intelligence score that is at least average. The XBASS produces a statistic, somewhat confusingly termed *G probability*, which is the probability that general intelligence (g) is at least average. The XBASS calculates the G probability by producing two weighted ratios from cognitive abilities that are considered strengths by the clinician; the XBASS then adds the two composites together and transforms them into a probability. The weights associated with each ability are: $G_c=.2355$, $G_f=.1870$, $G_{lr}=.1572$, $G_{wm}=.1152$, $G_v=.1167$, $G_a=.1029$, $G_s=.0864$. The weights were derived from prior studies which used factor analytic methods to determine the contribution of each ability to g . The cross-battery assessment manual recommends practitioners choose an ability to be a strength if it is no lower than $2/3$ of a standard deviation below the mean. The second stage in the calculation produces another ratio. In the second ratio, G_c , G_f , G_{lr} , and G_{wm} receive weights of .25 and are added together if it is chosen as a strength. Both values are

added together and divided by two to produce the G probability. For example, If G_c , G_f , G_{lr} , G_v , and G_s are chosen as strengths, the calculations will be as follows: $.2355 + .1870 + .1572 + .1167 + .0864 + .25 + .25 + .25 = 1.5328 / 2 = .7664$. Probabilities above .50 are regarded as likely to be normative. If G probabilities were below .50 even if congruent academic and cognitive deficits exist, the case was classified as Non SLD, because the deficit is not unexpected. In summary, if a reading based impairment was observed (Reading academic test score composite < 85) in conjunction with a deficient related CHC ability (Related CHC composite < 85) and the academic deficit is unexpected (Residual of g value prediction of academic performance $< 10^{\text{th}}$ percentile) and g is likely within the average or above range (G probability $> .50$) a classification of observed specific learning disorder was made. Otherwise a classification of Non SLD was made. Table 5 provides this information at a glance.

After observed cases are classified, the same process was conducted at the latent level. At this point, diagnostic test characteristics were calculated to determine the specificity, sensitivity, PPV and NPV of the cross-battery assessment method of specific learning disorder identification.

Table 4

Academic Tests, Elements of Reading, and Relevant Cognitive Abilities

Element of Reading	Relevant CHC Abilities
Reading Fluency	Crystallized Intelligence (<i>Gc</i>) Auditory Processing (<i>Ga</i>) Working Memory (<i>Gwm</i>)
Reading Decoding	Crystallized Intelligence (<i>Gc</i>) Auditory Processing (<i>Ga</i>) Long Term Memory (<i>Glr</i>) Working Memory (<i>Gwm</i>)
Reading Fluency	Long Term Memory (<i>Glr</i>) Processing Speed (<i>Gs</i>)
Comprehension	Processing Speed (<i>Gs</i>) Working Memory (<i>Gwm</i>) Long Term Memory (<i>Glr</i>) Crystallized Intelligence (<i>Gc</i>)

Summary

The current study aimed to answer the following research questions using simulated data based on the Cattell-Horn-Carroll Theory of Cognitive Abilities:

1. What is the diagnostic accuracy (specificity, sensitivity, PPV, and NPV) of cross-battery assessment in determining a diagnosis status of specific learning disorder (meaning, SLD or Not SLD) when the cross-battery assessment procedure is accurately simulated with a prevalence reflective of the referred population?

2. How much does increasing the number of subtests per cognitive composite increase the diagnostic accuracy of the specific learning disorder decision?

Table 5

Cross-Battery Assessment Specific Learning Disorder Criteria and Decisions Rules

SLD Condition	Decision Rule
At least one reading based impairment	<85
Related CHC reading impairment	<85
G probability likely average	>.50
Unexpected academic deficit, given g: Residual of prediction	< 10th Percentile

CHAPTER IV

RESULTS

Diagnostic Accuracy of the Cross-Battery Assessment Method

In any simulation study, the results depend on the decisions of the analyst and the explicit and implicit assumptions used to generate the data. The accuracy of cross-battery assessment as it is practiced in the real world cannot be evaluated directly because latent scores are unobservable. To avoid confusion with in vivo cross-battery assessment, I will refer to cross-battery assessment as I have modelled it in this study as cross-battery assessment. As seen in Table 6, the current study found accuracy results for cross-battery assessment similar those at the low end of Stuebing et al.'s results. With the current definitions, assumptions, and methods, 8% individuals in the population were classified as having a specific learning disorder. Although the current results suggest that cross-battery assessment has high specificity (.92), its sensitivity is low (.47). The low sensitivity of the cross-battery assessment method means that about half of the people with specific learning disorder were not diagnosed.

The implications of low specificity depend on which population is being assessed. In the general population the prevalence of specific learning disorder is low. Under the definition of specific learning disorder used in this study, the prevalence of specific learning disorder was about 8%, resulting in a PPV of .33. That is, if a person is selected at random from the population and diagnosed with specific learning disorder using cross-battery assessment, there is only about a one third probability that the person really does have specific learning disorder. However, as with the results in Stuebing et al. (2012), this low PPV is misleadingly low. When a person is

referred for an assessment because of academic problems, the prior probability of a learning disability is much higher than it would be for a person selected at random. As seen in Table 6 and

Figure 11, PPV rises to .72 when the base rate is 30% and to .93 when the base rate is 70%. At high base rates, NPV becomes dangerously low, making it hard to rule out specific learning disorder accurately.

Table 6

Overall Accuracy of Cross-Battery Assessment at Different Prevalence Levels

Accuracy Statistics	Prevalence			Stuebing et al. 3.7%–9.1%
	Current Results			
	8%	30%	70%	
Sensitivity	0.47	0.47	0.47	.44–.55
Specificity	0.92	0.92	0.92	.95–.97
Positive Predictive Value	0.33	0.72	0.93	.33–.53
Negative Predictive Value	0.95	0.80	0.43	.95–.98

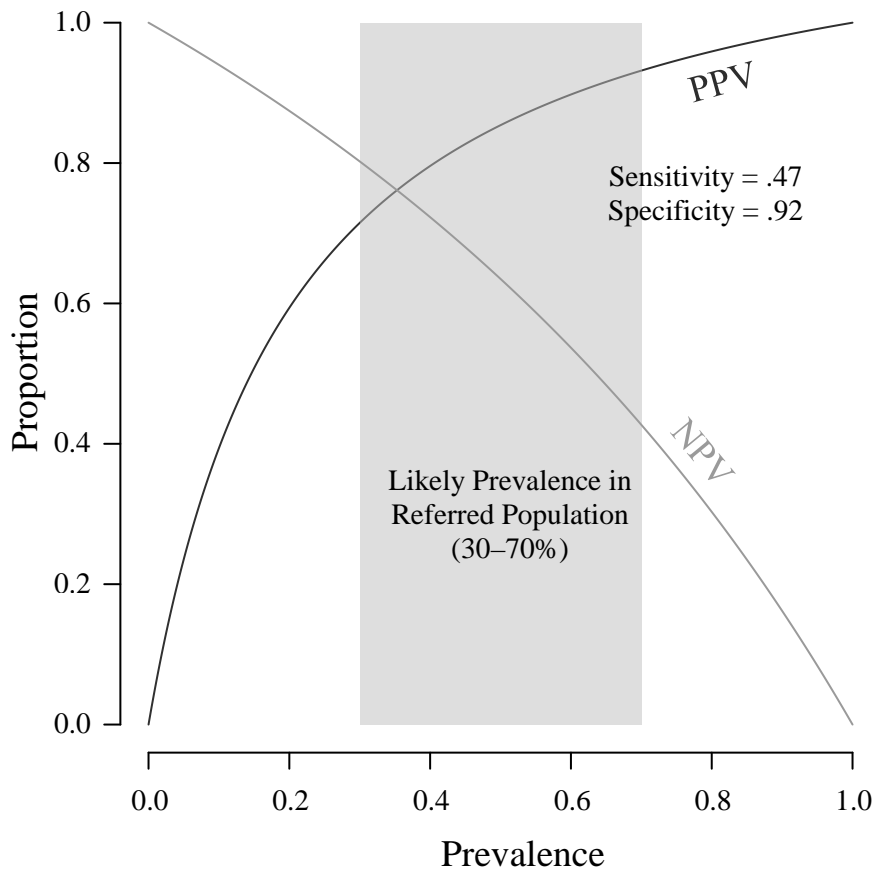


Figure 11. Overall accuracy of Cross-Battery Assessment as a function of prevalence.

Accuracy of Cross-Battery Assessment Academic Deficit Identification

I now turn to examining the accuracy of each step in the cross-battery assessment process. First, a decision must be made to determine if an academic deficit exists. At this level in reading assessment there are several outcomes: no deficit is observed, a deficit in sounding out phonologically regular words (Word Attack), a deficit in real-word decoding (Letter-Word Identification), a deficit in reading speed (Sentence Reading Fluency), a deficit in reading comprehension (Passage Comprehension), or any combination of reading deficits. The accuracy

of these decision is solely a function of the test reliability coefficient. Because the WJ IV Achievement tests tend to be reasonably reliable, the accuracy statistics were reasonably high. Table 7 summarizes all of diagnostic values for the academic assessments.

Table 7

Cross-Battery Assessment Diagnostic Accuracy of Academic Deficit Identification

Reading Ability	Sensitivity	Specificity	PPV	NPV	Prevalence
Word Attack	.79	.93	.79	.93	.25
Letter-Word Identification	.81	.94	.81	.94	.25
Sentence Reading Fluency	.82	.94	.83	.94	.25
Passage Comprehension	.75	.92	.75	.92	.25

Accuracy of Cross-Battery Assessment Cognitive Deficits Identification

At the cognitive level of assessment, a decision must be made if a cognitive deficit exists. Again, at this stage the current study compared the observed decisions made by the cross-battery assessment system to that of the latent level. Diagnostic accuracy statistics were then calculated. As seen in Table 8, abilities with the highest mean subtest loadings (see Table 3) have the best accuracy statistics (G_c and G_s). Abilities with lower subtest loadings have lower accuracy statistics in Table 8 (e.g., G_a and G_{lr}).

Table 8

Cross-Battery Assessment Diagnostic Accuracy of Cognitive Deficit Identification

CHC Ability	Sensitivity	Specificity	PPV	NPV	Prevalence
Gc	0.73	0.92	0.76	0.91	.26
Gf	0.61	0.88	0.67	0.86	.27
Ga	0.58	0.87	0.64	0.85	.28
Gv	0.63	0.89	0.68	0.87	.27
Glr	0.61	0.88	0.67	0.86	.28
Gwm	0.65	0.90	0.70	0.87	.27
Gs	0.67	0.90	0.72	0.88	.27

Accuracy of G Probability Above Average

The decision of cross-battery assessment in regards to G probabilities are categorical. The G probability is either above or below .50. Because this decision is categorical the diagnostic accuracy can be calculated. Cross-battery assessment' was found to have fairly accurate positive decisions (accurate at identifying cases above .50) with an obtained PPV of .93. However negative predictions suffered (accuracy of identifying cases below .50) with an obtained NPV of .69.

The reason for the lower NPV but higher PPV at this step is that the cross-battery assessment procedure for identifying G probability only uses cognitive strengths (i.e., scores average or better). This procedure biases the estimate of g upward. Under the assumptions modeled in this study, cross-battery assessment estimates of g were positively biased by .34 SD. This bias means that cross-battery assessment procedures are likely to lead to a judgment that g is at least average when it is not.

Accuracy of Cross-Battery Assessment Defined Unexpected Deficit

The final step in a cross-battery assessment diagnosis of specific learning disorder is determining whether an observed academic deficit in conjunction with an associated cognitive deficit is unexpected. This deficit is determined through the use of a regression model that predicts academic performance based on the G value. If the cases predicted performance is lower than what their G estimation predicts their performance to be, the deficit is unexpected. More specifically, if a cases standardized residual, is lower than the Z score critical threshold ($Z_{crit} = -1.28$) the deficit is unexpected. In other words, if the prediction of academic performance by the G value is below the 10th percentile, the deficit is unexpected. Comparing the observed decision to the latent level, diagnostic accuracy statistics were calculated. Reflective of the overall method, the NPV at this stage is high (.97), and the PPV is very low (.41-.47). Table 9 provides a summary of this information by academic test.

Table 9

Cross-Battery Assessment Diagnostic Accuracy of Unexpected Academic Deficits

	Sensitivity	Specificity	PPV	NPV	Prevalence
Word Attack	.59	.94	.47	.96	.08
Letter-Word Identification	.59	.94	.47	.96	.08
Ready Fluency	.58	.94	.46	.96	.08
Passage Comprehension	.51	.94	.41	.96	.08

Diagnostic Accuracy of Repeated Testing

The current study explored how the accuracy of decisions would change with repeated rounds of testing regardless of cross-battery assessment suggestion for follow up testing. Repeated testing was explored to investigate if the cross-battery assessment method of determining follow up testing actually provides accuracy above a static number of subtests.

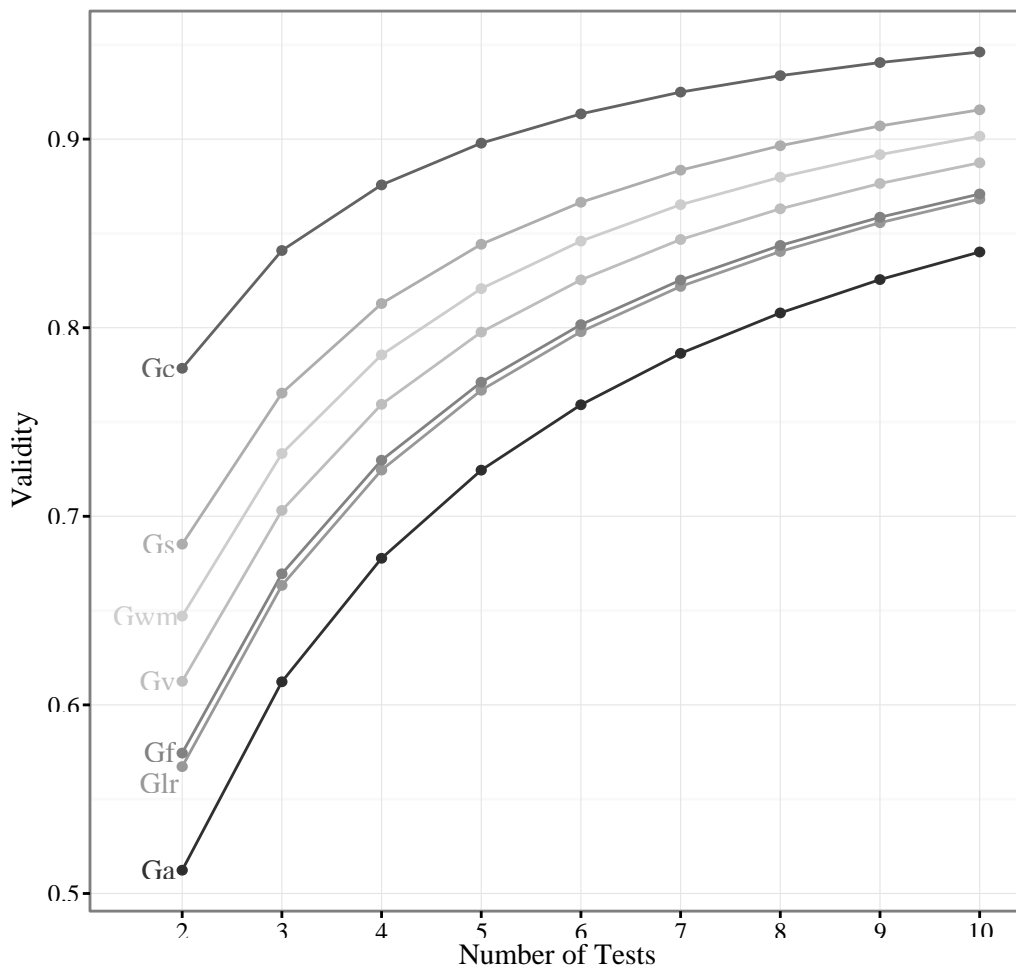


Figure 12. Validity of cognitive ability tests by number of tests in composite.

It is assumed that diagnostic accuracy would increase with more rounds of testing because each round of testing increases the validity of the cluster scores, as seen in Figure 12.

Diagnostic accuracy increases slowly with each round of testing, but with diminishing returns. Using regular composite scores rather than cross-battery assessment resulted in higher specificity but required five rounds of testing to reach the same level of sensitivity as cross-battery assessment. As seen in Table 10, each subsequent test adds both accuracy and validity; although there are diminishing returns with each additional test.

Table 10

Including More Subtests Per Cognitive Cluster Increases Diagnostic Accuracy

Statistic	<u>Accuracy</u>	<u>Number of Subtests Per Cognitive Cluster Score</u>								
	XBA	2	3	4	5	6	7	8	9	10
Sensitivity	0.47	0.44	0.45	0.46	0.47	0.48	0.49	0.49	0.5	0.5
Specificity	0.92	0.95	0.96	0.96	0.96	0.96	0.97	0.97	0.97	0.97
8% Prevalence										
PPV	0.33	0.44	0.47	0.5	0.51	0.53	0.54	0.55	0.56	0.56
NPV	0.95	0.95	0.95	0.96	0.96	0.96	0.96	0.96	0.96	0.96
30% Prevalence										
PPV	0.72	0.8	0.82	0.84	0.85	0.85	0.86	0.86	0.87	0.87
NPV	0.8	0.8	0.81	0.81	0.81	0.81	0.82	0.82	0.82	0.82
70% Prevalence										
PPV	0.93	0.96	0.96	0.97	0.97	0.97	0.97	0.97	0.97	0.97
NPV	0.43	0.42	0.43	0.43	0.44	0.44	0.44	0.45	0.45	0.45

Even with 10 subtests per cluster, sensitivity was only .50. Why is it so low? One of the reasons that increasing the number of tests per cognitive cluster did not result in high levels of sensitivity is that a specific learning disorder diagnosis does not just depend on the accuracy of the cognitive tests but also on the accuracy of the academic tests, which still were measured with a single test. To further understand the sources of inaccuracy with the repeated testing variant of cross-battery assessment, I will examine the accuracy of each diagnostic step. The accuracy of the achievement deficits is unchanged from Table 7.

Accuracy Cognitive Deficits Identification with Repeated Testing

One reason that repeated testing did not result in dramatic increases in diagnostic accuracy is that the validity of a composite score does not increase as quickly as its reliability when the number of subtests increases, especially when the subtest loadings on the latent abilities are low (see Figure 12). Table 11 displays how much diagnostic accuracy improves with repeated testing. the accuracy of identifying cognitive deficits. Repeated testing adds only modest benefits to both PPV and NPV.

Table 11

Accuracy of Identifying Specific Cognitive Deficits with Repeated Testing

Ability	Subtests	Sensitivity	Specificity	PPV	NPV	Prevalence
Gc	2	0.76	0.92	0.75	0.92	0.25
	3	0.80	0.93	0.79	0.93	0.25
	4	0.82	0.94	0.81	0.94	0.25
	5	0.84	0.94	0.83	0.95	0.25
	6	0.85	0.95	0.85	0.95	0.25
	7	0.86	0.95	0.86	0.95	0.25

(Table continues)

Ability	Subtests	Sensitivity	Specificity	PPV	NPV	Prevalence
<i>Ga</i>	8	0.87	0.96	0.87	0.95	0.25
	9	0.87	0.96	0.88	0.96	0.25
	10	0.88	0.96	0.89	0.96	0.25
	2	0.62	0.87	0.61	0.87	0.25
	3	0.66	0.89	0.66	0.89	0.25
	4	0.70	0.90	0.70	0.90	0.25
	5	0.72	0.91	0.72	0.91	0.25
	6	0.74	0.91	0.74	0.91	0.25
	7	0.76	0.92	0.76	0.92	0.25
	8	0.77	0.92	0.77	0.92	0.25
<i>Gwm</i>	9	0.79	0.93	0.78	0.93	0.25
	10	0.79	0.93	0.80	0.93	0.25
	2	0.69	0.89	0.68	0.90	0.25
	3	0.73	0.91	0.72	0.91	0.25
	4	0.76	0.92	0.76	0.92	0.25
	5	0.78	0.93	0.79	0.92	0.25
	6	0.79	0.93	0.80	0.93	0.25
	7	0.81	0.94	0.82	0.94	0.25
	8	0.82	0.94	0.83	0.94	0.25
	9	0.83	0.94	0.83	0.94	0.25
<i>Glr</i>	10	0.84	0.95	0.84	0.95	0.25
	2	0.64	0.88	0.65	0.88	0.25
	3	0.70	0.89	0.69	0.90	0.25
	4	0.73	0.91	0.72	0.91	0.25
	5	0.75	0.92	0.75	0.91	0.25
	6	0.77	0.92	0.76	0.92	0.25
	7	0.78	0.93	0.78	0.93	0.25
	8	0.79	0.93	0.80	0.93	0.25
	9	0.81	0.93	0.80	0.94	0.25
	10	0.82	0.94	0.81	0.94	0.25
<i>Gs</i>	2	0.70	0.90	0.70	0.90	0.25
	3	0.74	0.92	0.75	0.91	0.25
	4	0.77	0.93	0.78	0.92	0.25
	5	0.80	0.93	0.79	0.93	0.25

(Table continues)

Ability	Subtests	Sensitivity	Specificity	PPV	NPV	Prevalence
<i>G_f</i>	6	0.81	0.94	0.81	0.94	0.25
	7	0.82	0.94	0.83	0.94	0.25
	8	0.83	0.95	0.84	0.94	0.25
	9	0.84	0.95	0.85	0.95	0.25
	10	0.85	0.95	0.85	0.95	0.25
	2	0.64	0.88	0.65	0.88	0.25
	3	0.70	0.90	0.69	0.90	0.25
	4	0.73	0.91	0.73	0.91	0.25
	5	0.75	0.92	0.76	0.91	0.25
	6	0.77	0.92	0.77	0.92	0.25
<i>G_v</i>	7	0.78	0.93	0.79	0.93	0.25
	8	0.80	0.93	0.79	0.93	0.25
	9	0.81	0.94	0.81	0.94	0.25
	10	0.81	0.94	0.82	0.94	0.25
	2	0.66	0.89	0.68	0.88	0.26
	3	0.71	0.90	0.72	0.90	0.25
	4	0.74	0.91	0.74	0.91	0.25
	5	0.77	0.92	0.76	0.92	0.25
	6	0.78	0.93	0.79	0.93	0.25
	7	0.80	0.93	0.80	0.93	0.25
8	0.81	0.94	0.81	0.93	0.25	
9	0.82	0.94	0.82	0.94	0.25	
10	0.83	0.94	0.83	0.94	0.25	

Accuracy of G Probability Above Average with Repeated Testing

Diagnostic accuracy statistics were calculated for the accuracy of the G probability decision (G probability > .50 or not). Because the accuracy of the G probability decision is already moderately high, there is not a substantial increase in the diagnostic accuracy statistics. Although it is worth noting that the accuracy of the decision being non deficient (NPV) increases the more than the PPV. As can be seen in Table 12, the PPV results are similar to that of the

analogous step in cross-battery assessment. However, the NPV is higher, even with just two subtests per broad ability. The reason for this increased accuracy is that no low scores are discarded in the estimate of g . In contrast, cross-battery assessment discards low cluster scores when g is estimated. Instead of 75% of people having average or better g , cross-battery assessment identifies 79%, which is an overestimate. For this reason, the cross-battery assessment practice of omitting cognitive weakness from the estimate of g may be a source of diagnostic inaccuracy, at least under the assumption modelled in this study.

Table 12

Accuracy of Identifying Average or g Scores with Repeated Testing

Subtests	Sensitivity	Specificity	PPV	NPV	Prevalence
2	.95	.73	.91	.84	.75
3	.96	.75	.92	.86	.75
4	.96	.76	.92	.88	.75
5	.97	.77	.93	.89	.75
6	.97	.78	.93	.89	.75
7	.97	.78	.93	.90	.75
8	.97	.78	.93	.90	.75
9	.97	.79	.93	.90	.75
10	.97	.79	.93	.91	.75
XBA	.92	.73	.93	.70	.79

Accuracy of Unexpected Deficit

For a diagnosis of specific learning disorder, the observed academic deficit must be unexpected. Which is determined through the use of a regression model that predicts academic performance based on the G value. If the predicted performance is substantially lower than what is predicted from g, the deficit is unexpected. Cross-battery assessment defines unexpectedly low achievement as when the prediction error is in the bottom 10% of prediction errors. Compared to cross-battery assessment, the repeated testing variant of cross-battery assessment resulted in more accurate identification of unexpectedly low achievement because it estimated g more precisely and with less bias (see Table 12). The estimates of achievement are the same in both methods. However additional rounds of testing do not substantially increase the accuracy of identifying unexpectedly low achievement, as seen in Table 13.

Table 13

Accuracy of Identifying Unexpected Low Achievement with Repeated Testing

Academic Test	Subtests	Sensitivity	Specificity	PPV	NPV	Prevalence
Word Attack	2	0.61	0.96	0.61	0.96	0.10
	3	0.61	0.96	0.62	0.96	0.10
	4	0.62	0.96	0.62	0.96	0.10
	5	0.62	0.96	0.62	0.96	0.10
	6	0.62	0.96	0.62	0.96	0.10
	7	0.62	0.96	0.62	0.96	0.10
	8	0.62	0.96	0.62	0.96	0.10
	9	0.62	0.96	0.62	0.96	0.10
	10	0.62	0.96	0.62	0.96	0.10
	Letter Word Identification	2	0.63	0.96	0.62	0.96
3		0.63	0.96	0.63	0.96	0.10

(Table continues)

Academic Test	Subtests	Sensitivity	Specificity	PPV	NPV	Prevalence
	4	0.64	0.96	0.64	0.96	0.10
	5	0.64	0.96	0.64	0.96	0.10
	6	0.64	0.96	0.64	0.96	0.10
	7	0.64	0.96	0.64	0.96	0.10
	8	0.65	0.96	0.64	0.96	0.10
	9	0.65	0.96	0.64	0.96	0.10
	10	0.65	0.96	0.64	0.96	0.10
Reading Fluency						
	2	0.62	0.96	0.61	0.96	0.10
	3	0.62	0.96	0.62	0.96	0.10
	4	0.63	0.96	0.63	0.96	0.10
	5	0.63	0.96	0.63	0.96	0.10
	6	0.63	0.96	0.63	0.96	0.10
	7	0.63	0.96	0.63	0.96	0.10
	8	0.64	0.96	0.64	0.96	0.10
	9	0.64	0.96	0.64	0.96	0.10
	10	0.64	0.96	0.64	0.96	0.10
Passage Comprehension						
	2	0.54	0.95	0.54	0.95	0.10
	3	0.54	0.95	0.54	0.95	0.10
	4	0.54	0.95	0.54	0.95	0.10
	5	0.55	0.95	0.54	0.95	0.10
	6	0.55	0.95	0.55	0.95	0.10
	7	0.55	0.95	0.55	0.95	0.10
	8	0.55	0.95	0.55	0.95	0.10
	9	0.55	0.95	0.55	0.95	0.10
	10	0.55	0.95	0.55	0.95	0.10

CHAPTER V

DISCUSSION, IMPLICATIONS, AND LIMITATIONS

Accuracy of the Cross-Battery Assessment Method

There are over 70 million children in the United States. The Diagnostic and Statistical Manual of Mental disabilities cites the prevalence of specific learning disorder to be between 5 and 12 percent. Which means that approximately 3.5 Million to 8.4 Million children have this condition. With such a large number of children in need and the high level of training necessary; psychoeducational assessment is an expensive industry. Inaccurate identification not only adds to the cost of assessment on schools, but diverts already scarce resources from those who can benefit most.

A high NPV is considered more important in scenarios such as diagnostic decisions because a false negative is regarded as much more detrimental than a false positive. Positive diagnosis are typically followed up with further investigation, which makes a false positive more likely to be detected later. However, a negative diagnosis is frequently not followed up, so later detection is less likely.

The results of the current study found that cross-battery assessment had low sensitivity (.47) but higher specificity (.92). Which means that about half of those with specific learning disorder are identified by cross-battery assessment as not having specific learning disorder and about 8% of those without specific learning disorder are identified by cross-battery assessment as having specific learning disorder. The accuracy of diagnosis depends on the prevalence of specific learning disorder in the population being assessed. In the general population, specific learning disorder might have a prevalence around 8%. If the general population is assessed, the

PPV for cross-battery assessment is .33, meaning that only about one third of those diagnosed with a reading specific learning disorder by cross-battery assessment procedures truly have it. At a prevalence of 8%, the NPV is .95, meaning that the vast majority of those deemed by cross-battery assessment not to have specific learning disorder actually do not have it.

However, the general population is rarely assessed for specific learning disorder. People referred for specific learning disorder assessments certainly have a higher prevalence of specific learning disorders than does the general population. If we conservatively assume that between 30 and 70 percent of people assessed for specific learning disorder truly have specific learning disorder, then the PPV is much higher than .33, ranging from .72 to .93. Unfortunately, the NPV dips ever lower, the more prevalent specific learning disorder is in a referred population, ranging from .80 at a prevalence of 30% to .43 at a prevalence of 70%. This is an alarming rate of false negative diagnosis. False negative diagnoses withhold the important resources from children who need them.

Improving the Diagnostic Accuracy of Cross-Battery Assessment with Repeated Testing

The second goal of the study was to find out how many iterations of testing it takes to reach acceptable levels of diagnostic accuracy. The major result of this analysis is that increasing the number of subtests per cluster improves diagnostic accuracy but only by a little bit. Achieving high levels of accuracy would require impractical amounts of testing time (e.g., administering more than 10 subtests per broad ability).

A second major finding of this study was that using all the scores in a composite resulted in higher accuracy than the cross-battery assessment method of finding coherent clusters. Likewise, using all scores to estimate g is more accurate than the cross-battery assessment method of discarding cognitive weaknesses. Using all available test scores as opposed to determining score cohesion produces more accurate results. Therefore, identifying a target area of

weakness and performing multiple assessments for each associated cognitive ability will produce more accurate results while maintaining practicality in time constraints.

To achieve high levels of diagnostic accuracy with a manageable number of tests, it is likely that increasing the underlying reliability and validity of the tests is going to be necessary. In the meantime, given the results of this paper, it seems increasing the number of tests per factor to a higher but still practical number is warranted (e.g., about 3 or 4 tests per factor).

An important consideration for this study is that most diagnostic missteps occur near the cut scores for the various decision rules. That is, most errors are near misses, not gross missteps. Each time a dichotomous decision is made (e.g. *g* deficit or not, coherent or not, unexpected or not) accuracy suffers. A possible way to remedy the repeated process of dichotomous decision making would be to use a dimensional approach instead of the binary decision approach in cross-battery assessment.

Limitations

There are several important limitations of the current study. Simulation studies are no better than the accuracy of their assumptions. It is possible that the current finding that regular composite scores are more accurate than the cross-battery assessment coherent composites would not hold true if the distribution of scores is not multivariate normal, as assumed here. For example, from time to time a child might misunderstand the test or cease to try on a test. Often the examiner sees what is happening but a certain percentage of the time misunderstandings and low effort are undetected. In such cases, it would be more accurate to discard anomalously low scores as mandated by cross-battery assessment.

Other possible inaccuracies in the modeling process might be that if the distribution of subtest loadings is biased or incorrect, the underlying causal model might be incorrect as well.

This simulation study only modeled cross-battery assessment Reading based specific learning disorder Identification. A single type of specific learning disorder identification (reading)

was chosen in an attempt to provide clear and succinct results. This restricts the generalization of obtained accuracy past that of reading subtype specific learning disorder identification within the cross-battery assessment framework.

There are a number of cross-battery assessment procedures that could not be modeled in this study. For example, cross-battery assessment has procedures and steps of identification for considering cultural factors; The primary cultural factor being English as a second language. It would be difficult to model such considerations meaningfully with simulated data.

Finally, only reading-related learning disabilities were modelled in this study. Future studies should investigate the accuracy of cross-battery assessment specific learning disorder identification in writing and math domains as well as reading.

REFERENCES

- Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). The impact of the Cattell-Horn-Carroll theory on test development and interpretation of cognitive and academic abilities. In *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 185–202). New York: Guilford.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). Washington, DC: American Psychiatric Association.
- Antczak-Bouckoms, A., Tulloch, J., Bouckoms, A. J., Keith, D., & Lavori, P. (1990). *Diagnostic Decision Making. Anesthesia Progress*, 37, 161–165.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin*, 38, 592.
- Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, 40(3), 153–193.
- Charvat, J. L. (2005). How many school psychologists are there? *NASP Communiqué*, 33(6), 7–8.
- Flanagan, D. P., Ortiz, S., & Alfonso, V. C. (2015). *Cross-battery assessment software system (X-BASS)*. Hoboken: John Wiley.
- Flanagan, D. P., Ortiz, S., & Alfonso, V. C. (2008). Response to intervention (RTI) and cognitive testing approaches provide different but complementary data sources that inform specific learning disorder identification. *NASP Communiqué*, 36(5), 16–17.

- Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). New Jersey: John Wiley & Sons.
- Flanagan, D. P., & Schneider, W. J. (2016). Cross-Battery Assessment? XBA PSW? A case of mistaken identity: A commentary on Kranzler and colleagues' "Classification agreement analysis of Cross-Battery Assessment in the identification of specific learning disorders in children and youth". *International Journal of School & Educational Psychology*, 4(3), 137–145.
- Fletcher, J. M., Francis, D. J., Shaywitz, S. E., Lyon, G. R., Foorman, B. R., Stuebing, K. K., & Shaywitz, B. A. (1998). Intelligent testing and the discrepancy model for children with learning disabilities. *Learning Disabilities Research & Practice*, 13(4), 186–203.
- Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2006). *Learning disabilities: From identification to intervention*. New York: Guilford.
- Fuchs, L. S., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research and Practice*, 13(4), 204–219.
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2015). *mvtnorm: Multivariate normal and t distributions* [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=mvtnorm> (R package version 1.0-3)
- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57(5), 253–270.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Macmillan.
- Katz, G., & Lazcano-Ponce, E. (2008). Intellectual disability: Definition, etiological factors, classification, diagnosis, treatment and prognosis. *Salud Pública de México*, 50(Su2), s132–s141.
- Kaufman, A. S. (2009). *IQ testing 101*. New York: Springer Publishing Co.

- Kaufman, A. S., Lichtenberger, E. O., Fletcher-Janzen, E., & Kaufman, N. L. (2005). *Essentials of KABC-II assessment*. Hoboken, NJ: John Wiley & Sons.
- Kavale, K. A., Spaulding, L. S., & Beam, A. P. (2009). A time to define: Making the specific learning disability definition prescribe specific learning disability. *Learning Disability Quarterly*, *32*(1), 39–48.
- Keith, T. Z., & Reynolds, M. R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we’ve learned from 20 years of research. *Psychology in the Schools*, *47*(7), 635–650.
- McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.). *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–179). New York: Guilford Press.
- McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 136–181). New York: Guilford Press.
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*(1), 1–10.
- McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Woodcock-Johnson IV technical manual*. Rolling Meadows, IL: Riverside.
- McGrew, K. S., & Woodcock, R. (2001). *Technical manual. Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.
- Newton, J. H., & McGrew, K. S. (2010). Introduction to the special issue: Current research in Cattell-Horn-Carroll–based assessment. *Psychology in the Schools*, *47*(7), 621–634.
- Pauker, S. G., & Kassirer, J. P. (1980). The threshold approach to clinical decision making. *New England Journal of Medicine*, *302*(20), 1109–1117.

- Reynolds, M. R., Keith, T. Z., Fine, J. G., Fisher, M. E., & Low, J. A. (2007). Confirmatory factor structure of the Kaufman Assessment Battery for Children–Second Edition: Consistency with Cattell-Horn-Carroll theory. *School Psychology Quarterly*, 22(4), 511.
- Speece, D. L., & Shekitka, L. (2002). How should reading disabilities be operationalized? a survey of experts. *Learning Disabilities Research & Practice*, 17(2), 118–123.
- Sternberg, R. J. (2012). Intelligence. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(5), 501–511.
- Stuebing, K. K., Fletcher, J. M., Branum-Martin, L., & Francis, D. J. (2012). Evaluation of the technical adequacy of three methods for identifying specific learning disabilities based on cognitive discrepancies. *School Psychology Review*, 41(1), 3–22.
- United States Census (2016, September 9). U.S. and World Population Clock. Retrieved from <http://www.census.gov/popclock/>
- Woodcock, R. W., Johnson, M. B., & Mather, N. (1990). *Woodcock-Johnson psycho-educational battery–revised*. DLM Teaching Resources.

APPENDIX
SIMULATION CODE

```
# Required packages
library(purrr)
library(psych)
library(magrittr)
library(dplyr)
library(tidyr)
library(tibble)
library(lavaan)
library(broom)
library(caret)
library(dtplyr)
library(data.table)
library(ggrepel)
library(haven)
library(pander)
library(mvtnorm)
library(e1071)
library(robustHD)
library(caret)

# How many subtests per broad factor
kGx <- 10
# Sample Size
n <- 1000000

# Get CHC Loadings for subtests
d <- read.csv("chc_loadings_csv.csv")

# Cognitive Factors
cog <- c('Gc', 'Gs', 'Gwm', 'Glr', 'Gv', 'Ga', 'Gf')
dc <- filter(d, Ability %in% cog) %>% mutate(Ability = factor(Ability,
  levels = cog))

# Broad Ability loading parameters (mean and sd)
Gx1 <- read.csv("ConstructLoading.csv") %>%
  mutate(z1 = fisherz>Loading)) %>%
  select(Ability, z1) %>%
  group_by(Ability) %>%
```

```

summarize(zl_mean = mean(zl), zl_sd = sd(zl)) %>%
mutate(l_mean = fisherz2r(zl_mean))

# Function for generating loadings
fr2z <- function(l_mean, l_sd, n = 1) {
  k <- length(l_mean)
  fisherz2r(rnorm(n * k, fisherz(l_mean), l_sd))
}

# Create latent scores
# g to broad ability loadings
Glr <- dc$G_AVG
names(Glr) <- dc$Ability

# Function to make Latent Scores
Gx <- function(Glr, n = 1){
  k <- length(Glr)
  G <- matrix(rep(rnorm(n),k), ncol = k)
  e <- matrix(rnorm(n * k),ncol = k)
  e1 <- sqrt(1 - Glr ^ 2)
  Gxx <- G %**% diag(Glr) + e %**% diag(e1)
  as_data_frame(cbind(G[,1],Gxx) %>% set_colnames(c("G",names(Glr))))
}

# Make g and Broad Ability latent scores
dConstructScores <- Gx(Glr, n) %>% mutate(id = 1:n)

# Cog Subtests
# Every combination of id, Ability, and i
longdata <- data.table::CJ(id = 1:n,
                          Ability = factor(dc$Ability),
                          i = 1:kGx)

# Set key for Gx1 (Subtest loading parameters)
setDT(Gx1,key = "Ability")

# Set key for longdata
setkey(longdata,Ability)

# Merge Gx1 and longdata by Broad Ability
longdata <- longdata[Gx1]

# Long to wide restructuring of construction scores
gatherdConstructScores <- dConstructScores %>%
  select(-G) %>%
  gather(key = "Ability",value = "Gx",-id) %>%
  mutate(Ability = factor(Ability,cog))

```

```

# Merge construct scores into longdata
setkey(longdata,id,Ability)
setDT(gatherdConstructScores,key = c("id","Ability"))
gatherdConstructScores <- as.data.table(gatherdConstructScores)
longdata <- longdata[gatherdConstructScores]
rm(gatherdConstructScores)

# Generate random loadings
longdata[,l := fisherz2r(rnorm(n * kGx * 7,zl_mean,zl_sd))]

# Plot loading distribution by ability
if (n < 100001) ggplot(longdata ,aes(Ability,l)) + geom_violin(colour =
  NA, fill = rgb(1,1,1,.8)) + ylab("Loadings") + ylim(0,1) +
  geom_point(data = Gx1 %>% rename(l = l_mean),colour = "gray40")

# Subtest Errors
longdata[,ze := rnorm(n * kGx)]

# Subtest Scores
longdata[,Tx := l * Gx + sqrt(1 - l ^ 2) * ze]

# Winsorize to 4 SD and round scores to 1/15 of an SD
longdata <- mutate(longdata, Tx = robustHD::winsorize(round(Tx * 15, 0)
  / 15,standardized = TRUE, const = 4))

# Make widedata
widedata <- longdata[,.(id, Ability, i, Tx)]
widedata <- dcast(widedata,id + Ability ~ i, value.var = "Tx")
colnames(widedata) <- c("id","Ability",paste0("x",1:kGx))
wideloading <- longdata %>% as_data_frame() %>%
  select(id, Ability, i, l)
setDT(wideloading)
wideloading <- dcast(wideloading, id + Ability ~ i, value.var = "l")
wideloading <- select(wideloading, -id,-Ability)
colnames(wideloading) <- paste0("l",colnames(wideloading))

# Generate XBA Composites
widedata <- widedata %>%
  bind_cols(wideloading) %>%
  mutate(
    dif12 = abs(x1 - x2),
    Above120 = (x1 > (5 / 3)) & (x2 > (5 / 3)),
    Below80 = (x1 < (-5 / 3)) & (x2 > (-5 / 3)),
    Cohesive2 = (dif12 < 1) | Above120 | Below80
  )

# Restructure data (2 subtest XBA composites)
rm(wideloading)

```

```

widedata2 <- widenedata %>%
  filter(Cohesive2) %>%
  mutate(nTests = 2,
         Tx = (x1 + x2) / sqrt(2 + 2 * l1 * l2))

# Restructure data (3 subtest XBA composites)
widedata3plus <- widenedata %>%
  filter(!Cohesive2) %>%
  mutate(
    dif13 = abs(x1 - x3),
    dif23 = abs(x2 - x3),
    eq3 = dif13 == dif23,
    Cohesive3 = (dif13 < 1 | dif23 < 1))
widedata3 <- widenedata3plus %>%
  filter(Cohesive3) %>%
  mutate(nTests = 3,
         Tx13 = (x1 + x3) / sqrt(2 + 2 * l1 * l3),
         Tx23 = (x2 + x3) / sqrt(2 + 2 * l2 * l3),
         Higher13 = Tx13 > Tx23,
         Tx = if_else(eq3,
                      if_else(Tx13 > Tx23, Tx13, Tx23),
                      if_else(dif13 < dif23, Tx13, Tx23)))

# Restructure data (4 subtest XBA composites)
setDT(widedata3plus)
widedata4plus <- widenedata3plus[Cohesive3 == F]
widedata4plus[, ':='(dif14 = abs(x1 - x4), dif24 = abs(x2 - x4), dif34
                    = abs(x3 - x4))]
widedata4plus[, minDif := pmin(dif14, dif24, dif34)]
widedata4plus[, Cohesive4 := (minDif < 1)]
widedata4 <- widenedata4plus %>%
  filter(Cohesive4) %>%
  mutate(nTests = 4,
         Tx14 = (x1 + x4) / sqrt(2 + 2 * l1 * l4),
         Tx24 = (x2 + x4) / sqrt(2 + 2 * l2 * l4),
         Tx34 = (x3 + x4) / sqrt(2 + 2 * l3 * l4)) %>%
  rowwise() %>%
  mutate(imin = list((1:3)[minDif == c(dif14, dif24, dif34)]),
         Tx = max(c(Tx14, Tx24, Tx34)[imin]))

# Restructure data (5 subtest XBA composites)
widedata5plus <- widenedata4plus %>%
  filter(!Cohesive4) %>%
  mutate(
    dif15 = abs(x1 - x5),
    dif25 = abs(x2 - x5),
    dif35 = abs(x3 - x5),
    dif45 = abs(x4 - x5))

```

```

if (nrow(widedata5plus) > 0) {
  widedata5plus <- widedata5plus %>%
    rowwise() %>%
    mutate(minDif = min(dif15,dif25,dif35,dif45)) %>%
    ungroup() %>%
    mutate(Cohesive5 = (minDif < 1))
}
widedata5 <- widedata5plus %>%
  filter(Cohesive5) %>%
  mutate(nTests = 5,
         Tx15 = (x1 + x5) / sqrt(2 + 2 * 11 * 15),
         Tx25 = (x2 + x5) / sqrt(2 + 2 * 12 * 15),
         Tx35 = (x3 + x5) / sqrt(2 + 2 * 13 * 15),
         Tx45 = (x4 + x5) / sqrt(2 + 2 * 14 * 15))
if (nrow(widedata5) > 0) {
  widedata5 <- widedata5 %>%
    rowwise() %>%
    mutate(imin = list((1:4)[minDif == c(dif15,dif25,dif35,dif45)]),
           Tx = max(c(Tx15,Tx25,Tx35,Tx45)[imin]))
} else {
  widedata5$Tx <- double(0)
}

# Restructure data (6 subtest XBA composites)
widedata6plus <- widedata5plus %>%
  filter(!Cohesive5) %>%
  mutate(
    dif16 = abs(x1 - x6),
    dif26 = abs(x2 - x6),
    dif36 = abs(x3 - x6),
    dif46 = abs(x4 - x6),
    dif56 = abs(x5 - x6))
if (nrow(widedata6plus) > 0) {
  widedata6plus <- widedata6plus %>%
    rowwise() %>%
    mutate(minDif = min(dif16,dif26,dif36,dif46,dif56)) %>%
    ungroup() %>%
    mutate(Cohesive6 = (minDif < 1))
}
widedata6 <- widedata6plus %>%
  filter(Cohesive6) %>%
  mutate(nTests = 6,
         Tx16 = (x1 + x6) / sqrt(2 + 2 * 11 * 16),
         Tx26 = (x2 + x6) / sqrt(2 + 2 * 12 * 16),
         Tx36 = (x3 + x6) / sqrt(2 + 2 * 13 * 16),
         Tx46 = (x4 + x6) / sqrt(2 + 2 * 14 * 16),
         Tx56 = (x5 + x6) / sqrt(2 + 2 * 15 * 16))
if (nrow(widedata6) > 0) {

```

```

widedata6 <- widedata6 %>%
  rowwise() %>%
  mutate(imin = list((1:5)[minDif ==
    c(dif16,dif26,dif36,dif46,dif56)]),
    Tx = max(c(Tx16,Tx26,Tx36,Tx46,Tx56)[imin]))
} else {
  widedata6$Tx <- double(0)
}

# Restructure data (7 subtest XBA composites)
widedata7plus <- widedata6plus %>%
  filter(!Cohesive6) %>%
  mutate(
    dif17 = abs(x1 - x7),
    dif27 = abs(x2 - x7),
    dif37 = abs(x3 - x7),
    dif47 = abs(x4 - x7),
    dif57 = abs(x5 - x7),
    dif67 = abs(x6 - x7))
if (nrow(widedata7plus) > 0 ) {
  widedata7plus <- widedata7plus %>%
    rowwise() %>%
    mutate(minDif = min(dif17,dif27,dif37,dif47,dif57,dif67)) %>%
    ungroup() %>%
    mutate(Cohesive7 = (minDif < 1))
}
widedata7 <- widedata7plus %>%
  filter(Cohesive7) %>%
  mutate(nTests = 7,
    Tx17 = (x1 + x7) / sqrt(2 + 2 * 11 * 17),
    Tx27 = (x2 + x7) / sqrt(2 + 2 * 12 * 17),
    Tx37 = (x3 + x7) / sqrt(2 + 2 * 13 * 17),
    Tx47 = (x4 + x7) / sqrt(2 + 2 * 14 * 17),
    Tx57 = (x5 + x7) / sqrt(2 + 2 * 15 * 17),
    Tx67 = (x6 + x7) / sqrt(2 + 2 * 16 * 17))
if (nrow(widedata7) > 0) {
  widedata7 <- widedata7 %>%
    rowwise() %>%
    mutate(imin = list((1:5)[minDif ==
      c(dif17,dif27,dif37,dif47,dif57,dif67)]),
      Tx = max(c(Tx17,Tx27,Tx37,Tx47,Tx57,Tx67)[imin]))
} else {
  widedata7$Tx <- double(0)
}

```



```

# Merge XBA composites in one data.frame
XBA <-
  bind_rows(widedata2,widedata3,widedata4,widedata5,widedata6,wided
    ata7) %>% select(id:dif12,nTests,Tx) %>% select(-dif12)
rm(widedata2,widedata3,widedata4,widedata5,widedata6,widedata7,widedata
  3plus,widedata4plus,widedata5plus,widedata6plus,widedata7plus)
setDT(XBA, key = c("id","Ability"))
setorder(XBA, id, Ability)
setDT(dConstructScores, key = "id")

# Merge latent scores with XBA
XBA <- XBA %>% rename(TxXBA = Tx)
XBA <- XBA[setkeyv(melt(dConstructScores,id.vars = "id",variable.name =
  "Ability",value.name = "Gx")[Ability != "G"],cols =
  c("id","Ability"))]
wideXBA <- XBA %>% select(id,Ability,TxXBA) %>%
  mutate(Prefix = "Tx") %>%
  unite(Ability,Prefix,Ability) %>%
  spread(key = Ability, value = TxXBA)
widedata <- longdata %>%
  rename(Test = Ability) %>%
  select(id, Test, Tx, i) %>%
  dcast(id ~ Test + i ,value.var = "Tx")

# Read WJIV Norming data
d <- haven::read_spss("W4 Norming_Schneider 051815.sav") %>%
  select(SubjAgeMos,contains("_SS"))

# Specify age of cases
d$AgeYear <- d$SubjAgeMos / 12
d$Age <- d$AgeYear %>% cut(c(0,4,6,8,10,14,18,40,60,100))
AgeCat <- d$Age %>% unique
library(lavaan)

# Set as workable data frame
d <- as.data.frame((as.matrix(d %>% select(-AgeYear,-Age,-SubjAgeMos))
  - 100) / 15)

# Structural model of Cognitive and Academic abilities
m <- '
Ga =~ PHNPRO_SS + SEGMNT_SS + SNDBLN_SS
Gc =~ VOCAB_SS + ACKNOW_SS + ORLCMP_SS
Gs =~ PERSPD_SS + PAIRCN_SS
Glr =~ VAL_SS + STYREC_SS
Gwm =~ GWM3_SS + AUDMS_SS
WRDATK =~ 0.918 * WRDATK_SS
LWIDNT =~ 0.933 * LWIDNT_SS
RDGFLU =~ 0.940 * RDGFLU_SS

```

```

PSGCMP =~ 0.880 * PSGCMP_SS
WRDATK ~ Ga + Gc + Gwm
LWIDNT ~ WRDATK + Gc
RDGFLU ~ LWIDNT + Gs + Glr
PSGCMP ~ LWIDNT + RDGFLU + Gc
'

library(semPlot)
semfit <- lavaan::sem(m,d)
mstat <- lavInspect(semfit,"std")
Beta <- mstat$beta %>% as.data.frame()
psi <- mstat$psi
exonames <- c('Ga', 'Gc', 'Gwm', 'Gs', 'Glr')
endonames <- c('WRDATK', 'LWIDNT', 'RDGFLU', 'PSGCMP')

# Specify matrix properties
x <- select(dConstructScores,Ga,Gc,Gwm,Gs,Glr) %>% as.matrix()
Bet <- Beta[endonames,endonames] %>% as.matrix()
Gam <- Beta[endonames,exonames] %>% as.matrix()
library(mvtnorm)
zeta <- rmvnorm(n,sigma = psi[endonames,endonames])
colnames(zeta) <- endonames
Y = t(solve(diag(ncol(Bet)) - Bet) %**% (Gam %**% t(x) + t(zeta)))

# Restructure and select data
rawd <- data.frame(x,Y)
colnames(rawd) <- c(exonames,endonames)

# Academic abilities structural model
m <- '
WRDATK_SS ~ Ga + Gc + Gwm
LWIDNT_SS ~ WRDATK_SS + Gc
RDGFLU_SS ~ LWIDNT_SS + Gs + Glr
PSGCMP_SS ~ LWIDNT_SS + RDGFLU_SS + Gc
'

# Create XBA dataset adding construct scores for cognitive and academic
abilities
All_d <- dConstructScores %>% cbind(rawd %>% select(WRDATK:PSGCMP)) %>%
as_data_frame()
All_d <- All_d %>% left_join(wideXBA, by = "id")

# Add variation by Reliability coefficient
All_d$Tx_WRDATK <- 0.918 * All_d$WRDATK + sqrt(1 - 0.918 ^ 2) *
rnorm(n)
All_d$Tx_LWIDNT <- 0.933 * All_d$LWIDNT + sqrt(1 - 0.933 ^ 2) *
rnorm(n)
All_d$Tx_RDGFLU <- 0.940 * All_d$RDGFLU + sqrt(1 - 0.940 ^ 2) *
rnorm(n)

```

```

All_d$Tx_PSGCMP <- 0.880 * All_d$PSGCMP + sqrt(1 - 0.880 ^ 2) *
  rnorm(n)
filterw <- c(1, 1, 1, 1, 0, 0, 0)
w <- c(.1029,.2355,.1870,.1572,.0864,.1167,.1152)
fStrength <- function(x) x > -2/3
Strength <- apply(All_d %>% select(Gc:Gf),2,fStrength)
GAE <- Strength %**% w
Lw <- filterw * 0.25
LAE <- ((Strength %**% Lw) + (Strength %**% (w * filterw)) / sum(w *
  filterw)) / 2
All_d$g_Prob_Latent <- (LAE + GAE) / 2
rm(GAE,LAE,Strength,w,filterw,Lw)

# LD Latent
weakness <- function(x,threshold = (-2 / 3)) x < threshold

# Latent LEVEL I Identification
All_d$WRDATKDeficit <- weakness(All_d$WRDATK)
All_d$LWIDNTDeficit <- weakness(All_d$LWIDNT)
All_d$RDGFLUDeficit <- weakness(All_d$RDGFLU)
All_d$PSGCMPDeficit <- weakness(All_d$PSGCMP)

# Latent LEVEL III Identification
All_d$Gc_Deficit <- weakness(All_d$Gc)
All_d$Ga_Deficit <- weakness(All_d$Ga)
All_d$Gwm_Deficit <- weakness(All_d$Gwm)
All_d$Glr_Deficit <- weakness(All_d$Glr)
All_d$Gs_Deficit <- weakness(All_d$Gs)
All_d$Gf_Deficit <- weakness(All_d$Gf)
All_d$Gv_Deficit <- weakness(All_d$Gv)

# Latent ID Identification
All_d$G_NotCogDeficit_Latent <- All_d$g_Prob_Latent > .50

# Latent IF WRDATKDeficit & Ga OR Gc OR Gwm
All_d$WRDATK_CogDeficit <- ((All_d$Gc_Deficit) | (All_d$Ga_Deficit) |
  (All_d$Gwm_Deficit))

# Latent IF LWIDNTDeficit & Gc
All_d$LWIDNT_CogDeficit <- (All_d$Gc_Deficit | All_d$Ga_Deficit |
  All_d$Glr_Deficit | All_d$Gwm_Deficit)

# Latent IF RDGFLUDeficit & Gs OR Glr
All_d$RDGFLU_CogDeficit <- (All_d$Gs_Deficit | All_d$Glr_Deficit)

# Latent IF PSGCMPDeficit & Gc
All_d$PSGCMP_CogDeficit <- (All_d$Gc_Deficit | All_d$Gf_Deficit |
  All_d$Glr_Deficit | All_d$Gs_Deficit | All_d$Gwm_Deficit)

```

```

# Latent LEVEL IV Identification
All_d$AcademicDeficit <- (All_d$WRDATKDeficit | All_d$LWIDNTDeficit |
  All_d$RDGFLUDeficit | All_d$PSGCMPDeficit) &
  (All_d$G_NotCogDeficit_Latent)

# Critical z
zcrit <- qnorm(.1)

# Standardized residual lower than zcrit for Word Attack academic test
mWRDATK <- lm(WRDATK ~ G, All_d)
All_d$WRDATKUnexpectedLow <- rstandard(mWRDATK) <= zcrit

# Standardized residual lower than zcrit for Letter Word Identification
academic test
mLWIDNT <- lm(LWIDNT ~ G, All_d)
All_d$LWIDNTUnexpectedLow <- rstandard(mLWIDNT) <= zcrit

# Standardized residual lower than zcrit for Reading Fluency academic
test
mRDGFLU <- lm(RDGFLU ~ G, All_d)
All_d$RDGFLUUnexpectedLow <- rstandard(mRDGFLU) <= zcrit

# Standardized residual lower than zcrit for Passage Completion
academic test
mPSGCMP <- lm(PSGCMP ~ G, All_d)
All_d$PSGCMPUnexpectedLow <- rstandard(mPSGCMP) <= zcrit

# Restructure data
All_d <- All_d %>% rowwise %>% mutate(
  LDWRDATK_Latent =
    all(WRDATKDeficit,G_NotCogDeficit_Latent,WRDATK_CogDeficit,WRDATK
      UnexpectedLow),
  LDLWIDNT_Latent =
    all(LWIDNTDeficit,G_NotCogDeficit_Latent,LWIDNT_CogDeficit,LWIDNT
      UnexpectedLow),
  LDRDGFLU_Latent =
    all(RDGFLUDeficit,G_NotCogDeficit_Latent,RDGFLU_CogDeficit,RDGFLU
      UnexpectedLow),
  LDPSGCMP_Latent =
    all(PSGCMPDeficit,G_NotCogDeficit_Latent,PSGCMP_CogDeficit,PSGCMP
      UnexpectedLow)
)
All_d$SLD_Latent <- (All_d$LDPSGCMP_Latent | All_d$LDRDGFLU_Latent |
  All_d$LDLWIDNT_Latent | All_d$LDWRDATK_Latent)

##G Probability for XBA
d_XBA <- All_d %>% select(id, G, contains("Tx_"))

```

```

# XBA G composite loadings
filterw <- c(0, 1, 0, 1, 1, 0, 1)
w <- c(.1029,.2355,.1870,.1572,.0864,.1167,.1152)

# XBAIs cognitive ability test score a strength or weakness?
Strength <- (d_XBA %>% select(Tx_Ga:Tx_Gwm)) > -2/3
GAE <- Strength %**% w
Lw <- filterw * 0.25
LAE <- ((Strength %**% Lw) + (Strength %**% (w * filterw)) / sum(w *
  filterw)) / 2
g_prob_observed <- (LAE + GAE) / 2
isSufficient <- function(x) ifelse(x > -2 / 3,1,0)
d_Sufficient <- d_XBA %>% select(Tx_Ga:Tx_Gwm) %>%
  mutate_each(funs(isSufficient)) %>% as.matrix
nSufficient <- rowSums(d_Sufficient) > 3

# Restructure as matrix
d_G <- d_XBA %>% select(Tx_Ga:Tx_Gwm) %>% as.matrix
sd_G <- ((d_G %>% rowSums) / 7) %>% sd
d_XBA$G_XBA <- rowSums(d_Sufficient * d_G) / (sd_G *
  rowSums(d_Sufficient))
d_XBA$G_XBA[!nSufficient] <- NA
# XBA Create True or False list of weakness by Academic test
d_XBA$WRDATKDeficit_XBA <- weakness(d_XBA$Tx_WRDATK)
d_XBA$LWIDNTDeficit_XBA <- weakness(d_XBA$Tx_LWIDNT)
d_XBA$RDGFLUDeficit_XBA <- weakness(d_XBA$Tx_RDGFLU)
d_XBA$PSGCMPDeficit_XBA <- weakness(d_XBA$Tx_PSGCMP)

# XBA LEVEL III Identification
# Create True or False list of weakness by cognitive ability
d_XBA$Gc_Deficit_XBA <- weakness(d_XBA$Tx_Gc)
d_XBA$Ga_Deficit_XBA <- weakness(d_XBA$Tx_Ga)
d_XBA$Gwm_Deficit_XBA <- weakness(d_XBA$Tx_Gwm)
d_XBA$Glr_Deficit_XBA <- weakness(d_XBA$Tx_Glr)
d_XBA$Gs_Deficit_XBA <- weakness(d_XBA$Tx_Gs)
d_XBA$Gf_Deficit_XBA <- weakness(d_XBA$Tx_Gf)
d_XBA$Gv_Deficit_XBA <- weakness(d_XBA$Tx_Gv)

# XBA ID Identification
d_XBA$G_NotCogDeficit_XBA <- g_prob_observed > .50

# XBA IF WRDATKDeficit & Ga OR Gc OR Gwm
d_XBA$WRDATK_CogDeficit_XBA <- ((d_XBA$Gc_Deficit_XBA) |
  (d_XBA$Ga_Deficit_XBA) | (d_XBA$Gwm_Deficit_XBA))

```

```

# XBA IF LWIDNTDeficit & Gc
d_XBA$LWIDNT_CogDeficit_XBA <- (d_XBA$Gc_Deficit_XBA |
  d_XBA$Ga_Deficit_XBA | d_XBA$Glr_Deficit_XBA |
  d_XBA$Gwm_Deficit_XBA)

# XBA IF RDGFLUDeficit & Gs OR Glr
d_XBA$RDGFLU_CogDeficit_XBA <- (d_XBA$Gs_Deficit_XBA |
  d_XBA$Glr_Deficit_XBA)

# XBA IF PSGCMPDeficit & Gc
d_XBA$PSGCMPCogDeficit_XBA <- (d_XBA$Gc_Deficit_XBA |
  d_XBA$Gf_Deficit_XBA | d_XBA$Glr_Deficit_XBA |
  d_XBA$Gs_Deficit_XBA | d_XBA$Gwm_Deficit_XBA)

# XBA LEVEL IV Identification
d_XBA$AcademicDeficit_XBA <- (d_XBA$WRDATKDeficit_XBA |
  d_XBA$LWIDNTDeficit_XBA | d_XBA$RDGFLUDeficit_XBA |
  d_XBA$PSGCMPCogDeficit_XBA) & (d_XBA$G_NotCogDeficit_XBA == FALSE)

# XBA Standardized residual lower than zcrit for word attack academic
  test
mWRDATK <- lm(Tx_WRDATK ~ G_XBA, d_XBA)
ze <- rstandard(mWRDATK)
low <- data.frame(id = as.numeric(names(ze)),WRDATKUnexpectedLow_XBA =
  (ze <= zcrit))
setDT(low,key = "id")
setDT(d_XBA,key = "id")
d_XBA <- merge(d_XBA,low,all.x = TRUE, all.y = FALSE)
d_XBA$WRDATKUnexpectedLow_XBA[is.na(d_XBA$WRDATKUnexpectedLow_XBA)] <-
  FALSE

# XBA Standardized residual lower than zcrit for letter word
  identification academic test
mLWIDNT <- lm(Tx_LWIDNT ~ G_XBA, d_XBA)
ze <- rstandard(mLWIDNT)
low <- data.frame(id = as.numeric(names(ze)),LWIDNTUnexpectedLow_XBA =
  (ze <= zcrit))
setDT(low,key = "id")
setDT(d_XBA,key = "id")
d_XBA <- merge(d_XBA,low,all.x = TRUE, all.y = FALSE)
d_XBA$LWIDNTUnexpectedLow_XBA[is.na(d_XBA$LWIDNTUnexpectedLow_XBA)] <-
  FALSE

# XBA Standardized residual lower than zcrit for reading fluency
  academic test
mRDGFLU <- lm(Tx_RDGFLU ~ G_XBA, d_XBA)
ze <- rstandard(mRDGFLU)

```

```

low <- data.frame(id = as.numeric(names(ze)),RDGFLUUnexpectedLow_XBA =
  (ze <= zcrit))
setDT(low,key = "id")
setDT(d_XBA,key = "id")
d_XBA <- merge(d_XBA,low,all.x = TRUE, all.y = FALSE)
d_XBA$RDGFLUUnexpectedLow_XBA[is.na(d_XBA$RDGFLUUnexpectedLow_XBA)] <-
  FALSE

# XBA Standardized residual lower than zcrit for passage comprehension
academic test
mPSGCMP <- lm(Tx_PSGCMP~ G_XBA, d_XBA)
ze <- rstandard(mPSGCMP)
low <- data.frame(id = as.numeric(names(ze)),PSGCMPUnexpectedLow_XBA =
  (ze <= zcrit))
setDT(low,key = "id")
setDT(d_XBA,key = "id")
d_XBA <- merge(d_XBA,low,all.x = TRUE, all.y = FALSE)
d_XBA$PSGCMPUnexpectedLow_XBA[is.na(d_XBA$PSGCMPUnexpectedLow_XBA)] <-
  FALSE

# XBA LD (TRUE or FALSE)
d_XBA <- d_XBA %>% rowwise %>% mutate(
  LDWRDATK_XBA = all(WRDATKDeficit_XBA,G_NotCogDeficit_XBA
    ,WRDATK_CogDeficit_XBA,WRDATKUnexpectedLow_XBA),
  LDLWIDNT_XBA =
    all(LWIDNTDeficit_XBA,G_NotCogDeficit_XBA,LWIDNT_CogDeficit_XBA,L
      WIDNTUnexpectedLow_XBA),
  LDRDGFLU_XBA =
    all(RDGFLUDeficit_XBA,G_NotCogDeficit_XBA,RDGFLU_CogDeficit_XBA,R
      DGFLUUnexpectedLow_XBA),
  LDPSGCMP_XBA =
    all(PSGCMPDeficit_XBA,G_NotCogDeficit_XBA,PSGCMP_CogDeficit_XBA,P
      SGCMPUnexpectedLow_XBA)
)

# XBA Any SLD Diagnosis
All_d$SLD_XBA <- (d_XBA$LDPSGCMP_XBA | d_XBA$LDRDGFLU_XBA |
  d_XBA$LDLWIDNT_XBA | d_XBA$LDWRDATK_XBA)

# XBA Diagnostic Accuracy
ctable <- with(All_d ,
  table( SLD_XBA, SLD_Latent ))
confuseM <- confusionMatrix(ctable,positive = "TRUE")
confuseM

# XBA Diagnostic accuracy statistics
Sens <- confuseM$byClass["Sensitivity"]
Spec <- confuseM$byClass["Specificity"]

```

```

PPV_overall <- confuseM$byClass["Pos Pred Value"]
NPV_overall <- confuseM$byClass["Neg Pred Value"]

# Regular Composites
## Regular Cognitive Composites
### Slow but hyper-correct method
# Create Composite SD
dCSD <- longdata %>%
  mutate(li = paste0("l_",i)) %>%
  select(id,Ability,li,l) %>%
  spread(key = li, value = l) %>%
  unite(Abilityid,Ability,id)
CompositeSD <- function(l) {
  l <- as.matrix(l)
  CSD <- numeric(kGx - 1)
  for (i in 2:kGx) {
    r <- l[1:i] %*% t(l[1:i])
    diag(r) <- 1
    CSD[i - 1] <- sqrt(sum(r))
  }
  matrix(CSD,ncol = 1)
}
dCompositeSD <- dCSD %>% select(-Abilityid) %>% apply(1,CompositeSD)
  %>% t %>% set_colnames(paste0("CSD_",2:kGx))

# Create Composite Sums
dCSUM <- longdata %>%
  mutate(li = paste0("T_",i)) %>%
  select(id,Ability,li,Tx) %>%
  spread(key = li, value = Tx) %>%
  unite(Abilityid,Ability,id)
CompositeSUM <- function(Tx) {
  Tx <- as.matrix(Tx)
  CSUM <- numeric(kGx - 1)
  for (i in 2:kGx) {
    CSUM[i - 1] <- sum(Tx[1:i])
  }
  matrix(CSUM,ncol = 1)
}
dCompositeSUM <- dCSUM %>% select(-Abilityid) %>% apply(1,CompositeSUM)
  %>% t %>% set_colnames(paste0("C_",2:kGx))
Cz <- as_data_frame(dCompositeSUM / dCompositeSD)

# Composite Scores
CompositeScores <- data_frame(Abilityid = dCSUM$Abilityid) %>%
  cbind(Cz) %>%
  separate(Abilityid,c("Ability","id")) %>%
  gather("k","v",-Ability,-id) %>%

```



```

unite(test,Ability,k) %>%
spread(test,v) %>%
mutate(id = as.numeric(id)) %>%
setDT(key = "id")
dCS <- dConstructScores %>% select(id,G:Gf) %>% setDT(key = "id")
dCS <- dCS[widedata][CompositeScores]

# Composite IQ
# Function for calculating the sums of IQ
fIQSum <- function(k) scale(longdata[i < k + 1,.(S = sum(Tx)),id]$S)
ObservedIQ <- as.data.frame((sapply(1:kGx,fIQSum)))
colnames(ObservedIQ) <- paste0("IQ_",1:kGx)
ObservedIQ$id <- 1:n
setDT(ObservedIQ, key = "id")
dCS <- dCS[ObservedIQ]

## Validity Coefficients
ValidityCoefficient <- function(lc) cor(cbind(dCS %>% select_(lc) , dCS
      %>% select(contains(paste0(lc,"_C_")))))[1,-1]
valCoef <- sapply(cog,ValidityCoefficient)[paste0(cog[1],"_C_",2:kGx),]
      ^ 2
rownames(valCoef) <- 2:kGx
valCoef <- as_tibble(valCoef)
valCoef$k <- 2:kGx
valCoef <- valCoef %>% select(k,Gc:Gf)
dvalCoef <- valCoef %>% gather(k = "Gx",value = "Validity",-k)
valCoefLabel <- dvalCoef %>% filter(k == 2)
dvalCoef %>% ggplot(aes(k,Validity,color = Gx)) + geom_line() +
      geom_point() + geom_label_repel(data = valCoefLabel,aes(x = k,y =
      Validity, label = Gx)) + theme(legend.position = "none") +
      xlab("Number of Tests") + scale_x_continuous(limits = c(2, kGx
      ),breaks = 2:kGx)
dvalCoef$Validity <- formatC(dvalCoef$Validity, 2, format = "f") %>%
      gsub(pattern = "0.", replacement = ".")
dvalCoef %>% spread(key = "Gx", value = "Validity") %>%
      knitr::kable(caption = "Validity coefficients by subtests")

## LD Regular Composites (CS), Non SLD forced repeated testing method
dCS$WRDATKDeficit_CS <- All_d$WRDATKDeficit
dCS$LWIDNTDeficit_CS <- All_d$LWIDNTDeficit
dCS$RDGFLUDeficit_CS <- All_d$RDGFLUDeficit
dCS$PSGCMPDeficit_CS <- All_d$PSGCMPDeficit
dCS$Tx_WRDATK <- All_d$Tx_WRDATK
dCS$Tx_LWIDNT <- All_d$Tx_LWIDNT
dCS$Tx_RDGFLU <- All_d$Tx_RDGFLU
dCS$Tx_PSGCMP <- All_d$Tx_PSGCMP
dCS$SLD_Latent <- All_d$SLD_Latent
dCS$Tx_WRDATKDeficit_CS <- weakness(dCS$Tx_WRDATK)

```

```

dCS$Tx_LWIDNTDeficit_CS <- weakness(dCS$Tx_LWIDNT)
dCS$Tx_RDGFLUDeficit_CS <- weakness(dCS$Tx_RDGFLU)
dCS$Tx_PSGCMPDeficit_CS <- weakness(dCS$Tx_PSGCMP)
for (kCS in 2:kGx) {
  for (a in c("Gc","Ga","Gwm","Glr","Gs","Gs","Gf","Gv")) {
    dCS[,paste0(a,"_Deficit_CS_",kCS)] <- weakness(dCS %>%
      select_(paste0(a,"_C_",kCS)))
  }

  # ID Identification
  dCS[,paste0("G_NotCogDeficit_CS_",kCS)] <- dCS %>%
    select_(paste0("IQ_",kCS)) > -2/3

  # IF WRDATKDeficit & Ga OR Gc OR Gwm
  dCS[,paste0("WRDATK_CogDeficit_CS_",kCS)] <-
    (dCS[,paste0("Gc_Deficit_CS_",kCS), with = F] |
     dCS[,paste0("Ga_Deficit_CS_",kCS), with = F] |
     dCS[,paste0("Gwm_Deficit_CS_",kCS), with = F])

  # IF LWIDNTDeficit & Gc
  dCS[,paste0("LWIDNT_CogDeficit_CS_",kCS)] <-
    (dCS[,paste0("Gc_Deficit_CS_",kCS), with = F] |
     dCS[,paste0("Ga_Deficit_CS_",kCS), with = F] |
     dCS[,paste0("Glr_Deficit_CS_",kCS), with = F] |
     dCS[,paste0("Gwm_Deficit_CS_",kCS), with = F])

  # IF RDGFLUDeficit & Gs OR Glr
  dCS[,paste0("RDGFLU_CogDeficit_CS_",kCS)] <-
    (dCS[,paste0("Gs_Deficit_CS_",kCS), with = F] |
     dCS[,paste0("Glr_Deficit_CS_",kCS), with = F])

  # IF PSGCMPDeficit & Gc
  dCS[,paste0("PSGCMPCogDeficit_CS_",kCS)] <-
    (dCS[,paste0("Gc_Deficit_CS_",kCS), with = F] |
     dCS[,paste0("Gf_Deficit_CS_",kCS), with = F] |
     dCS[,paste0("Glr_Deficit_CS_",kCS), with = F] |
     dCS[,paste0("Gs_Deficit_CS_",kCS), with = F] |
     dCS[,paste0("Gwm_Deficit_CS_",kCS), with = F])

  # LEVEL IV Identification
  dCS[,paste0("AcademicDeficit_",kCS)] <-
    (dCS[,paste0("WRDATKDeficit_CS"), with = F] |
     dCS[,paste0("LWIDNTDeficit_CS"), with = F] |
     dCS[,paste0("RDGFLUDeficit_CS"), with = F] |
     dCS[,paste0("PSGCMPCDeficit_CS"), with = F]) &
    (dCS[,paste0("G_NotCogDeficit_CS_",kCS), with = F] == FALSE)

```

```

# Standardized residual lower than zcrit for academic test Word Attack?
mWRDATK <- lm(as.formula(paste0("Tx_WRDATK ~ IQ_",kCS)), dCS)
dCS[,paste0("WRDATKUnexpectedLow_CS_",kCS)] <- rstandard(mWRDATK) <=
  zcrit

# Standardized residual lower than zcrit for academic test Letter Word Identification?
mLWIDNT <- lm(as.formula(paste0("Tx_LWIDNT ~ IQ_",kCS)), dCS)
dCS[,paste0("LWIDNTUnexpectedLow_CS_",kCS)] <- rstandard(mLWIDNT) <=
  zcrit

# Standardized residual lower than zcrit for academic test reading fluency?
mRDGFLU <- lm(as.formula(paste0("Tx_RDGFLU ~ IQ_",kCS)), dCS)
dCS[,paste0("RDGFLUUnexpectedLow_CS_",kCS)] <- rstandard(mRDGFLU) <=
  zcrit

# Standardized residual lower than zcrit for academic test passage comprehension?
mPSGCMP <- lm(as.formula(paste0("Tx_PSGCMP ~ IQ_",kCS)), dCS)
dCS[,paste0("PSGCMPUnexpectedLow_CS_",kCS)] <- rstandard(mPSGCMP) <=
  zcrit

# Regular Composite LD (TRUE or FALSE)
for (aa in c("WRDATK", "LWIDNT", "RDGFLU", "PSGCMP")) {
  dCS[,paste0("LD",aa,"_CS_",kCS)] <- dCS[,paste0(aa,"Deficit_CS"),
    with = F] & dCS[,paste0("G_NotCogDeficit_CS_",kCS), with = F] &
  dCS[,paste0(aa,"_CogDeficit_CS_",kCS), with = F] &
  dCS[,paste0(aa,"_CogDeficit_CS_",kCS), with = F] &
  dCS[,paste0(aa,"UnexpectedLow_CS_",kCS), with = F]
}
dCS[,paste0("SLD_CS_",kCS)] <- dCS[,paste0("LDPSGCMP_CS_",kCS), with
  = F] | dCS[,paste0("LDRDGFLU_CS_",kCS), with = F] |
  dCS[,paste0("LDLWIDNT_CS_",kCS), with = F] |
  dCS[,paste0("LDWRDATK_CS_",kCS), with = F]
}

# Diagnostic Accuracy of Regular Composites
# PPV
ppv_CS <- function(i) confusionMatrix(table(dCS[,paste0("SLD_CS_",i),
  with = F][[1]],dCS[,"SLD_Latent", with = F][[1]]),positive =
  "TRUE")$byClass["Pos Pred Value"]

# Confusion matrix
confusion_CS <- function(i)
  confusionMatrix(table(dCS[,paste0("SLD_CS_",i), with =

```

```

      F[[1]],dCS[,"SLD_Latent", with = F[[1]]),positive =
      "TRUE")$byClass[1:5]
ConfusionMatrixCS <- sapply(2:kGx,confusion_CS)
colnames(ConfusionMatrixCS) <- 2:kGx
dppv_CS <- data.frame(k = 2:kGx, PPV = sapply(2:kGx,ppv_CS))
ggplot(dppv_CS, aes(k,PPV)) + geom_line() + geom_point() +
  geom_text(aes(label = gsub(formatC(PPV,2,format = "f"),pattern =
  "0.",replacement = ".")),vjust = 1.5) + xlab("Number of Subtests
  per Cognitive Composite") + ylab("Positive Predictive Value")
knitr::kable(ConfusionMatrixCS, caption = "Accuracy Statistics for
  Regular Composites")

# Accuracy of Academic Achievement Deficits
# Academic Achievement Variables
aav <- c("WRDATK","LWIDNT","RDGFLU","PSGCMP")

# Academic Achievement Variable Names
aaNames <- c("Word Attack", "Letter-Word Identification", "Ready
  Fluency", "Passage Comprehension")

# Accuracy of Diagnostic Steps (XBA)
## Accuracy of XBA Cognitive Deficits
rbind(table(All_d$Gc_Deficit, d_XBA$Gc_Deficit_XBA) %>%
  confusionMatrix(positive = "TRUE") %$% byClass[1:5],
  table(All_d$Gf_Deficit, d_XBA$Gf_Deficit_XBA) %>%
  confusionMatrix(positive = "TRUE") %$% byClass[1:5],
  table(All_d$Ga_Deficit, d_XBA$Ga_Deficit_XBA) %>%
  confusionMatrix(positive = "TRUE") %$% byClass[1:5],
  table(All_d$Gv_Deficit, d_XBA$Gv_Deficit_XBA) %>%
  confusionMatrix(positive = "TRUE") %$% byClass[1:5],
  table(All_d$Glr_Deficit, d_XBA$Glr_Deficit_XBA) %>%
  confusionMatrix(positive = "TRUE") %$% byClass[1:5],
  table(All_d$Gwm_Deficit, d_XBA$Gwm_Deficit_XBA) %>%
  confusionMatrix(positive = "TRUE") %$% byClass[1:5],
  table(All_d$Gs_Deficit, d_XBA$Gs_Deficit_XBA) %>%
  confusionMatrix(positive = "TRUE") %$% byClass[1:5]) %>%
  set_rownames(c("Gc", "Gf", "Ga", "Gv", "Glr", "Gwm", "Gs")) %>%
  knitr::kable(digits = 2, caption = "Accuracy of Cognitive Deficits
  (XBA Method)")

## Accuracy of Any Cognitive Deficit (XBA)
acdf <- function(aa) table(All_d[paste0(aa,"_CogDeficit")][[1]],
  d_XBA[,paste0(aa,"_CogDeficit_XBA")][[1]]) %>%
  confusionMatrix(positive = "TRUE") %$% byClass[1:5]
sapply(aav,acdf) %>% t %>%
  set_rownames(aaNames) %>%
  knitr::kable(digits = 2, caption = "Accuracy of Any Cognitive Deficit
  for Specific LD Diagnoses (XBA Method)")

```

```

## Accuracy of Unexpected Low Achievement (XBA Method)
ulaf <- function(aa) table(All_d[paste0(aa,"UnexpectedLow")][[1]],
  d_XBA[,paste0(aa,"UnexpectedLow_XBA")][[1]]) %>%
  confusionMatrix(positive = "TRUE") %$$ byClass[1:5]
sapply(aav,ulaf) %>% t %>%
  set_rownames(aaNames) %>%
  knitr::kable(digits = 2, caption = "Accuracy of Unexpected Low
  Achievement for Specific LD Diagnoses (XBA Method)")

# Accuracy of XBA Specific LD
aaf <- function(aa) table(All_d[paste0("LD",aa,"_Latent")][[1]],
  d_XBA[,paste0("LD",aa,"_XBA")][[1]]) %>% confusionMatrix(positive
  = "TRUE") %$$ byClass[1:5]
sapply(aav,aaf) %>% t %>%
  set_rownames(aaNames) %>%
  knitr::kable(digits = 2, caption = "Accuracy of Specific LD Diagnoses
  (XBA Method)")

# Accuracy of Diagnostitic Steps (Regular Composites)
# Accuracy of Absence of G Deficit
gf <- function(i) table(All_d$dG_NotCogDeficit_Latent,
  dCS[,paste0("G_NotCogDeficit_CS_",i),with = F][[1]]) %>%
  confusionMatrix(positive = "TRUE") %$$ byClass[1:5]
rbind(table(All_d$dG_NotCogDeficit_Latent, d_XBA$dG_NotCogDeficit_XBA)
  %>% confusionMatrix(positive = "TRUE") %$$
  byClass[1:5],sapply(2:kGx,gf) %>% t) %>%
  set_rownames(c("XBA G", paste0("IQ_", 2:kGx))) %>%
  knitr::kable(digits = 2, caption = "Accuracy of Absence of G Deficits
  (XBA Method & Regular Composites)")

# Accuracy of Cognitive Deficits (Regular Composites)
cogf <- function(cog,i) table(All_d[paste0(cog,"_Deficit")][[1]],
  dCS[,paste0(cog,"_Deficit_CS_",i),with = F][[1]]) %>%
  confusionMatrix(positive = "TRUE") %$$ byClass[1:5]
for (cog in c("Gc","Ga","Gwm","Glr","Gs","Gs","Gf","Gv"))
  print(cbind(Subtests = 2:kGx,map2(cog,2:kGx,cogf) %>%
  simplify2array() %>% t %>% as.data.frame()) %>%

  knitr::kable(digits = 2, caption = paste0("Accuracy of ",cog,"
  Deficit (Regular Composites)"))

# Accuracy of Any Cognitive Deficit for Specific LD (Regular
  Composites)
acdif <- function(aa,i) table(All_d[paste0(aa,"_CogDeficit")][[1]],
  dCS[,paste0(aa,"_CogDeficit_CS_",i),with = F][[1]]) %>%
  confusionMatrix(positive = "TRUE") %$$ byClass[1:5]
cbind(Subtests = 2:kGx,map2("WRDATK",2:kGx,acdif) %>% simplify2array()
  %>% t %>% as.data.frame()) %>%

```

```

knitr::kable(digits = 2, caption = "Accuracy of Any Cognitive Deficit
for Word Attack LD (Regular Composites)")
cbind(Subtests = 2:kGx,map2(aav[2],2:kGx,acdif) %>% simplify2array()
%>% t %>% as.data.frame()) %>%
knitr::kable(digits = 2, caption = "Accuracy of Any Cognitive Deficit
for Letter-Word Identification LD (Regular Composites)")
cbind(Subtests = 2:kGx,map2(aav[3],2:kGx,acdif) %>% simplify2array()
%>% t %>% as.data.frame()) %>%
knitr::kable(digits = 2, caption = "Accuracy of Any Cognitive Deficit
for Reading Fluency LD (Regular Composites)")
cbind(Subtests = 2:kGx,map2(aav[4],2:kGx,acdif) %>% simplify2array()
%>% t %>% as.data.frame()) %>%
knitr::kable(digits = 2, caption = "Accuracy of Any Cognitive Deficit
for Passage Comprhension LD (Regular Composites)")

# Accuracy of Unexpected Low Achievement (Regular Composites)
ulaif <- function(aa,i) table(All_d[paste0(aa,"UnexpectedLow")][[1]],
dCS[,paste0(aa,"UnexpectedLow_CS_",i),with = F][[1]]) %>%
confusionMatrix(positive = "TRUE") %$$ byClass[1:5]
for (ii in 1:4) print(cbind(Subtests = 2:kGx,map2(aav[ii],2:kGx,ulaif)
%>% simplify2array() %>% t %>% as.data.frame()) %>%
knitr::kable(digits = 2, caption =
paste0("Accuracy of Unexpected Low Achievement for
",aaNames[ii]," LD (Regular Composites)")))

# Accuracy of Specific LD Diagnoses (Regular Composites)
aaf <- function(aa,i) table(All_d[paste0("LD",aa,"_Latent")][[1]],
dCS[,paste0("LD",aa,"_CS_",i),with = F][[1]]) %>%
confusionMatrix(positive = "TRUE") %$$ byClass[1:5]
for (a in 1:4) print(cbind(Subtests = 2:kGx,map2(aav[a],2:kGx,aaf) %>%
simplify2array() %>% t %>% as.data.frame()) %>%
knitr::kable(digits = 2, caption =
paste0("Accuracy of ",aaNames[a]," LD (Regular Composites)")))

```