2012

Preliminary Analysis of the Geriatric Intelligence Test

Adam W. Fominaya
Cleveland State University

Recommended Citation

Fominaya, Adam W., "Preliminary Analysis of the Geriatric Intelligence Test" (2012). *ETD Archive*. 596.
https://engagedscholarship.csuohio.edu/etdarchive/596

PRELIMINARY ANALYSIS OF THE GERIATRIC INTELLIGENCE TEST

ADAM W. FOMINAYA

Bachelor of Arts in Psychology

The University of North Carolina Wilmington

May, 2010

Submitted in partial fulfillment of the requirements for the degree

MASTER OF ARTS IN PSYCHOLOGY

at the

Cleveland State University

January, 2012

This thesis has been approved

for the Department of PSYCHOLOGY

and the College of Graduate Studies by:

_____

Thesis Chairperson, Amir Poreh, Ph. D.

_____

Department & Date

_____

Committee Member, Boaz Kahana, Ph. D.

_____

Department & Date

_____

Committee Member, Leslie Fisher, Ph. D.

_____

Department & Date

ACKNOWLEDGEMENTS

I owe considerable gratitude to my research advisor and mentor, Amir Poreh, for his devotion to this project, my education and my future career goals. Mr. Stewart Collins and the research participants, who generously volunteered their time, also deserve considerable appreciation. Without their help and guidance, this project would not have been possible.

I would also like to thank the multitude of professors, mentors, advisers, family and friends who have been enormously influential in my education. Thank you to Elaine Ashby, Vaughn Crowl, James Johnson, Nora Noel, Antonio Puente, Jeffery Toth, Karen Daniels, Paul Buongiorno, Michael Horvath, Boaz Kahana, Anne-Marie Hodge, my parents, and my two very supportive brothers

PRELIMINARY ANALYSIS OF THE GERIATRIC INTELLIGENCE TEST

ADAM W. FOMINAYA

ABSTRACT

The Wechsler Abbreviated Scale of Intelligence was not designed specifically for older adults. Many potential confounding factors occur with greater frequency in the elderly population and these may make the WASI a less appropriate measure of general cognitive functioning. This preliminary study aimed to develop the Geriatric Intelligence Test (GIT). Items were piloted on older adults (>80) who also completed two WASI Subtests. Preliminary results show that multiple GIT subtests are strong predictors of WASI subtest scores and showed good internal consistency. Results will be used to develop a more parsimonious version of the test.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

Versions of the Wechsler test have existed since 1939. The Wechsler-Bellevue was based on previously created tests dating back to as early as 1890. Modern tests include: The Wechsler Adult intelligence Scale – Fourth Edition (WAIS-IV)and The Wechsler Abbreviated Scale of Intelligence (WASI). The WAIS-IV is an IQ test designed for individuals ages 16-89. The WASI spans an even wider range of ages, 6-89. Older adults (65+) made up only 27% of the normative sample for the WAIS-IV, and just 20% of the normative sample for the WASI (Psychological Corporation, 2008; 1999). Given the age range of the test and national demographics, these proportions are relatively fair. However, there are two main problems: First, the baby boomer generation is beginning to enter senior citizenship and, as such, older adults will soon make up a larger proportion of the population of the United States than in previous generations. Second, life expectancy and quality of life may both continue to increase as medical knowledge and care improve.

The intelligence tests which are being used today were designed to cover a wide age range. Because of this, the unique concerns of older adults, which may not be related to cognitive functioning, were not directly considered. Roughly half of adults over 65 are

diagnosed with arthritis (Center for Disease Control and Prevention, 2010). The

prevalence of age related macular degeneration increases with age and is estimated to

increase substantially over the next generation (The Eye Diseases Prevalence Research

Group, 2004). Tremors, hearing loss, glaucoma, cataracts and numerous other deficits

are common in the elderly, are unrelated to cognitive performance, and are not

adequately addressed by intelligence tests today.

The present study is a preliminary investigation of The Geriatric Intelligence Test

(GIT). This test was designed to address the specific concerns related to the assessment

of cognitive functioning in older adults. The GIT is designed to be a measure of verbal

knowledge, verbal reasoning and perceptual reasoning abilities. Working memory and

processing speed are not assessed by the GIT. These processes are heavily interrelated

and difficult to disentangle, especially in older adults. As such, test developers

determined that they should not be included as a measure of general intellectual

functioning in older adults. The primary goal of this investigation is to pilot early tests to

determine their appropriateness for further study. Each test will be evaluated for its

internal consistency and ability to predict WASI subtest scores. The results of this study

are preliminary and will be used to determine which tests should be discarded, which

should be redesigned, and which should be kept as they are for further future study.

This portion of this report is meant to provide a background to intelligence testing

and will highlight the areas of concern with respect to intelligence testing in older adults.

As such, the examination will begin with an examination of the early history of

intelligence testing and the influence of early tests on modern measures. It is important

to understand the history behind the development of intelligence tests. Today's tests are

closely tied to the earliest theories and designs. Understanding the progression of these tests can help to avoid making mistakes previously made. GIT developers argue that the Army Alpha Test designs can be adapted to fit the needs of the proposed test.

Next, an examination of the possible influences of somatic, cognitive, and psychiatric factors on cognitive performance testing will be presented. Somatic factors include degradations in sensory acuity and motor output. Cognitive factors include those cognitive domains known to decline as a result of the normal aging process. Lastly, psychiatric factors are reviewed and the challenges to accurate testing are discussed.

Following this, the test design will be reviewed and specific criteria for item development are considered and presented. The influence of the Army Alpha Tests on the current GIT is considerable. Many of the tests designs are quite similar to those created by Yerkes (or perhaps, more accurately, by Alfred Binet and Arthur Otis). Modern measures utilize administration procedures with timed measures and open ended questioning. To more directly measure each construct, the GIT utilizes multiple-choice close ended questioning.

There are several benefits to the GIT test design and administration procedures. First, administration is easy, which decreases the influence of the examiner as there is less variability in the presentation of the test to examinees. Scoring is simplified and objective. There is no inter-rater disagreement. The shorter format decreases the influence of fatiguing factors.

Perhaps the greatest benefit to the multiple choice format is its adaptability to computer. Utilizing technology will allow the GIT to be shared with clinicians across the

country with the anonymous data transmitted back to the test developers for analysis and continuous redevelopment of norms.

The multiple-choice format also allows for more accurate translations into most other languages. This opens the possibility of sharing the test across the world along with a comparison of results among and within cultures. Clinicians will be able to quickly and accuratelyestimate general intellectual functioning. Test developers can regularly redevelop norms and can offer predictions of WASI or WAIS test scores based on continuously revised regression equations.

In part, this study is meant to reexamine the possibility of using Yerkes based administration procedures and test designs to predict scores on Wechsler based tests. The eventual product is meant to be a widely distributed, multilingual collection of tests to be shared online (through secure web-programs) for mass data collection. The cost is anticipated to be substantially lower than Wechsler based tests increasing the availability of the test to clinicians. By offering the test online, developers hope to acquire data from more rural areas rather than the urban based normative sample provided by the Wechsler tests. This test is designed to be more accurate, more accessible, more malleable to the needs of clinicians (across the world), less expensive, easier to administer, and easier to score.

CHAPTER II

LITERATURE REVIEW

2.1 The Early History of Intelligence Testing and Early Contributions to Modern Tests

Intelligence testing is one of the oldest lines of research in modern psychology. J.M. Cattell, Alfred Binet, Francis Galton and Charles Spearman (among others) were making their mark on intelligence testing within the first 25 years of the establishment of Wundt's lab in 1879. This is a brief review of the early history of intelligence testing and its influence on modern day IQ assessment and an overview of many of the major contributors to the field. This examination will begin with Wilhelm Wundt's students: Francis Galton, J.M. Cattell, and Charles Spearman and their contributions to the field through test design, statistics, and theory. Next, Alfred Binet and those who carried on his research – including Arthur Otis and Lewis Terman – will be discussed. After this, Robert Mearns Yerkes and the Army Alpha and Beta Tests (designed to be administered to World War I army recruits) will be described. Finally, this report will conclude with a review of David Wechsler's contribution to the field of intellectual assessment. Each of these individuals influenced the future of the field. Many borrowed tests, ideas, or

theories from their predecessors. In other cases, tests were designed to deviate from the status quo. These contributions can still be seen in present day tests.

Francis Galton.

Francis Galton believed in the existence of "a general mental ability with biological underpinnings, as a product of the evolutionary process (Jensen, 1994)." Further, Galton believed that this general ability could be measured objectively through measures such as discrimination and reaction time (RT) (Jensen, 1994). Galton's theory stated that people take in information through their senses, "and thus the most intelligent people must have the best developed senses (Kaufman, 2000)." Both of these ideas heavily influenced researchers of the era, namely J.M. Cattell and Charles Spearman.

Galton made a sizeable contribution to modern day intelligence testing, being "the first to demonstrate that the [normal distribution] could be applied to human psychological attributes, including intelligence (Plucker, 2007)." He was the first to use percentiles to measure "relative standing on various measurements (Plucker, 2007)". He also "invented the measures of bivariate correlation and regression (further developed by Karl Pearson; Jensen, 2002)". His work had a substantial impact on the development of modern scoring methods used on tests like the WAIS-IV, as evidenced by the fact that every index score is reported based on the normal distribution and percentile ranks.

Galton is, perhaps, best known for his stance on eugenics. "It seemed obvious and even unarguable to Galton that, from a eugenic viewpoint, superior mental and behavioural capacities, as well as physical health, are advantageous, not only to an individual but for the well-being of society as a whole (Jensen, 2002)."

J.M. Cattell.

One of the earliest players in the field was J.M. Cattell, a student of Wilhelm Wundt's (Aiken, 1999). Cattell was adamant that psychophysical measurement was essential to the field of psychology as a whole. In his highly influential article in 1890, Cattell wrote, "Psychology cannot attain the certainty and exactness of the physical sciences, unless it rests on a foundation of experiment and measurement (Cattell, 1890)."

Cattell was heavily influenced by his time studying with Francis Galton. Cattell's largest contribution to intelligence testing comes from Mental Tests and Measurements, published in 1890. Many of the tests described in this publication were already being used by Galton in Europe, and were simply transferred to the United States. As well, many of these tests involve RT or discrimination, following Galton's reasoning that these could be objective measures of a general mental ability. Galton is cited multiple times in Cattell's historic article (Cattell, 1890). The ten tests Cattell used were:

1. Dynamometer Pressure.

2. Rates of Movement.

3. Sensation-areas.

4. Pressure causing Pain.

5. Least Noticeable difference in Weight.

6. Reaction-time for Sound.

7. Time for naming Colours.

8. Bi-section of a 50 cm. Line

9. Judgment of 10 seconds time.

10. Numbers of Letters remembered on once Hearing.

Here we can see some contributions to modern day IQ assessment. The final test of Cattell's battery involves repeating letters back to the participant. This is quite similar to Digit Span, which is used in the modern WAIS-IV (Psychological Corporation, 2008). Like the modern day administration of the digit-span test, Cattell is careful about the speed or rate that the letters are presented. Cattell also states the importance of avoiding the use of vowels so that syllables cannot be formed (Cattell, 1890). More than one Wechsler test also includes the Letter-Number Sequencing subtest. The letters used in Letter-Number Sequencing are, likewise, always consonants (Psychological Corporation, 2008).

In his description of the tenth task, Cattell states that he begins by naming six letters. If the participant cannot accurately repeat six letters, Cattell will reduce the number of letters from six to five, and then to four, until the participant successfully completes a trial (Cattell, 1890). This is, seemingly, the genesis of the reverse criteria, which is included in every test in the WAIS-IV (Psychological Corporation, 2008).

Clark Wissler.

In 1901, utilizing the newly created Pearson's correlation formula (Freed & Freed, 1992), Wissler examined the data collected by Cattell (his mentor) and found few or only weak relationships between the different tests and little relationship between overall performance on Cattell's tests and academic performance. In essence, Wissler's data suggested that Cattell's Mental Tests and Measures were not measuring one construct but rather many; and furthermore these many constructs, were unrelated to intelligence (as measured by academic performance). Despite having a monumental

influence to the field of psychology, most of Wissler's professional life was spent as an anthropologist (Plucker, 2007).

Charles Spearman.

Charles Spearman undertook the task of debunking Wissler's findings. He did this, in part, by explaining that error in measurement always underestimates the degree of relationships between two variables. This is called correlation attenuation. Essentially, any obtained measurement is comprised of two additive components, a true score and random error. "It follows that the total variance of [an obtained measurement] consists of true-score variance plus the error variance (Jensen, 1994)." Since error is random, the error of one variable cannot be correlated with the error of another variable. "Spearman realized that, in evaluating the obtained correlation between variables, one must take into account the proportion of the total variance of each variable that consists of true-score variance. This proportion became known as the reliability coefficient (Jensen, 1994)." These concepts are the fundamentals of classical test theory.

In sum, Spearman discovered that measurements on Cattell's Mental Tests and Measures included some error. If these values were analyzed using a Pearson's Correlation Formula, without correcting for attenuation of the correlation coefficient, it would appear as though each of the tests were weakly correlated or uncorrelated with one another. After statistically correcting for this Correlation Attenuation, the true relationships between scores emerged and were much more consistent with previous conceptualizations of a general mental ability. He did all of this while still a graduate student in Wundt's lab (Jensen, 1994).

After seeing the degree to which these measures were related, Spearman became increasingly convinced that Galton's conceptualization of a general mental ability did indeed exist. "Not only were the sensory and RT tests themselves substantially intercorrelated, but they were substantially correlated with independent estimates of the subjects' level of intelligence (Jensen, 1994)."

In addition to this, Arthur Jensen calls Spearman "arguably the most distinguished figure in the history of British psychology… [he] was the first systematic psychometrician and father of what is known today as classical test theory (Jensen, 1994)." He was also a pioneer of factor analysis (Plucker, 2007).

Arthur Binet.

Arthur Binet was a French psychologist who believed that intelligence could not be measured by elementary cognitive processes and demanded that good intelligence tests would examine more complex processes (Brody, 2000). In 1905, Binet publishedNew Methods for the Diagnosis of the Intellectual Level of Subnormals (Binet, 1916). This publication outlined the tests developed by Binet and Theodore Simon to "measure the intellectual capacity of a child" (Binet, 1916). This collection of tests was eventually called the Binet-Simon Scale (Plucker, 2007). Some of the tests included on the Binet-Simon Scale share characteristics with the modern day WAIS-IV. Repetition of Three Figures is a task that is similar to Digit-Span Forward. Comparison of Known Objects from Memory may also examine processes such as those addressed by the Similarities subtest. This test also has some resemblance to the Picture Concepts test on the Wechsler Intelligence Scale for Children (WISC). Exercise of Memory on Pictures includes elements that can be found on the Wechsler Memory Scale (WMS), on which an

individual must recite the names of items seen earlier in the test. Drawing a Design from Memory is nearly identical to the same test on the WMS. In fact, Figure 1 and figure 2 are found on today's Wechsler Memory Scale with only minor changes. The test titled Resemblances of Several Known Objects Given from Memory is nearly identical to the Similarities test on the WAIS-IV. Reply to an Abstract Question is remarkably similar to the Comprehension test on the WAIS-IV. Definition of Abstract Terms is quite similar to Comprehension, being little different than Comparison of Known Objects from Memory described above (Binet, 1905).

Wasserman and Tulsky (2005) go so far as to say, "[Alfred Binet] may justifiably be called the father of cognitive and intellectual assessment". The Binet-Simon was the first modern day intelligence test (Wasserman &Tulskey, 2005), and as illustrated above, many of the subtests have influenced modern day tests. In some cases, modern tests seem to have been entirely adopted from the Binet-Simon.

Arthur Otis.

Arthur Otis is not generally considered to be one of the major contributors to the field of psychology. Indiana University maintains a webpage which includes biographies of leading contributors to the development of intelligence theory and testing practices, and as of November 2010, Arthur Otis is not among the 78 names listed in the alphabetical index. Yet his impact is noteworthy. Gary Robertson called Arthur Otis "a particularly ingenious developer of test item formats (Robertson, 1994)." Otis "made group intelligence testing a reality in schools (Robertson, 1994)". Along with David Wechsler, Arthur Otis is credited with "the development and implementation of the deviation IQ (Wasserman &Tulsky, 2005)." Eventually, substantial portions of his Otis

Group Intelligence Scale were incorporated into the highly influential Army Alpha Tests. In fact, "nearly half of the Army Alpha Tests came from the work of Arthur Otis (Wasserman &Tulsky, 2005)." Through his influence on the Army Alpha and Beta Tests, Otis' designs also influenced a number of modern tests. A version of the Analogies tests appears on the present day Graduate Record Examinations and often appears on the Scholastic Aptitude Test. The Comprehension test on the WAIS-IV still includes some proverbs, another of Otis' tests. The Arithmetic Subtest on the WAIS-IV seems to owe its roots to the Otis Group Intelligence Scales. The Otis Scales included a test titled Geometric Figures. This likely influenced the WAIS-IV subtest Visual Puzzles. Otis' tests were designed to be given to a group. Because of this they were easily adaptable to the Army's group examination requirements. Otis' design differed from the Binet-Simon in that respect, as the Binet-Simon was administered to only one participant at a time.

Lewis Terman.

Lewis Terman was Arthur Otis' mentor. Terman helped Otis publish his Group Intelligence Scale and was responsible for bringing this scale to the Committee on the Psychological Examination of Recruits, which developed the aforementioned Army Alpha and Beta Tests (Wasserman &Tulsky, 2005). In addition, Terman developed the Stanford-Binet, an adaptation of the Binet-Simon for American populations (Plucker, 2007). Furthermore, he adapted the Binet-Simon Scales to a group format and titled them the Stanford Achievement Test. The original Wechsler-Bellevue was developed specifically in response to specific concerns about the Stanford-Binet (Pastorino, &Doyle-Portillo, 2009).

Robert Mearns Yerkes.

In 1917, Robert Mearns Yerkes, then president of the American Psychological Association, was named The Chairman of the Committee for the Psychological Examination of Recruits and given the task of overseeing psychological testing for United States Army Recruits during World War I. This committee determined that the most appropriate way to classify these recruits was on the basis of intellect (Yerkes, 1921). These tests were named the Army Alpha and Army Beta Tests. They were developed, in large part, from the existing tests of the day. As mentioned above, Arthur Otis' methodology became particularly important because he had previously developed tests to be administered to groups.

There are a substantial number of tests and individual items which appeared on the Army Alpha Tests that still exist today. Many of the familiar tests have been mentioned earlier in this examination since the Army Alpha Tests were much more of an amalgamation of preexisting assessments. However, there are a number of actual test items which appear on the Army Alpha that can still be seen today. Individual test items include (Yerkes, 1921):

Likenesses and differences.

1. In what way are the eye and the ear alike? (appears on the Similarities Subtest in the WAIS-IV; Psychological Corporation, 2008)

Logical memory.

1. Anna Thompson of South Boston, employed as a scrub woman in an office building, reported at the city hall that she had been held up on State Street the night before and robbed of about five dollars. She had four little

children and the rent was due. The officers made up a purse for her.

(Psychological Corporation, 2009)

Comprehension test.

1. Why does land in the city cost more than land in the country?

2. Why should people have to pay taxes?

3. If you picked up a pocket book on the road with a hundred dollars in it, what would you do to find the owner?

4. What should you do if you find a sealed, stamped, and addressed envelope in the street?

Each of these items is comparable or identical to an item in the WAIS-IV (Psychological Corporation, 2008).

Orientational [sic] information.

The questions, "What day is it? What month? What day of the month? What year?" appear on the Brief Cognitive Status Exam in the Wechsler Memory Scale (Psychological Corporation, 2009). Subjects were asked to name the days of the week backward. Also, subjects were asked to name the months of the year backward. These instructions appear on the Brief Cognitive Status Exam in the Wechsler Memory Scale (Psychological Corporation, 2009)

Memory for designs.

Two of the items from this test are seen, unaltered, in the Wechsler Memory Scale – fourth edition and are shown in Figure 3 and Figure 4 (Psychological Corporation, 2009). In addition, the Vocabulary Test from the Army Alpha is almost identical to the administration of that same exam in the WAIS-IV (Yerkes, 1921; Psychological

Corporation, 2008). Digits Forward had previously appeared on the Binet-Simon Scales, Digits Backward appear on the Army Alpha Tests (Binet, 1916; Yerkes, 1921). The only new manipulation in the Wechsler administration is Digit-Span Sequencing. Cube Construction is nearly identical to Block Design from Wechsler's Tests, and Arithmetical Reasoning is identical in design to the Arithmetic subtest in the WAIS-IV (Yerkes, 1921; Psychological Corporation, 2008).

David Wechsler.

"The link between the World War I practical innovations and current tests and practices in the clinical assessment of intelligence is summed up by two words: David Wechsler (Kaufman, 2000)." Wasserman and Tulsky (2005) wrote: "It appears that Wechsler's strength was not in writing and developing items. Instead, Wechsler was a master at synthesizing tests and materials that were already in existence". This was precisely how the Wechsler-Bellevue scale was originally created in 1939. Wechsler met Yerkes after volunteering to "score the Army Alpha Tests (Plucker, 2007)." He later became familiar with the Stanford-Binet after agreeing to administer it to recruits who had performed poorly on the Army Alpha Tests (Plucker, 2007). As illustrated throughout this report, it is from these tests that the Wechsler-Bellevue was developed in 1939. The Wechsler-Bellevue quickly overtook the Stanford-Binet as "the most widely used test of intelligence... After decades of refinement and redesign, Wechsler's intelligence tests continue to dominate intellectual assessment" among many major fields of psychology (Wasserman &Tulsky, 2005). As mentioned above, Wechsler was also the co-creator of the deviation IQ (still used today).

Conclusion.

The goals of this review were to illustrate the exchange of ideas in the early history of intelligence testing, in order to demonstrate the ways in which the development of intelligence tests ultimately contributed to psychology as a science through the creation of new statistics, measurements, and procedures. This report covered many key figures in the development of intelligence tests today, but it was certainly not exhaustive and many key names were not mentioned. This examination was meant to demonstrate that very early ideas have endured into modern day testing procedures and administration. Ideas that flowed through the pen of Charles Spearman in 1904 continue to be applied throughout psychology. Tests items designed by Alfred Binet continue to be placed in front of participants today. When considering standardized testing in schools, IQ testing for enrichment placement, and testing for cognitive deficits and developmental delays, it is hard to imagine an individual in the United States not being affected by the work of these early psychologists.

CHAPTER III

DEFINING AND MEASURING INTELLIGENCE IN OLDER ADULTS

3.1 Defining and Measuring Intelligence

Defining intelligence has been a recurrent problem in the field of intellectual assessment. Definitions in 1921 ranged from "the ability to learn or having learned to adjust oneself to the environment" to "the capacity to acquire capacity" (Wasserman &Tulsky, 2005). Wechsler's theory of intelligence stated:

Intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment. It is global because it characterizes the individual's behavior as a whole; it is an aggregate because it is composed of elements or abilities which, though not entirely independent, are qualitatively differentiable (Wechsler, 1975, as cited in Wasserman &Tulsky 2005).

Sternberg and Detterman (1986) published a collection of essays which suggested that there is still no consensus definition of intelligence to that point. Later John Horn wrote:

Unresolvable issues should be recognized as unresolvable, and defining human

intelligence is definitely one such issue... Given such unsolvable problems, we can never expect to know the precise nature of intelligence. We can know some and learn more. The problem of understanding intelligence is one of building construct validity, which requires the building of a scientific theory (Horn, 1991).

Separate from a definition for the construct is a theory to describe intelligence. The first is meant to explain intelligence and the way it is applied to the world; the latter is meant to explain those variables of which the construct is comprised. One popular theory of intelligence is the Cattell-Horn-Carroll (CHC) Model. This is the theory on which most current intelligence tests are based (Keith & Reynolds, 2010). A broad description of the CHC model explains that it is combination of two theories. First, the *Gf-Gc* model (Fluid Reasoning-Crystallized Intelligence respectively) developed by Cattell and Horn. Second, the Carroll Three-Stratum Theory which posits that *g* (General Intelligence) is hierarchical and that *Gf* and *Gc,* along with other factors, load onto *g* (Keith & Reynolds, 2010).

Alternatively, the Wechsler Adult Intelligence Scale is based on the idea that general intelligence (*g*) is composed of exactly four factors. These include, Verbal Comprehension, Perceptual Reasoning, Working Memory and Processing Speed. These factors act as the four indices used for determining one's Full Scale IQ (FSIQ). Despite selecting separate factors, Carroll's three-stratum approach is maintained in the Wechsler model (Lichtenberger & Kaufman, 2009).

The Wechsler Abbreviated Scale of Intelligence is based solely on the Verbal Comprehension and Perceptual Reasoning indices and removes from consideration the Working Memory and Processing Speed Indices (Lichtenberger & Kaufman, 2009). The

GIT maintains the Three Stratum model but reduces the total number of factors loading

onto *g* to two.

General intellectual functioning has not been correlated to a specific brain area.

Intellect may be an important area of discussion for children and young adults as it can be

a potential predictor of academic or occupational success. In older adults, this is less of

an issue. Clinical neuropsychologists may question the relevance of an IQ score in an

elderly patient. Certain WAIS subtests have been shown to be sensitive to dementia.

Other subtests appear to be unable to differentiate between demented and non-demented

patients (Izawa, Urakami, Kojima, Ohama, 2009). The GIT will, after completion, be

capable of providing an estimated IQ score. This is not the primary purpose of this test,

despite its name. These tests are designed to serve one of four functions, (a)

differentiation between demented and non-demented older adults, (b) indicated the

severity of a dementia, (c) detect dementia earlier in its course and predict later cognitive

decline, (d) estimate premorbid functioning.

The goals of the GIT with respect to dementia are not the focus of this present

study. This criterion validity must be established by a subsequent study. The purpose of

the present study is to design the tests such that the influence of confounding factors is

limited or removed.

Premorbid functioning can be estimated using a number of techniques. Certain

cognitive domains show resilience to the effects of neurodegenerative diseases. If these

domains remain unaffected, measuring them in a demented patient should yield a good

estimation of previous functioning. The difficulty is that this approach does not consider

a person's relative abilities prior to the onset of a neurodegenerative disease. Consider a

patient who was average in areas related to crystallized intelligence yet above average or superior with respect to perceptual reasoning abilities. The deleterious effects of the disease may manifest as decreased scores on perceptual reasoning tasks. This could show that the patient's current functioning is average with respect to perceptual reasoning abilities. The patient's crystallized intelligence would appear to, likewise, be in the average range. Hence, a clinician might erroneously conclude that the patient was functioning at average levels and may miss the relative decline for this particular patient. This example takes a narrow view of the process. In reality, socioeconomic status, academic performance, occupation, and considerable other data are considered before arriving at a final estimation of premorbid functioning. However, premorbid functioning can be estimated by objective test data and accurately estimating this important variable can make a considerable difference in a patient's final diagnosis. This is an area in which certain WAIS subtests have performed quite well (Izawa, et al, 2009; Lanham &Misukanis, 1999; Paolo, Ryan, &Tröster, 1997). The GIT test is expected to be similarly capable of estimating premorbid functioning.

3.2 Justification for the Development of this New Intelligence Test

To date, there is no well known, widely distributed, easy to attain and administer intelligence test for the elderly population. Often, elderly clients are tested using the Wechsler Adult Intelligence Scales (WAIS) or Wechsler Abbreviated Scale of Intelligence (WASI).   These tests are designed for a wide age range which reaches as young as 16 for the WAIS-IV or only 6 years old for the WASI. There are numerous factors relating to old age which suggest that it is more appropriate to consider the elderly

population separately for the purposes of intellectual assessment. This would involve the development of an intelligence test specifically created to be administered to older individuals. The Geriatric Intelligence Test (GIT) proposed here is in its incipience; however, the GIT is designed to be sensitive to those aspects of life as an older adult which may act as confounds in neuropsychological testing in older adults. These concerns and problems relating to the sampling procedures used to develop the WAIS-IV and WASI are discussed.

The development of the deviation IQ was a monumental step in the assessment of cognitive abilities. When a test is developed, it is purported to measure a set of variables. Test developers and consumers too often assume that an individual's deviation from the mean is a reflection of deficits (or strengths) with respect to those intended variables. However, should confounding variables be present, deficits with respect to those confounds could decrease performance on the test. A clinician may erroneously assume that a decreased score is a reflection of a particular cognitive deficit, when it may, instead, be caused by an extraneous variable.

This is accounted for in classical test theory. Error is a part of nearly every measurement and should be evenly distributed. The reason that error is a larger issue in older adults is as follows: An observed score is comprised of true score and error. However the relative contributions of each could be different for older adults than they are for younger counterparts. Restated, older adults may be more heterogeneous with respect to the sources of error and the degree to which they impact an observed score. With regard to the WAIS and WASI, test developers assume that the sources of error for younger adults are identical to the sources of error in elderly adults. Inversely, it is

assumed that the tests measure the same latent variable to the same degree in both younger and older adults. This assumption may not be true (Salthouse&Saklofske, 2010).

The contention presented herein is that error plays a greater role in scores observed in older adults than it may in younger adults. In essence, the proportion of an observed score which is reflective of the latent construct decreases in older adults as more confounding variables are introduced or as confounding deficits become more pronounced. Such sources of error in older adults could be eye related diseases, hearing loss, motor impairment, arthritis, decreases in cognitive functioning as a result of normal aging, decreases in cognitive performance from sources other than those purported to be measured by a particular test, or psychiatric influences such as geriatric depression. A review of potential confounding factors as well as evidence of cognitive heterogeneity in older adults is presented.

Eye diseases and poor vision are important considerations when testing an older adult as performance on a given item can likely be influenced by the patient's ability to receive the necessary sensory input. If a patient cannot see the item or a salient part of the item, that individual will be less likely to provide a correct response. This is an important source of measurement error in the elderly population as a number of age-related eye diseases exist, including age-related macular degeneration and cataracts, and could act as confounds. As the size of the elderly population grows, these eye diseases are, understandably, being seen in greater frequency (Sterns, 2009). Test items which have subtle but important components will be more difficult for older adults.

In addition to eye related diseases, hearing loss is another concern in testing in the

elderly. Wingfield (2005) reports that presbycusis, or age-related hearing loss, is "not uncommon". Causes of presbycusis could include conductive or sensorineural hearing loss. Hearing loss could play an important role in many of the WAIS subtests, but would likely be most apparent on tasks such as digit span and arithmetic. On the digit span task, the examiner is not permitted to repeat the auditory stimuli. On the arithmetic test, the examiner is permitted to repeat the item once (upon request), though it is a speeded task and timing does not stop for repetition of the item (Psychological Corporation, 2008).

Manual or manipulative dexterity is also of concern in the elderly. Older adults complete manual dexterity tasks more slowly and have been shown to readjust hand positions more often than younger adults on certain tasks. In addition, older adults have demonstrated excessive grip force, a delay in grip force adjustments and have shown deficits related to adjustment of grip force in response to differences in load (Diermayr, McIsaac, & Gordon, 2011). Some of these deficiencies could influence scores on WAIS subtests. Pencil and paper subtests such as Symbol Search or Coding, which require speeded motor output from the patient could show decreased scores caused by factors unrelated to the purported construct, speed of information processing. The Block Design subtest is likely most susceptible to measurement error related to degradation of manual dexterity. This task requires a speeded motor output with penalties for slower completion times as well as rotation errors and gaps between manipulated blocks (Psychological Corporation, 2008).

Li and Lindenberger(2002) also present evidence of an interaction between sensory and sensorimotor deficits and age-related cognitive decline. Li and Lindenberger(2002)demonstrate that a substantial portion of the variance in cognitive

performance in older adults can be explained by sensory and sensorimotor factors. These cross-sectional findings were partially supported by additional longitudinal studies, however some inconsistencies were reported. Experimental designs utilizing manipulations of level of sensory load (i.e., controlled manipulations of signal to noise ratios) as well as dual-task paradigms both suggest a relationship between age-related cognitive decline and sensory and/or sensorimotor decline. In short, a decrease in signal to noise ratios leads to age-related decreases in performance. As well, age-related deficits in dual-task paradigms were also observed. Therefore, older adults show greater cognitive decline when sensory or sensorimotor resources are more heavily taxed. Two hypotheses have been presented to explain this relationship, namely, the common factor versus shared resources hypotheses (Li and Lindenberger, 2002).

There is general agreement in the literature that working memory and processing speed are highly interrelated in the elderly, though they are still viewed, in general, as separate constructs. The literature also suggests wide acceptance that these constructs have a considerable influence on many cognitive performance measures (Kirasic, Allen, Dobson, & Binder 1996;Nettlebeck& Burns, 2010;Park et al., 2002; Salthouse, 1992;Salthouse, 2000). An exhaustive review of this complex theoretical question is beyond the scope of the present review. Below, processing speed and working memory are discussed as separate entities (this is the way that these constructs are treated in the WAIS-IV factor structure) with the understanding that these constructs appear to be highly interrelated and difficult to disentangle.

Cognitive variables may also act as confounds on specific WAIS or WASI tests. In particular, processing speed is known to decline as a result of normal aging (Salthouse,

2000). This is reflected in the norms for these subtests, as older adults tend to perform more poorly on speeded tasks than their younger counterparts. However speeded tasks are found in three out of four of WAIS-IV indices. Speeded tasks are not limited to the Processing Speed Index. The Visual Puzzles subtest, a speeded task of mental/visual construction, loads onto the Perceptual Reasoning Index. The Block Design subtest, a speeded visuomotor/visuospatial construction task, loads onto the Perceptual Reasoning Index. And, the Arithmetic subtest, a speeded measure of attention and working memory, loads onto the Working Memory Index (Psychological Corporation, 2008). The Block Design subtest additionally awards a bonus for completing the task more quickly. Deficits in processing speed as a result of normal aging could decrease an individual's performance in each of these tests irrespective of the intended latent variables.

Working memory also declines as a function of normal aging (Belleville, Peretz, &Malenfant, 1996; Park, et al. 2002). Subtests which involve a dual task component can be found on three of the four indices and are not limited to the Working Memory Index, namely, Block Design, Visual Puzzles, and Coding. The Block Design subtest involves both motor and visuospatial abilities. Visual Puzzles includes a working memory component in addition to the visuospatial requirements. Coding may involve a memory component in addition to the motor and processing speed requirements.

Psychiatric factors are reported to influence scores on cognitive performance tests as well (Butters, et al. 2000; Elderkin-Thompson et al., 2003;Herrmann, Goodwin, &Ebmeier, 2007). The authors deemed that, outside of treatment with reported remission of symptoms, there is little that can be done with respect to methodological or procedural manipulations which could attenuate this concern.

With regard to sampling, the primary issue with the WAIS-IV and WASI tests is the remarkably low number of subjects from the elderly population that were used to norm scores for those age ranges. The WAIS-IV, for example, created 13 age bands with five of these being in the 65+ age range (and not exceeding age 90). 200 subjects were used in the 65-69:11 (65 - 69 and 11 months) age band. 100 were used for each of the eldest age bands coming to a total of 600 subjects ages 65 to 90:11 (a range of 25 years). However, the same number of subjects sampled was for ages 16-24:11 (a range of only 9 years) (Psychological Corporation, 2008). These norms were designed to closely mirror the demographics of the most recent census. However, the number of subjects sampled at the eldest ages is disproportionately low with respect to the current age distribution. This will be exacerbated as the Baby Boomers continue to enter senior citizenship.

Further, norms for these tests were established using a normal or healthy population. Item analyses and examinations of factor structures may not reveal the true relationships between test items, subtests, and indices as they exist in clinical populations. Early items on certain subtests may show very little variability in healthy individuals. However, the heterogeneity of clinical populations may manifest with lower scores as well as greater variability on earlier test items.

CHAPTER IV

THE GERIATRIC INTELLIGENCE TEST

4.1 General Test Design

The Geriatric Intelligence Test (GIT) is designed to effectively respond to the shortcomings of the Wechsler tests as they are applied to the elderly population. The test was carefully designed to be easier to administer, easier to train examiners, shorter in length, and less expensive to distribute. The GIT is intended to be less fatiguing. Timed tasks were eliminated, and under the new format, issues relating to confirmation bias and subjectivity in scoring are removed. These changes should improve the testing experience for both examiner and examinee while providing a clearer picture of the individual's intelligence and are explained in greater detail herein.

Stated above, The GIT was crafted in such a way as to be easier to score and eliminate confirmation bias. These goals are accomplished by designing each test in multiple-choice or limited-choice format. By doing this, the possibility of confirmation bias is eliminated since the examiner is not involved in discerning the degree of correctness of a particular response. The response is either correct or incorrect and that decision is not made by the examiner in the room during testing. Not only does this

decrease confirmation bias but also eliminates inter-rater disagreement. Also, training clinicians to administer and score the test is far easier and less expensive ultimately making the test more portable.

An additional goal of this test is to make it more clinically relevant. By providing more concise instructions and tests with limited possible responses, a clinician is freed from many former reading, writing and speaking duties and is encouraged to observe the client. Allowing the clinician to keep their eyes and attention on the client throughout the exam should allow for a more relevant and accurate behavioral observations.

Because of the considerable literature supporting degradation of processing speed and working memory as a result of normal aging, as well as evidence that these cognitive domains are highly interrelated, test developers for the GIT have made the decision to remove those indices from those constructs intended to be measured. Overlap with these domains is minimized, in part, by removing timed tasks. Patients are encouraged to respond but may take as much time as they need on any particular item.

Lastly, the final goal of this test, and the feature which will likely make it much more useful and available than the WAIS-IV, is to computerize it. This will ensure immediate scoring and application of regression formulas to provide instantaneous scores on the GIT as well as estimations of performance on a multitude of other tests. Beyond this, if the test is provided to clinicians through an online medium, anonymous data can be collected *en masse* as clinicians are making use of the test. This increases the external validity of the data collected and provides researchers with considerable and robust data.

4.2 Subtest Development

The remainder of this chapter will focus on the criteria upon which test items were designed and the goals of the new test. Rationale behind the decisions such as inclusion or exclusion of each test item will be discussed in general, rather than on an item-by-item basis. Each subtest had specific rules and tendencies which test developers attempted to maintain throughout all items of a particular subtest. Deviations from general rules will be indicated where such practices were intentional. Recommendations for the final criteria for included items as well as the order in which items are presented will be provided in the discussion section.

Six subtests were developed, three were anticipated to be predictive of the verbal comprehension index; three were anticipated to be predictive of the perceptual reasoning index. The verbal measures designed were (a) a vocabulary subtest, (b) an analogies subtest, (c) an information subtest. The perceptual reasoning measures were (a) a matrices subtest, (b) a complex geometrical figure construction subtest, (c) a test for hidden shapes. Each test is discussed below.

The vocabulary subtest is partially designed based on Bowles and Salthouse (2008) findings that multiple-choice format vocabulary tests showed the least decline in scores as a function of age compared to three other vocabulary tests utilizing various methods. As well, Blatt (1959) concluded that recognition based vocabulary tests may be a better measure of premorbid intellectual functioning than recall vocabulary tests. This is a common use for the vocabulary subtest of the WAIS (Lanham and Misukanis, 1999). Many stimuli were drawn directly from the Army Alpha tests (Yerkes, 1921).

Stimuli varied by asking the participant to provide a synonym or provide a definition. Criteria for response options varied depending on the characteristic of the particular stimuli. For vocabulary items in which the stimulus was a noun, most or all of the response options were also nouns. This relationship is likewise maintained with the other parts of speech. When the stimulus word carried a particular connotation, positive or negative, an effort was made to provide at least one option which appeared to have the opposite connotation. In general, most items (with the exclusion of some items earlier in the test) have at least one absurd or wholly unrelated possible choice.

The analogies subtest has, perhaps the most succinct description as nearly all items, stimuli and response sets were drawn directly from the Army Alpha tests with very few additions by the authors (Yerkes, 1921). Those additions which were included were designed simply on the criteria that they be relatively different from existing test items in terms of content and/or type of solution.

The information subtest was designed to attempt to balance the type of content being addressed. The WAIS-IV information subtest asks several questions related to geography and history with very little variation in the type of content being assessed. Two questions ask the participant to name on which continent a particular stimulus resides. History related questions trend toward asking for the names or descriptions of specific people. For the development of the GIT information test, many items were drawn from the Army Alpha tests with stimulus and response set generally maintained. An effort was made to ensure that specific questions were categorized by type of content assessed. A general category was included for items which did not clearly fit into another category. Most items in the general category were Army Alpha items.

Categories were: (a) geography, (b) history, (c) science, (d) arts and mythology. A concerted effort was made to minimize repetition of content. The intention was to ensure that test items were more independent while still existing within the same cognitive domain. As well, this is a recognition based test meant to be a predictor of premorbid intelligence.

The matrices subtest was designed to be more complex than WAIS-Matrices. The grids in which items were presented varied by number and orientation. Answers followed a logical pattern presented within the grid and were deemed to be the only possible option from the response set which could adequately maintain the pattern. Stimuli were designed to be quite large. Stimuli very rarely varied based on color. When color variation was included as part of the stimuli or response set, measures were taken to ensure that colors could be seen clearly. Items which may have been difficult for some participants to see were removed.

The geometric figures subtest was designed to reduce the number of items required by only asking the participant to mentally construct the puzzle from two pieces instead of three (as mandated by the WAIS-IV). Almost all items in the response set were constructed from pieces of the original stimulus, however there was only one combination of two response set items which could adequately be mentally constructed to recreate the original stimulus. Rules governing the use of color variations which were discussed in the matrices subtest are maintained throughout this subtest as well. In addition, when color variations were utilized in this subtest, a concerted effort was made to ensure that there were multiple ways to solve a particular item. An item including small colored circles can be solved based on color variations, relative position, or

number of circles. Only one combination of response set items would yield the correct number of circles presented in the original stimulus.

The hidden shapes test asks the user to find a shape hidden amongst other shapes. The shape was hidden in only one of the response set items. Stimuli early in the test are designed to be quite poorly hidden. Later, more difficult, items utilize more complex interferences as well as interference stimuli which are highly similar in feature and size to the intended stimuli but which included or excluded at least one important feature.

Based on the methodological deviations from the WAIS-IV, one interesting way in which the GIT may improve our understanding of intelligence and premorbid functioning is by comparing the two tests in a multitude of clinical and non-clinical populations. Premorbid functioning may be better estimated by the GIT. Two of the six subtests are designed specifically with that purpose in mind. That said, Lanham and Misukanis (1999) suggest that tests such as a reading, vocabulary or information test should not be the only source of information for a clinician to estimate premorbid intelligence. Utilizing a more structured test, we may be better able to estimate premorbid intelligence more directly in both verbal and non-verbal domains that the WAIS can offer. This is a hypothesis which can be explored following final test design and release.

4.3 Goals

While a major critique of the Wechsler Tests is the size and recruitment of the normative sample collecting an adequate nation-wide sample is beyond the scope of the immediate project. The intention is to collect preliminary data on a relatively small

sample such that a more parsimonious test can be generated. Specifically, this study identifies those subtests which best predict WASI subtests (Vocabulary and Block design, independent of one another). It also aims to identify and remove problematic items, including items which are inconsistent with the aforementioned item development criteria, items which few individuals in our sample correctly completed, and items which were identified by the participants as difficult to see. Finally, performance on a given item across participants will be used to determine that item's difficulty rating and final position in the test. Easier items will occur earlier in the test; more difficult items will appear later.

CHAPTER V

METHOD

5.1 Participants

Participants were recruited from a local independent living apartment complex. While the initial goal of the study was to recruit individuals from ages 65 to 90 and higher, this community only provided participants over the age of 79. The mean age of individuals included in regression analyses was 87.4375 with the youngest being 80 (n=2) and the eldest being 92 (n=1). The eldest age range for WASI norms was used for individuals over age 89. This was deemed to be a benefit rather than detriment despite deviation from initial proposal. Sampling in the eldest age cohort is often quite difficult and recruitment practices are often exhaustive and rarely fruitful. Individuals recruited for this study were functioning fairly well and mostly living on their own with little or no assistance.

5.2 Materials

WASI and GIT materials were presented to participants in a quiet room on a flat surface with ample space free of interruptions. Participants were offered breaks between subtests if needed. The testing environment was deemed to be of high quality.

5.3 Attrition

Most of the earliest participants elected to discontinue the testing session before all the intended data had been collected. Examiners elected to shorten the GIT tests to between 15 and 25 items per subtest. These earliest participants could not be included in regression analyses of WASI subtest scores (they did not complete these tests in most cases). Their responses were considered for the item-by-item analysis regarding the difficulty of each item. After decreasing the length of the GIT test, most participants completed all GIT and WASI subtests, however some participants did still elect to discontinue. This was not necessarily surprising considering the length of the test in its current form and the fact that there was no compensation, monetary or otherwise, for participants.

Test administration was counterbalanced such that participants were either given the WASI or the GIT test first, with the other test following. This was to battle against a potential order effect. The order of the test was predetermined for all individuals and administrators were not aware, and could not control which participants would receive which test in which order.

Testing was generally completed in between 60 and 120 minutes. Participants were not timed on any tasks. However, they were encouraged to provide a response

when excessive latency to respond was observed. Participants were encouraged to provide a response to all GIT items even when they were unsure. Most participants complied with this directive on most items. Omitted items were counted incorrect and were included in later item analyses. Two administrators were used. Due to scheduling constraints, administrators did not see an equal number of participants.

GIT instructions were presented on the page and participants were given practice trials on all tests. Administrators were permitted to read all instructions to the participants. Likewise, participants could read the instructions themselves. This decision was left to the participant. Administrators were instructed that they were permitted to repeat any instructions presented at the onset of a subtest at any point during the subtest. Providing instructions beyond the scope of those presented at test onset was discouraged. On items in which written words were presented, administrators were permitted to read any and all items to the participants. Because reading the words to the participant was permitted, correcting mispronunciation of words was also permitted. Therefore, if an individual audibly mispronounced a word, the administrator was permitted to correct the pronunciation.

5.4 Administration Procedures

Administrators would record responses and turn pages for the participant so as to protect the participant from discouragement for incorrect responses and to protect the quality of the testing materials. Some participants attempted to turn the pages themselves. During pilot testing, participants who attempted this would inadvertently skip pages and miss answering particular items. Administrators were encouraged to turn

the pages for the participant. Standardized procedures were used for WASI administration.

5.5 Statistical Method

Multiple analyses were necessary to begin the development of this test. The primary goal of this study was parsimony. The purpose was to identify the best subtests and the best items and remove all others. This is a fine balance considering that more items and tests may be better able to predict a common underlying construct. However as previously stated, the ultimate goal of the development of this test is brevity as well as accuracy.

The first step was to identify the best predictors of WASI-Vocabulary and WASI-Block Design test performance. Next, item difficulty was established on those measures identified. Finally, internal consistency needed to be established.

Internal consistency cannot be assessed on this test using Cronbach's Alpha due to the variable difficulty of each item. This gradation of item difficulty renders Cronbach's Alpha meaningless. Because of this, odd-even reliability (a type of split-half reliability) was used to determine internal consistency. Items were scaled for difficulty in order to ensure that items of a similar difficulty level were distributed between each of the split-halves in a relatively even manner. This was completed by modifying the SPSS syntax such that the investigator, rather than the statistical program, determined the split-halves. The SPSS syntax used for the analogies and matrices analyses is presented below:

Analogies.

```
DATASET ACTIVATE OddEvenAnalogies.
RELIABILITY
  /VARIABLES=handle hour food above januaryindiana
picture establish bold abide framework moon granary
seed go windows pupil jurors lion peninsula whale
tiger city
  /SCALE('ALL VARIABLES') ALL
  /MODEL=SPLIT
  /STATISTICS=DESCRIPTIVE SCALE
  /SUMMARY=TOTAL.
```

Vocabulary.

```
DATASET ACTIVATE OddEvenVocab.
RELIABILITY
  /VARIABLES=anachronism sapient juxtapose avarice
mordant deluge never priceless gelatinous obsequious
perfunctory    complot eyelash philanthropy afloat
orange dilapidated
  /SCALE('ALL VARIABLES') ALL
  /MODEL=SPLIT
  /STATISTICS=DESCRIPTIVE SCALE
  /SUMMARY=TOTAL.
```

Information.

```
DATASET ACTIVATE OddEveninfo.
RELIABILITY
  /VARIABLES=apple diamonds denim perjury tokyomandela
silo greekmonalisa river artichoke xylophone kilowatt
turqouise eight cutlass thyroid silkroad plants
midatlanticridge
  /SCALE('ALL VARIABLES') ALL
  /MODEL=SPLIT
  /STATISTICS=DESCRIPTIVE SCALE
  /SUMMARY=TOTAL.
```

Matrices.

```
DATASET ACTIVATE OddEvenMatrix.
RELIABILITY
  /VARIABLES=ten eight three five thirteen six
seventeen sixteen two four nine eleven seven fourteen
twelve fifteen
  /SCALE('ALL VARIABLES') ALL
```

```
   /MODEL=SPLIT
   /STATISTICS=DESCRIPTIVE SCALE
   /SUMMARY=TOTAL.
```

Geometric Construction.

```
DATASET ACTIVATE OddEvenGeo.
RELIABILITY
   /VARIABLES=two five eight thirteen ten fourteen
fifteen seven seventeen one three six twelve nine
eleven four eighteen sixteen
   /SCALE('ALL VARIABLES') ALL
   /MODEL=SPLIT
   /STATISTICS=DESCRIPTIVE SCALE
   /SUMMARY=TOTAL.
```

Hidden Shapes.

```
DATASET ACTIVATE OddEvenhidden.
RELIABILITY
   /VARIABLES=fourteen eighteen nine seventeen twelve
one three five seven ninteen eleven thirteen sixteen
ten eight two four six fifteen
   /SCALE('ALL VARIABLES') ALL
   /MODEL=SPLIT
   /STATISTICS=DESCRIPTIVE SCALE
   /SUMMARY=TOTAL.
```

Notice that the variable names for the nonverbal measures are written numbers.

These represent the order in which the participants were provided the stimulus. This does

not represent item difficulty. As stated above, items were reordered by difficulty prior to

running the split-half reliability analyses.

Variables were entered such that odd numbered variables were listed first, and

even numbered variables listed last. SPSS includes approximately half of the variables in

the first split-half based on the order in which they are listed above. Zero variance items

were excluded from this analysis but will be included in the final test as zero variance

items are considered to be a constant. However, these items are likely to show some

variability when given to a larger sample particularly if that sample is a clinical population. Spearman-Brown and Guttman Split-Half Coefficients are reported.

Only tests which were identified as good predictors of WASI subtest performance were retained for further analysis and development. The remaining tests were not considered in the remaining analyses as these tests either did not predict or did not uniquely predict subtest performance on the WASI and will likely be discarded or redesigned.

CHAPTER VI

RESULTS

OLS regression was completed to identify the best predictors of WASI-Vocabulary performance. Using stepwise entry method, the GIT-Analogies test was indicated as the best predictor of WASI-Vocabulary performance, β=1.075, $t(14) = 4.003$, p=.001. GIT-Analogies also explained a significant proportion of the variance in WASI-Vocabulary performance, R^2=.534, $F(1,14) = 16.028$, p=.001. GIT-Vocabulary did not show a significant unique correlation with WASI-Vocabulary β=.369, $t(14) = 1.949$, p=.073. GIT-Information also did not show a significant unique correlation with WASI-Vocabulary β= -.003, $t(14) = -.013$, p= .990.

A separate OLS regression was completed to identify the best predictors of WASI-Block Design performance. Using the stepwise entry method, the GIT-Matrices subtest was indicated as the best predictor of WASI-Block Design performance, β=2.203, $t(14) = 3.132$, p=.007. GIT-Matrices also explained a significant proportion of the variance in WASI-Block Design performance, R^2=.412, $F(1,14) = 9.810$, p=.007. GIT-Geometric Construction did not show a significant unique correlation with WASI-Block

Design, β=.074, $t(14)$ = .310, p=761. GIT-Hidden Shapes also did not show a significant unique correlation with WASI-Block Design, β= -.097, $t(14)$ = -.419, p=.682

For the GIT-Analogies subtest, the Guttman Split-Half Coefficient was .906 with equal and unequal length Spearman-Brown Coefficient of .908. For GIT-Matrices, the Guttman Split-Half coefficient was .763, and the equal and unequal length Spearman-Brown Coefficients were both .764. For GIT-Vocabulary the Guttman Split-Half Coefficient was .592, and the equal and unequal length Spearman-Brown Coefficients were both .616. For the GIT-Information subtest, the Guttman Split-Half Coefficient was .634 with both equal and unequal length Spearman-Brown Coefficients of .634. For the GIT-Geometric Construction subtest, the Guttman Split-Half Coefficient was .203 with both equal and unequal length Spearman-Brown Coefficients of .211. For the GIT-Hidden Shapes subtest, the Guttman Split-Half Coefficient was .739 with equal and unequal length Spearman-Brown coefficients of .744.

CHAPTER VII

DISCUSSION

7.1 General Discussion

This is the first step in a long process of test development. The final product is intended to be a short, accurate, easily administered test of cognitive performance. This test should minimize the influence of extraneous factors as much as possible. It will also be designed such that it can be readily adapted to computer for administration on touch screen or tablet PCs. The goal of the immediate study is simply to begin to identify tests which may be able to be used in the final version of the GIT. All results presented herein should be replicated. Correlations should be validated against other measures in a more heterogenous population. Internal consistency should be reexamined following any modifications to items or subtests. Results presented herein are meant only to guide the next step in the development of this test.

7.2 Statistical Review and Outcomes

There is a notable caveat about the statistical analyses. The probability of type I error would be .95 for any individual multiple regression analysis with alpha set at .05.

When more than one multiple regression analysis is performed, the alpha inflates across analyses. Hence, the probability of not rejecting the null erroneously for this analysis is: $(1 - alpha)^k$ where k indicates the number of comparisons made: $(1 - .05)^2 = .9025$. Thus, the likelihood of type I error is 9.75% rather than the typical 5%. This alpha inflation was intentional in this case and was permitted for two reasons: First, this is a preliminary study. The goal of the study was to identify potential predictors. It was deemed necessary to allow alpha to inflate as the bigger issue in this study is type II rather than type I error at this exploratory stage. Second, smaller sample sizes decrease the likelihood of finding an effect. Inversely, smaller sample sizes increase the likelihood of type II error. In addition, smaller sample sizes preclude the use of the more appropriate technique, multivariate GLM. While the reasons for inflating alpha are justified, reporting said inflation is also imperative. The influence of an order effect may be an important consideration and should be methodologically extricated using counterbalancing.

The GIT-Matrices subtest was a good predictor of WASI-Block Design performance. GIT-Analogies was a good predictor of WASI-Vocabulary performance. The other four subtests were not identified as significant unique predictors of WASI subtest performance. Moving forward, GIT-Matrices and GIT-Analogies will be retained and two new subtests will be designed to better predict subtest performance on specific subtests of the WAIS-IV.

Internal consistency was gleaned using odd-even reliability analysis. GIT-Analogies was shown to have excellent internal consistency. GIT-Matrices also showed

very good internal consistency. After minor modifications, both subtests appear to be ready for the next phase of development, discussed below.

7.3 Limitations and future directions

There were a number of limitations to this study. First, the sample was fairly small and highly homogenous. This was acceptable as the test length was quite considerable and a major goal of the study was to develop a more parsimonious test. The next step in this process is to design and pilot two new subtests to conclude with a four subtest battery. Because of the length of time required to complete a number of items, it is not recommended that both correlative work and internal consistency analyses be undertaken within the same study. The present study suffered considerable attrition because of the length of the testing session. When correlative studies are completed, shorter, more recent measures should be used for comparison such as WAIS-IV subtests or subtests included in the revised WASI (which has not been released at the time of writing this report). Monetary incentives may be considered if available. Course extra credit may be offered to older adults taking courses at a university participating in the study. Finally, study investigators may reach out to clinicians to inquire about the possibility of expanding the testing to clinical populations which are receiving psychological services. Additional demographics may be collected to examine potential covariates. This may help to develop of a battery which can be used as a standalone instrument or can be used to quickly predict WAIS performance.

The final goal of this project is to develop a four subtest cognitive performance test for older adults. This test will be made available to clinicians in a secure online

format such that anonymous data is received by study investigators for the purposes of

norming and further investigation. Whereas clinicians will have the opportunity to

quickly assess general cognitive functioning and GIT performance along with

demographic information can be used to estimate WAIS scores.

REFERENCES

Aiken, L. R. (1999). Origins and developments.In *Human differences*. Mahwah, NJ: Lawrence ErlbaumAssociates, Inc.

Belleville, S., Peretz I., and Malenfant D. (1996) Examination of the working memory components in normal aging and in dementia of the Alzheimer type. *Neuropsychologia*. 34 (3), 195-207.

Binet, A. (1916). New methods for the diagnosis of the intellectual level of subnormals. In E. S.Kite (Trans.), *The development of intelligence in children*. Vineland, NJ: Publications of The Training School at Vineland. (Originally published 1905 in *L'AnnéePsychologique, 12*, 191-244.)

Blatt, S. J. (1959). Recall and Recognition Vocabulary: Implications for intellectual deterioration. *Archives of General Psychiatry*. 1, 473-476.

Bowles, R. P., and Salthouse, T. A. (2008) Vocablary test format and differential relations to age. *Psychology and Aging*.23 (2).366-376.

Brody, N. (2000). History of theories and measurements of intelligence.In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 16-33). New York: Cambridge University Press.

Butters, M. A., Becker J. T., Nebes, R. D., Zmuda, M. D., Mulsant, M. D., Pollockk, B. G., andReynolds III, C. F. (2000).Changes in cognitive functioning following treatment of late-life depression.*American Journal of Psychiatry*. 157 (12), 1949-1954.

Cattell, J. M. (1890). Mental tests and measurements. *Mind, 15*, 373-380.

Center for Disease Control and Prevention (2010).Arthritis-Related Statistics.In

    Arthritis. Retrieved from:

    http://www.cdc.gov/arthritis/data_statistics/arthritis_related_stats.htm

Diermayr, G., McIsaac, T. L., Gordon, A. M. (2011). Finger force coordination

    underlying objectmanipulation in the elderly – a mini-review. *Gerontology*.57

    (3). 217-227. doi: 10.1159/000295921

Edwards, A.J. (1996) Wechsler, David (1896-1981). In R. J. Sternberg (Ed.),

    Encyclopedia of human intelligence (pp. 1134-1136). New York: Macmillan

    Publishing Company.

Elderkin-Thompson, V., Kumar A., Bilker, W. B., Dunkin, J., J., Mintz, J., Moberg, P. J.,

    Mesholam, R. I., and Gur, R. E. (2003).Neuropsychological deficits among

    patients with late-onset minor and major depression.*Archives of Clinical

    Neuropsychology*. 18 (5), 529-549.

The Eye Diseases Prevalence Research Group (2004) Prevalence of Age-Related Macular

    Degeneration in the United States.*Archives of ophthalmology*. 122, 564-572.

Freed, S. A., & Freed, R. S. (1992). Clark Wissler: September 18, 1870–August 25, 1947.

    Washington, DC: National Academy Press.

Jensen, A. R., (2002). Galton's legacy to research on intelligence.*Journal of Biosocial

    Science*, 34,145-172.

Jensen, A. R., (1994). Spearman, Charles Edward (1863-1945). In R. J. Sternberg (Ed.),

    Encyclopedia of human intelligence (pp. 1009-1014). New York: Macmillan

    Publishing Company.

Kaufman, A.S., (2000). Tests of intelligence. In R. J. Sternberg (Ed.), *Handbook of

human intelligence (pp. 445-473). New York: Cambridge University Press.

Herrman, L. L., Goodwin, G. M., Ebmeier, K. P. (2007). The cognitive neuropsychology

of depression in the elderly. *Psychological Medicine*. 37 (12), 1693-1702.

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In: K. S.

McGrew, J. K. Werder, R. W. Woodcock (Eds), *Woodcock-Johnson technical*

manual: A reference on theory and current research (pp. 197-246). Allen, TX:

DLM Teaching Resources.

Izawa, Y., Urakami, K., Kojima, T., Ohama, E. (2009). Wechsler Adult Intelligence

Scale, 3rd Edition (WAIS-III): usefulness in the early detection of Alzheimer's

disease. *YonagoActamedica*. 52, 11-20.

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.

Kirasic K. C., Allen, G. L., Dobson S. H., Binder K. S. (1996). Aging, cognitive

resources, and declarative learning. *Psychology and Aging*. 11 (4), 658-670.

Lanham, R. A., and Misukanis, T. (1999). Estimating Premorbid Intelligence:

Determining change in cognition following brain injury.*Psychological*

Assessment.3 (3).

Li, K.Z.H, and Lindenberger (2002) Relations between aging and sensory/sensorimotor

cognitive functions. *Neuroscience and Biobehavioral Reviews*. 26, 777-783.

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

York, NY: Wiley.

Nettlebeck, T., Burns, N. R. (2010) Processing speed, working memory and reasoning

ability from childhood to old age. *Personality and Individual Differences*. 48 (4), 379-384. doi: doi:10.1016/j.paid.2009.10.032.

Paolo, A. M., Ryan, J. J., and Tröster, A. I. (1997) Estimating premorbid WAIS-R intelligence in the elderly: an extension and cross validation of new regression equations. *Journal of Clinical Psychology*.53 (7).647-656.

Park, D. C., Lautenschlager, G., Hadden T., Davidson N. S., Smith A. D., Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*. 17 (2) 299-320. doi: 10.1037//0882-7974.17.2.299.

Pastorino, E. E., & Doyle-Portillo, S. M., (2009). Cognition, language and intelligence: How dowe think?.*What is Psychology*? (pp. 285-327). Belmont, CA: Thomas Wadsworth.

Plucker, J. A. (Ed.). (2007). Human intelligence: Historical influences, current controversies, teaching resources. Retrieved [November 27, 2010], from http://www.indiana.edu/~intell

Psychological Corporation. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI) manual*. San Antonio, TX: Author.

Psychological Corporation. (2008). *Wechsler Adult Scale of Intelligence – fourth edition (WAIS-IV) administration and scoring manual*. San Antonio, TX: Author.

The Psychological Corporation (2009). Wechsler Memory Scale fourth edition:

Administrativeand scoring manual. San Antonio, TX: Pearson

Robertson, G. J., (1994). Quick measures of intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 885-888). New York: Macmillan Publishing Company.

Salthouse, T. A. (1992) Influence of processing speed on adult age differences in working

   memory. *ActaPsychologica (Amst.)*. 79 (2), 155-170.

Salthouse, T. A. (2000) Aging and measures of processing speed.*Biological Psychology*.

   54 (1).35-54.

Salthouse, T.A., and Saklofske, D.H. (2010). Do the WAIS-IV tests measure the same

   aspects of cognitive functioning in adults under and over 65?. In L. G. Weiss, D.

H. Saklofske, D. Coalson, & S. E. Raiford (Eds.), WAIS–IV clinical use and

   interpretation. New York, NY: Elsevier.Sternberg, R. J., &Detterman, D. K.

   (Eds). (1986). What is intelligence? Contemporary viewpoints on its nature and

   definition. Norwood, NJ: Ablex

Strens, G. (2010). Score of the problem and demographic shift in population: visual

   disease incidence and prevalence in the elderly population. *Geriatric*

   Opthamology. (1-5). doi: 10.1007/b137372_1.

Wingfield, A., Tun, P. A., and McCoy, S. L. (2002). Hearing loss in older adulthood:

   What is it and how it interacts with cognitive performance. *Current Directions in*

   psychological science. 14 (3).144-148.

Wasserman, J. D., &Tulsky, D. S., (2005) A history of intelligence assessment.In D. P.

   Flannigan, & P. L. Harrison (Eds.),*Contemporary intellectual assessment:*

   Theories, tests, and issues (2nd ed.) (pp. 3-22). New York: The Guilford Press.

Yerkes, R. M. (1921). Psychological Examining in the United States Army (Memoirs of

   the National Academy of Sciences, Vol. 15). Washington, DC: Government

   Printing Office.

APPENDIX A
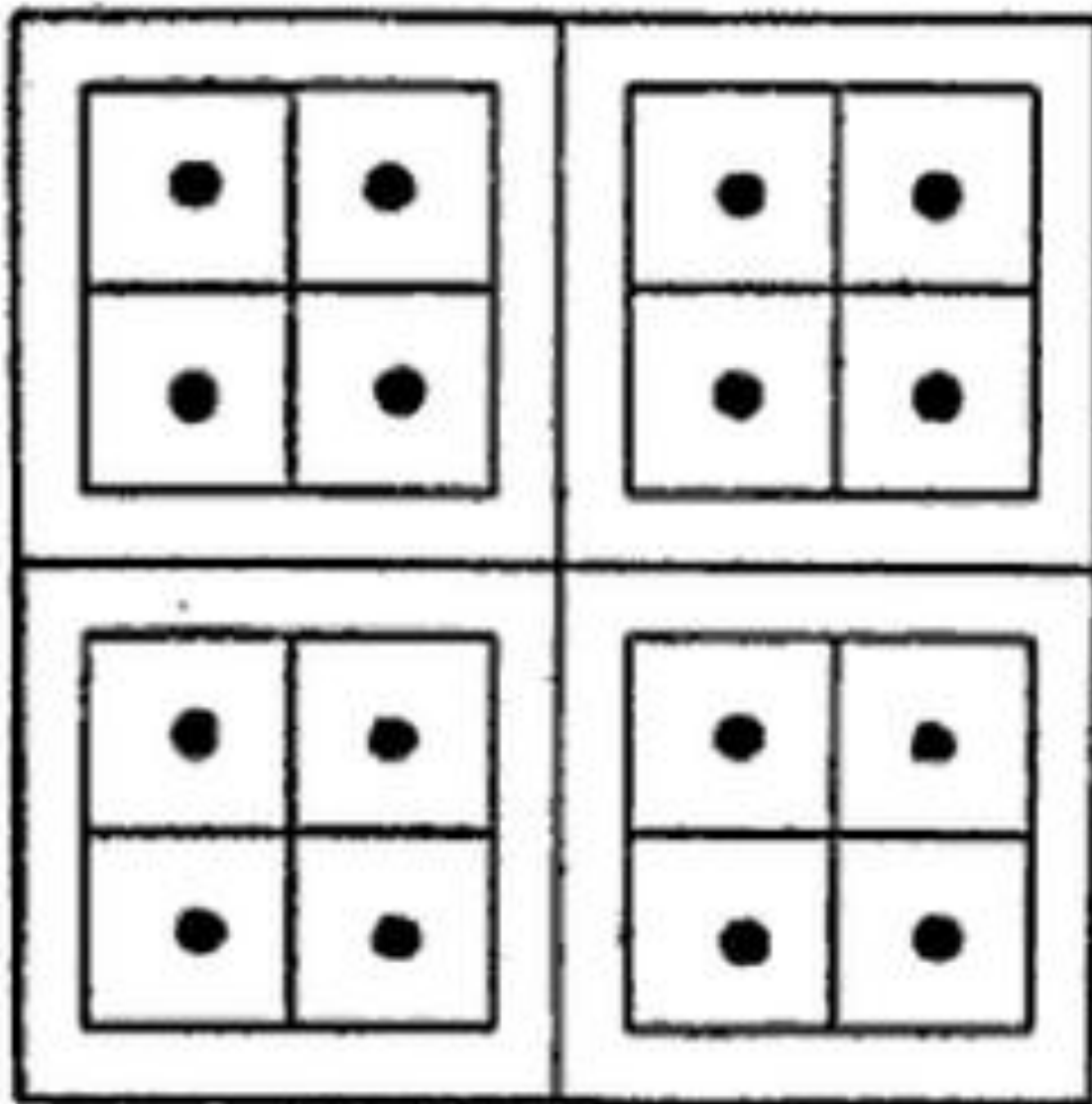
Stimuli from the original Binet-Simon Test
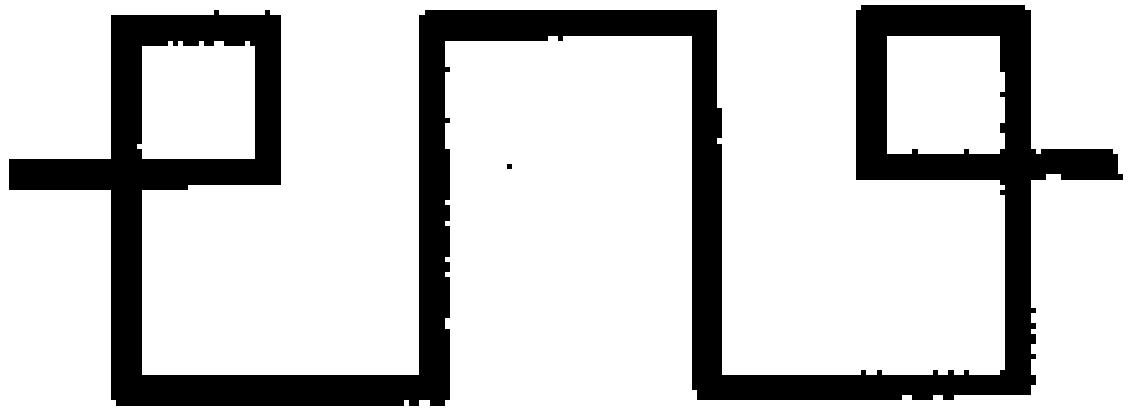
Stimulus from the Binet-Simon test later seen in the Wechsler Memory Scale (Binet, 1916; Psychological Corporation, 2009).
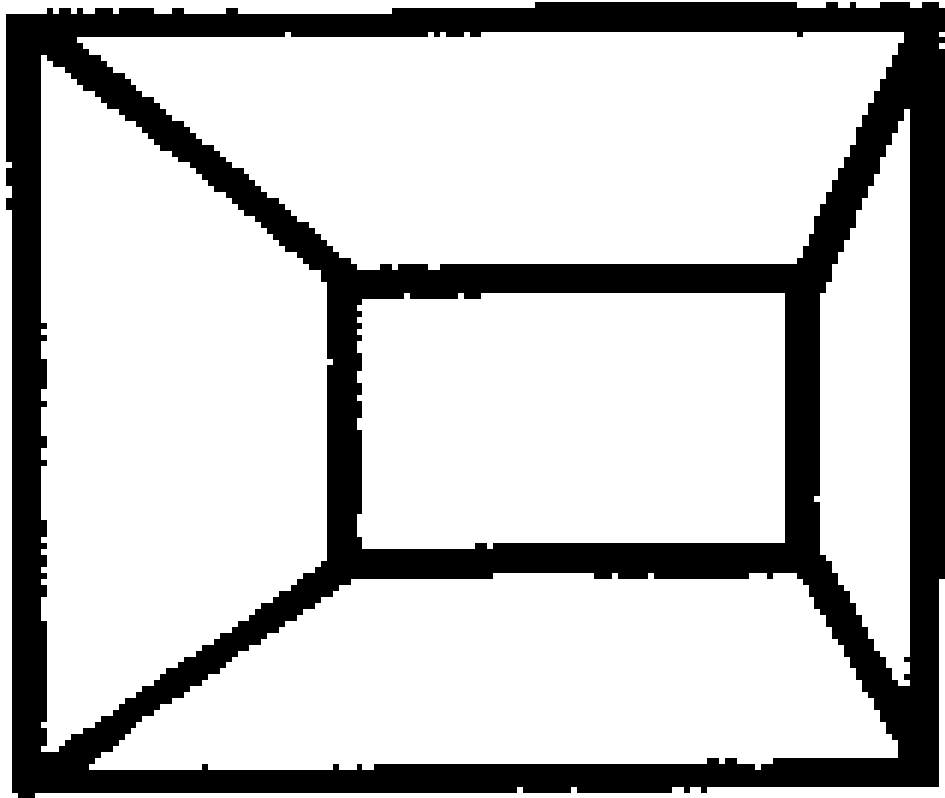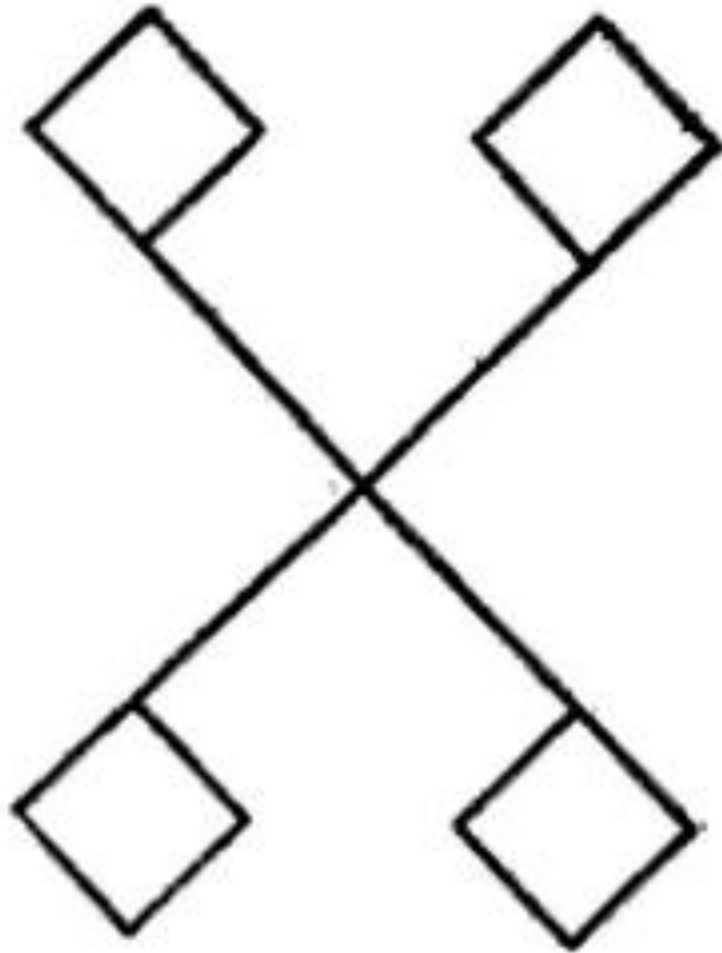
Stimulus from the Binet-Simon test later seen in the Wechsler Memory Scale (Binet, 1916; Psychological Corporation, 2009).

APPENDIX B

STIMULI FROM THE ARMY ALPHA AND ARMY BETA TESTS

Stimulus from the Army Alpha and Army Beta tests later seen in the WMS-IV (Yerkes, 1921; Psychological Corporation, 2009).

Stimulus from the Army Alpha and Army Beta tests later seen in the WMS-IV (Yerkes, 1921; Psychological Corporation, 2009).