The Pennsylvania State University

The Graduate School

College of Education

VALIDITY OF DIAGNOSTIC DECISION MAKING IN A SCHOOL PSYCHOLOGY TRAINING CLINIC

A Dissertation in

School Psychology

by

Megan M. Gammelmo

© 2015 Megan M. Gammelmo

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

August 2015
The dissertation of Megan Gammelmo was reviewed and approved* by the following:

Robert Hale  
Professor of Education (Educational & School Psychology)  
Dissertation Advisor  
Chair of Committee

Shirley Woika  
Associate Professor of Education (School Psychology)

Liza Conyers  
Associate Professor of Education (Rehabilitation and Human Services)

Marianne Hillemeier  
Professor of Health Policy and Administration and Demography

Barbara Schaefer  
Program Coordinator, Associate Professor of Education

* Signatures on file with the Graduate School
ABSTRACT

The present study addresses the nature of diagnostic decision making in a school psychology training clinic. The use of psychoeducational testing and ethical practice is necessary for the reliable and valid placement of students into diagnostic classifications. When making decisions, school psychologists must consider a myriad of information from a number of sources. Variables such as overall intellectual functioning, academic achievement, rating scale scores, referral questions, developmental delays, and existing medical conditions are expected to be weighted into a school psychologist’s decision to classify students in many disability categories. Variables such as race, age, sex, and where the evaluation was completed are not expected to be weighted or used by school psychologists. Through the use of discriminant analysis, it was determined that overall intellectual functioning, achievement scores, and referral question were used by school psychologists, while other available variables were not used in the decision making process.
# TABLE OF CONTENTS

List of Tables ...................................................................................................................... v
List of Figures ..................................................................................................................... vi

Chapter I: Introduction ........................................................................................................ 1
  Why is Classification Important ..................................................................................... 2

Chapter II: Literature Review ............................................................................................. 5
  Classification ..................................................................................................................... 5
  Psychoeducational Testing .............................................................................................. 12
  Types of Validity ............................................................................................................. 20
  Group Representation Bias ............................................................................................ 23
  Decision Making ............................................................................................................. 29
  Types of Decision Making in School Psychology ......................................................... 32
  Sources of Bias in Decision Making ............................................................................. 34
  Key Studies ..................................................................................................................... 35
  Research Questions and Hypotheses ........................................................................... 39

Chapter III: Method ............................................................................................................ 40
  Participants ..................................................................................................................... 40
  Procedure ....................................................................................................................... 40
  Data Analysis ................................................................................................................ 41

Chapter IV: Results ............................................................................................................. 43

Chapter V: Discussion ......................................................................................................... 47
  Limitations and Future Directions ............................................................................... 48

References ......................................................................................................................... 57
List of Tables

Table 1: Descriptive Statistics of Discriminant Analysis Independent Variables and Diagnosis ........51

Table 2: Discriminant Analysis Summary Table .................................................................52

Table 3: Significant Function Structure Matrix .................................................................53

Table 4: Cross-Tabulation of Classification Results .........................................................54

Table 5: Discriminant Analysis Classification Results .......................................................55
List of Figures

Figure 1: Canonical Discriminant Functions of Diagnostic Categories………………………56
Chapter I: INTRODUCTION

According to the National Association of School Psychologists (NASP; n.d.) a school psychologist is a professional who "help[s] children and youth succeed academically, socially, behaviorally, and emotionally. They collaborate with educators, parents, and other professionals to create safe, healthy, and supportive learning environments that strengthen connections between home, school, and the community for all students" (p. 1). In order to achieve this goal, school psychologists engage in a variety of activities. Chief among these is psychoeducational testing. A school psychologist spends approximately half (48%) of their work day engaging in testing (Bramlett, Murphy, Johnson, Willingsford, & Hall, 2002). Psychoeducational testing, diagnostic classification, and educational placements are a large and very important part of what a school psychologist does. The process of psychoeducational testing can be a life altering, positive or negative, process for students. School psychologists must assure the process of psychoeducational testing follows ethical and valid practices.

School psychologists are thoroughly trained in the background, use, and interpretation of many psychological and educational measures. Emphasis is placed on assuring that the measures chosen for use are reliable and valid for the purposes for which they are used (e.g., measuring intelligence, achievement, or behavior). In order for school psychologists to be able to arrive at valid classifications for students with whom they are working, they must follow a rigorous progression of clinical training. Training programs in school psychology approach this rigorous progression in differing manners.

First, for the training program used in the current study, school psychologists in training begin by observing case planning, testing, and decision making procedures. They complete coursework centered around representative assessments and focus on individual mastery of skills
and fluency of skills. During this coursework, school psychologists in training are provided direct instruction on the decision making process (e.g., what should be included, testing scores or background information, and what should not, race or gender). They engage in didactic seminars in which they observe and eventually participate in demonstrating case planning and instrument selection, testing complications, and arrival at an eligibility decision. School psychologists then begin working with their own clients; at first with peer help and heavy supervision. Eventually, when mastery of skills is demonstrated, school psychologists in training begin to work more independently. At this point, diagnostic or eligibility decisions are being made individually (with minimal supervision) or in the context of multidisciplinary teams (MDT). This culminating clinical experience prepares those in training to be independently practicing school psychologists.

**Why is Classification Important**

A major outcome of formal assessment and the testing process is to determine an appropriate educational disability classification. The benefits of classification are debated. Proponents of classification argue that classification allows for quick communication between professionals, organization of an individual’s complex behaviors, and access to services specific to a disability (Lander, 2010; Sattler, 2008). Those who argue against such a system argue that diagnoses can be stigmatizing, do little to help the client improve, provide an excuse for manageable behaviors or symptoms (Lander, 2010; Sattler, 2008), and are also fraught with issues of bias.

Classification of students into diagnostic categories involves both the use of testing instruments and human judgment. Human judgment must come into play in the evaluation of students by necessity. The research literature has repeatedly shown that human judgment may be
profitably used as an input variable in the classification process, but should never be used as the mechanism to form the classifications (Waller, Yonce, Grove, Faust, & Lenzenweger, 2006). Therefore, human judgment is one method of creating variables used to form diagnostic categories. The use of human judgment in the diagnostic process should end prior to the final diagnostic step. Once all information is gathered by use of multiple methods, including human judgment, empirical or mechanical techniques should then be used to determine final diagnostic classifications (Waller et al., 2006). Since 1927, the research literature has consistently indicated that the mechanical process of combining test information to arrive at a diagnosis is valid and more sound than human judgment (Dawes, Faust, & Meehl, 1989). In this, the actuarial approach, decisions are made solely through the empirical relationship of predictor to outcome variables. Indeed, adding human judgment to the act of forming a diagnosis decreases the reliability and validity of the diagnostic process (Dawes et al., 1989).

School psychology training does not solely focus on the mechanics of testing or the mechanics of data combination to arrive at a diagnosis, but it also includes clinical judgment in the process. Rules and policies almost enforce a diagnostic methodology where human judgment is used as part of the final step in considering and combining the information to make a diagnostic classification for a student. In practice, actuarial and human judgments are used in tandem to arrive at decisions (Dawes et al., 1989). Multidisciplinary team meetings, an important part of PL 94-142 and the Individuals with Disabilities Education Improvement Act, provide a convenient component in the diagnostic process where clinical judgment can be introduced and which, as indicated previously, will necessarily decrease the validity of the diagnostic process. It is therefore important to tease out the influence of human judgment from the instrument mechanic. This research is designed to see if variables that come from human
judgment and are obviously not appropriate in the differential diagnosis of students (e.g., race, age, gender) are being used in the diagnosis and classification of students when they are evaluated at a training program for school psychologists at a major university.
Chapter II: LITERATURE REVIEW

Classification

Diagnostic classification in educational settings has both negative and positive aspects. Although there are negatives to diagnoses, the benefits outweigh the costs; especially when diagnoses are used consistently and appropriately (Skiba, Knesting, & Bush, 2002). One of the benefits to classification in public schools is that it allows disabled and gifted students to receive appropriate services. However, caution is needed when placing disabled students in special education programs because once a student enters special education, he/she is more than likely going to be in special education for the duration of his/her educational experience (Huebner & Cummings, 1985). In fact, exit rates from special education range from only four to 12% (Hayes, 2006). There is also some evidence that special education placement may be stigmatizing (Sattler, 2008). Adults and peers can more readily affix stereotypes to the labeled individual which can lead to discrimination. These students are at risk of being ostracized by peers and ridiculed due to their differences. Of course, if a student is wrongly diagnosed and is placed in a special education program, the benefits of classification decrease and the potential risks greatly increase.

Classification systems with clear and objective rules are the most likely to produce more reliable and valid assignment (Sattler, 2008). In order to assist professionals in making the most accurate classification possible, the definitional criteria set forth by the Individuals with Disabilities Education Improvement Act (IDEA, 2004) and the definitional criteria outlined in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; American Psychological Association, 2013) are often used. The classification categories that are provided by these two systems are approved for financial support from insurance companies and
departments of special education. If a school psychologist were to use some other classification category, other than those outlined by these two systems, a client would not likely receive any services because that “new classification” would not be recognized. According to the IDEIA, there are 13 categories of exceptionalities available to school age children and approximately 12% of the U.S. school-aged population is served through one or more of these categories (Kamphaus et al., 2000).

Below are the definitions of the 13 disability categories as provided by the IDEIA (IDEIA, 2004).

(1)

(i) Autism means a developmental disability significantly affecting verbal and nonverbal communication and social interaction, generally evident before age three, that adversely affects a child's educational performance. Other characteristics often associated with autism are engagement in repetitive activities and stereotyped movements, resistance to environmental change or change in daily routines, and unusual responses to sensory experiences.

(ii) Autism does not apply if a child's educational performance is adversely affected primarily because the child has an emotional disturbance, as defined in paragraph (c)(4) of this section.

(iii) A child who manifests the characteristics of autism after age three could be identified as having autism if the criteria in paragraph (c)(1)(i) of this section are satisfied.

(2) Deaf-blindness means concomitant hearing and visual impairments, the combination of which causes such severe communication and other developmental and educational
needs that they cannot be accommodated in special education programs solely for children with deafness or children with blindness.

(3) Deafness means a hearing impairment that is so severe that the child is impaired in processing linguistic information through hearing, with or without amplification that adversely affects a child's educational performance.

(4)

(i) Emotional disturbance means a condition exhibiting one or more of the following characteristics over a long period of time and to a marked degree that adversely affects a child's educational performance:

(A) An inability to learn that cannot be explained by intellectual, sensory, or health factors.

(B) An inability to build or maintain satisfactory interpersonal relationships with peers and teachers.

(C) Inappropriate types of behavior or feelings under normal circumstances.

(D) A general pervasive mood of unhappiness or depression.

(E) A tendency to develop physical symptoms or fears associated with personal or school problems.

(ii) Emotional disturbance includes schizophrenia. The term does not apply to children who are socially maladjusted, unless it is determined that they have an emotional disturbance under paragraph (c)(4)(i) of this section.

(5) Hearing impairment means an impairment in hearing, whether permanent or fluctuating, that adversely affects a child's educational performance but that is not included under the definition of deafness in this section.
(6) Mental retardation means significantly subaverage general intellectual functioning, existing concurrently with deficits in adaptive behavior and manifested during the developmental period, that adversely affects a child's educational performance.

(7) Multiple disabilities means concomitant impairments (such as mental retardation-blindness or mental retardation-orthopedic impairment), the combination of which causes such severe educational needs that they cannot be accommodated in special education programs solely for one of the impairments. Multiple disabilities does not include deaf-blindness.

(8) Orthopedic impairment means a severe orthopedic impairment that adversely affects a child's educational performance. The term includes impairments caused by a congenital anomaly, impairments caused by disease (e.g., poliomyelitis, bone tuberculosis), and impairments from other causes (e.g., cerebral palsy, amputations, and fractures or burns that cause contractures).

(9) Other health impairment means having limited strength, vitality, or alertness, including a heightened alertness to environmental stimuli, that results in limited alertness with respect to the educational environment, that--

(i) Is due to chronic or acute health problems such as asthma, attention deficit disorder or attention deficit hyperactivity disorder, diabetes, epilepsy, a heart condition, hemophilia, lead poisoning, leukemia, nephritis, rheumatic fever, sickle cell anemia, and Tourette syndrome; and

(ii) Adversely affects a child's educational performance.

(10) Specific learning disability. (i) General. Specific learning disability means a disorder in one or more of the basic psychological processes involved in understanding or in using
language, spoken or written, that may manifest itself in the imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations, including conditions such as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia.

(ii) Disorders not included. Specific learning disability does not include learning problems that are primarily the result of visual, hearing, or motor disabilities, of mental retardation, of emotional disturbance, or of environmental, cultural, or economic disadvantage.

(11) Speech or language impairment means a communication disorder, such as stuttering, impaired articulation, a language impairment, or a voice impairment, that adversely affects a child's educational performance.

(12) Traumatic brain injury means an acquired injury to the brain caused by an external physical force, resulting in total or partial functional disability or psychosocial impairment, or both, that adversely affects a child's educational performance. Traumatic brain injury applies to open or closed head injuries resulting in impairments in one or more areas, such as cognition; language; memory; attention; reasoning; abstract thinking; judgment; problem-solving; sensory, perceptual, and motor abilities; psychosocial behavior; physical functions; information processing; and speech. Traumatic brain injury does not apply to brain injuries that are congenital or degenerative, or to brain injuries induced by birth trauma.

(13) Visual impairment including blindness means an impairment in vision that, even with correction, adversely affects a child's educational performance. The term includes both partial sight and blindness. (IDEIA, 2004)
If a student meets definitional criteria for one of the above categories, he/she must meet a second criterion to be considered eligible for special education services. For each category, the student’s educational performance must also be significantly affected. In order to achieve this goal, there are typical procedures used to arrive at disability decisions. These include behavioral observations, review of records, interviews, rating scales, direct assessment (e.g., achievement, intelligence, curriculum based assessment; CDC, 2015; Zirkel, 2011). Other professionals, such as an occupational therapist or speech language therapist or medical doctor, can also be utilized during assessment.

Given that not all clients are seeking services through their school and that the IDEIA’s definitions are very short, the DSM-5 (2013; the DSM-IV-TR used prior to 2013) is also used by school psychologists to aid in determining disabilities. Many of the DSM-5 diagnoses are used when working with school age children. At times, the DSM-5 is used to aid in determining whether a student meets definitional criteria and can be categorized under IDEIA. Examples of disorders that may fall into this category are major depressive disorder, attention deficit hyperactivity disorder, generalized anxiety disorder, and obsessive compulsive disorder. DSM-5 (2013) criteria for autism are illustrated below.

A. Persistent deficits in social communication and social interaction across multiple contexts, as manifested by the following, currently or by history (examples are illustrative, not exhaustive; see text):

1. Deficits in social-emotional reciprocity, ranging, for example, from abnormal social approach and failure of normal back-and-forth conversation; to reduced sharing of interests, emotions, or affect; to failure to initiate or respond to social interactions.
2. Deficits in nonverbal communicative behaviors used for social interaction, ranging, for example, from poorly integrated verbal and nonverbal communication; to abnormalities in eye contact and body language or deficits in understanding and use of gestures: to a total lack of facial expressions and nonverbal communication.

3. Deficits in developing, maintaining, and understanding relationships, ranging, for example, from difficulties adjusting behavior to suit various social contexts; to difficulties in sharing imaginative play or in making friends; to absence of interest in peers. Specify current severity: Severity is based on social communication impairments and restricted, repetitive patterns of behavior.

B. Restricted, repetitive patterns of behavior, interests, or activities, as manifested by at least two of the following, currently or by history (examples are illustrative, not exhaustive; see text):

1. Stereotyped or repetitive motor movements, use of objects, or speech (e.g., simple motor stereotypes, lining up toys or flipping objects, echolalia, idiosyncratic phrases).

2. Insistence on sameness, inflexible adherence to routines, or ritualized patterns of verbal or nonverbal behavior (e.g., extreme distress at small changes, difficulties with transitions, rigid thinking patterns, greeting rituals, need to take same route or eat same food every day).

3. Highly restricted, fixated interests that are abnormal in intensity or focus (e.g., strong attachment to or preoccupation with unusual objects, excessively circumscribed or perseverative interests).

4. Hyper- or hyporeactivity to sensory input or unusual interest in sensory aspects of the environment (e.g., apparent indifference to pain/temperature, adverse response to specific
sounds or textures, excessive smelling or touching of objects, visual fascination with lights or movement). Specify current severity: Severity is based on social communication impairments and restricted, repetitive patterns of behavior (see Table 2).

C. Symptoms must be present in the early developmental period (but may not become fully manifest until social demands exceed limited capacities, or may be masked by learned strategies in later life).

D. Symptoms cause clinically significant impairment in social, occupational, or other important areas of current functioning.

E. These disturbances are not better explained by intellectual disability (intellectual developmental disorder) or global developmental delay. Intellectual disability and autism spectrum disorder frequently co-occur; to make comorbid diagnoses of autism spectrum disorder and intellectual disability, social communication should be below that expected for general developmental level. (pp. 50-51)

**Psychoeducational Testing**

Psychoeducational testing is often done as a part of a formal assessment process completed by a school psychologist (Hale et al., 2001) in order to gather information to make decisions regarding the client (Sattler, 2008). In order for a school-aged student to receive services in school, he or she must be assessed by a qualified professional. Examples of individuals meeting the threshold as a qualified professional include school psychologists, speech language pathologists, or remedial reading teachers (IDEIA, 2004). In addition, if the service is to be financially supported, the student must be given a recognized classification.

Given the large number of students being assessed for special education services, testing is an increased necessity (Kamphaus, Petoskey, & Rowe, 2000). In almost all special education
referrals, norm-referenced testing is completed as a part of the evaluation (Sattler, 2008). This includes IQ tests, achievement tests, and rating scales. Tests and rating scales are important in differentiating students in many exceptionality categories. In the following sections, some of the most important and frequently used tests and scales used by school psychologists are discussed. These will be the tests and scales that are expected to be used by the psychologists in training in the current study.

**Frequently used rating scales.** In order to obtain a standardized measure of a student’s functioning in many settings, rating scales are often administered to a student’s parents and teachers, as well as to the student. Rating scales can be broken down into two general types: broadband and specific. They are used typically to evaluate emotional, behavioral and/or adaptive functioning.

**Broadband rating scales.** Broadband rating scales are designed to get a comprehensive measure of a student’s emotional, behavioral, or adaptive functioning. The Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 2004) is often used to this end.

The BASC (Reynolds & Kamphaus, 2004) is on its second edition and can be administered to parents, teachers, and students. Depending on the respondent and the age of the student, many scale scores can be obtained. The parent and teacher forms are for students aged two to 21 and provide information on Activities of Daily Living, Adaptability, Aggression, Anxiety, Attention Problems, Atypicality, Conduct Problems, Depression, Functional Communication, Hyperactivity, Leadership, Learning Problems, Social Skills, Somatization, Study Skills, and Withdrawal. The self-report version is designed for students aged 6 through college age and provides information on Alcohol Abuse, Anxiety, Attention Problems, Attitude to School, Attitude to Teachers, Atypicality, Depression, Hyperactivity, Interpersonal Relations,
Locus of Control, Relation with Parents, School Maladjustment, Self-Esteem, Self-Reliance, Sensation Seeking, Sense of Inadequacy, Social Stress, and Somatization.

**Adaptive behavior rating scales.** For adaptive functioning, two scales are often used: the Vineland Adaptive Behavior Scales – Second Edition (Vineland; Sparrow, Cicchetti, & Balla, 2005) and the Adaptive Behavior Assessment System – Second Edition (ABAS; Harrison & Oakland, 2003). The Vineland provides information regarding Communication (receptive, expressive, written), Daily Living Skills (personal, domestic, community), Socialization (interpersonal relationships, play and leisure time, coping skills, Motor Skills (fine, gross), and Maladaptive Behaviors (internalizing, externalizing, other). Both the Vineland and the ABAS can be used for individuals from birth to adulthood. The ABAS provides information in the areas of Conceptual skills, Practical skills, Community Use, Home Living, Self-Care, Social, Functional Academics, Community, Leisure, Health and Safety, Self-Direction, and Work.

**Disorder specific rating scales.** Specific rating scales are designed to measure only one area of functioning for a student. These scales are given when a specific area of concern is discovered. These scales may be given to help assess a variety of issues, such as anxiety, depression, autism, or ADHD. In the area of anxiety, scales such as the Revised Children’s Manifest Anxiety Scale (RCMAS; Reynolds & Richmond, 2000) or the Beck Anxiety Inventory (BAI; Beck & Steer, 1993), among others, are often given.

For depression the Beck Depression Inventory (BDI; Beck et al., 1961) or the Children’s Depression Inventory 2 (CDI 2; Kovacs, 2010) are often used. For Autism, rating scales frequently used include the Autism Spectrum Rating Scales (ASRS; Goldstein & Naglieri, 2009) and the Childhood Autism Rating Scale, Second Edition (CARS2; Schopler, Van Bourgondien,
Wellman, & Love, 2010). The Conners 3 ADHD Index (Conners, 2008) is often used to measure levels of ADHD symptoms.

**Frequently used IQ and achievement tests.** Intelligence and achievement tests are norm referenced tests that allow professionals to compare a client’s performance to a nationally representative sample (Lander, 2010). Through testing, a client’s intellectual ability or level of academic achievement, at that specific point in time, can be directly, rather than indirectly, ascertained (Haney, 1981; Lander, 2010). Given the important nature of intelligence and achievement testing, tests with high reliability estimates are required. A school psychologist’s training emphasizes this point and indicates standards that should be followed. For individual decision making, reliability estimates must meet a minimum threshold of .90 (Sattler, 2008). Any psychometric test that does not meet this high standard is not recommended for use for individual decision making. While there are many different intelligence and achievement tests available, only a few have become the standards of choice. These are discussed below, along with their reliability estimates.

The Wechsler scales are the most commonly used instruments to measure intelligence and academic achievement (Brooks, 2010; Brown & Morgan, 1991; Kamphaus et al., 2000). For use with school age populations, three Wechsler intelligence tests are most frequently used: Wechsler Intelligence Scale for Children (WISC; Wechsler, 2004, 2014), Wechsler Preschool and Primary Scales of Intelligence (WPPSI; Wechsler, 2012), and the Wechsler Adult Intelligence Scales (WAIS; Wechsler, 2008). The WISC is currently on its fifth edition (Wechsler, 2014) and is designed for use with children ages six to 16. Despite the WISC being on its fifth edition, information from the fourth edition (Wechsler, 2004) is presented below due to the fact that the fifth edition was not used in any of the evaluations in the present study given
its recent release date. The WISC-IV consists of 10 core subtests that yield an overall IQ (FSIQ) and four index scores. The subtests are Block Design, Similarities, Digit Span, Picture Concepts, Coding, Vocabulary, Letter-Number Sequencing, Matrix Reasoning, Comprehension, and Symbol Search. Supplemental subtests include Picture Completion, Cancellation, Information, Arithmetic, and Word Reasoning. The index scores are the Working Memory Index ($\alpha = .92$), Processing Speed Index ($\alpha = .88$), Verbal Comprehension Index ($\alpha = .94$), and the Perceptual Reasoning Index ($\alpha = .92$). The internal reliability coefficient of the FSIQ is .97.

The WPPSI (Wechsler, 2012) is currently on its fourth edition and is designed for use with children aged two and a half to seven years, seven months. For children up to age three, there are six core subtests that yield a FSIQ and three index scores. These subtests include Receptive Vocabulary, Picture Naming, Block Design, Object Assembly, Picture Memory, and Zoo Locations. The Primary Index Scales are the Verbal Comprehension Index ($\alpha = .94$), the Visual Spatial Index ($\alpha = .89$), and the Working Memory Index ($\alpha = .93$). For the younger age group, the FSIQ has an internal consistency estimate of .96. For the children four and older, there are 10 core subtests that yield a FSIQ and five index scores. The subtests are Information, Similarities, Block Design, Object Assembly, Matrix Reasoning, Picture Concepts, Picture Memory, Zoo Locations, Bug Search, and Cancellation. The five index scores include Verbal Comprehension ($\alpha = .94$), Visual Spatial ($\alpha = .90$), Fluid Reasoning ($\alpha = .93$), Working Memory ($\alpha = .91$), and Processing Speed ($\alpha = .86$). The FSIQ scores’ internal consistency estimates for this age group are .96.

The WAIS (Wechsler, 2008) is also on its fourth edition and is designed for people over the age of 16. A FSIQ and a General Ability score are calculated using 10 subtests, in addition to four other index scores. The subtests of the WAIS are Similarities, Vocabulary, Information,
Digit Span, Arithmetic, Block Design, Matrix Reasoning, Visual Puzzles, Symbol Search, and Coding. The index scores include: the Processing Speed Index (α = .90), the Working Memory Index (α = .94), the Verbal Comprehension Index (α = .96), and the Perceptual Reasoning Index (α = .95). The FSIQ score derived from the WAIS has an internal consistency estimate of .98.

The Wechsler Individual Achievement Test – Third Edition (WIAT-III; Wechsler, 2009) is the most frequent achievement test that accompanies the three Wechsler intelligence scales. The WIAT is designed for use with students from preschool to 12th grade (an adult/college supplement is also available) and includes 16 subtests. It also yields a variety of information about a student’s academic achievement. The subtests are Listening Comprehension, Early Reading Skills, Reading Comprehension, Math Problem Solving, Alphabet Writing Fluency, Sentence Composition, Word Reading, Essay Composition, Pseudoword Decoding, Numerical Operations, Oral Expression, Oral Reading Fluency, Spelling, and Math Fluency in Addition, Subtraction, and Multiplication. The WIAT scoring allows for several composite scores: Oral Language (α = .91), Total Reading (α = .97), Basic Reading (α = .98), Reading Comprehension and Fluency (α = .92), Written Expression (α = .94), Mathematics (α = .96), and Math Fluency (α = .94). The WIAT’s overall Achievement score has an internal consistency estimate of .98.

Other measures in the Wechsler scales include the Wechsler Abbreviated Scale of Intelligence (WASI), Wechsler Memory Scales (WMS), Wechsler Nonverbal (WNV), and several others.

The Kaufman series is also used to assess student’s intelligence and achievement skills. The Kaufman Assessment Battery for Children (KABC; Kaufman & Kaufman, 2004) is on its second edition and is designed for children aged three to 18. It offers 20 subtests that measure five broad components of intellectual functioning. The subtests for the KABC are Triangles, Face Recognition, Pattern Reasoning, Block Counting, Conceptual Thinking, Story Completion,
Rover, Gestalt Closure, Word Order, Number Recall, Hand Movements, Pattern Reasoning, Story Completion, Atlantis, Atlantis Delayed, Rebus, Rebus Delayed, Riddles, Expressive Vocabulary, and Verbal Knowledge. These subtests can be used in many combinations to create five broad ability scores. These broad abilities are Visual Processing ($\alpha = .92$), Short-Term Working Memory ($\alpha = .91$), Fluid Reasoning ($\alpha = .88$), Long-Term Memory ($\alpha = .91$), and Crystalized Intelligence ($\alpha = .90$). An overall intelligence score can be obtained as well ($\alpha = .95$).

The Kaufman Test of Educational Achievement (KTEA; Kaufman & Kaufman, 2014) is designed for use with individuals four to 25. It offers 19 subtests designed to tap an individual’s reading ($\alpha = .95$), math ($\alpha = .97$), and written achievement ($\alpha = .94$). These subtests are Phonological Processing, Math Concepts & Applications, Letter & Word Recognition, Math Computation, Nonsense Word Decoding, Writing Fluency, Silent Reading Fluency, Math Fluency, Reading Comprehension, Written Expression, Associational Fluency, Spelling, Object Naming Facility, Reading Vocabulary, Letter Naming Facility, Listening Comprehension, Word Recognition Fluency, Oral Expression, and Decoding Fluency. In addition to reading, writing, and math, the KTEA provides supplemental scores for Sound-Symbol ($\alpha = .96$), Decoding ($\alpha = .98$), Reading Fluency ($\alpha = .93$), Reading Understanding ($\alpha = .94$), Oral Language ($\alpha = .86$), Oral Fluency ($\alpha = .72$), Comprehension ($\alpha = .92$), Expression ($\alpha = .89$), Orthographic Processing ($\alpha = .91$), and Academic Fluency ($\alpha = .91$). An overall achievement score can also be obtained ($\alpha = .98$).

The Kaufman series also offers instruments such as the Kaufman Survey of Early Academic and Language Skills (KSEALS), the Kaufman Brief Intelligence Test (KBIT), and the Kaufman Short Neurological Assessment Procedure (KSNAP).
The Woodcock-Johnson IV Test of Cognitive Abilities (WJ-IV Cog; Schrank, Mather, & McGrew, 2014) allows administration of 18 subtests: Oral Vocabulary, Number Series, Verbal Attention, Letter-Pattern Matching, Phonological Processing, Story Recall, Visualization, General Information, Concept Formation, Numbers Reversed, Number-Pattern Matching, Nonword Repetition, Visual-Auditory Learning, Picture Recognition, Analysis-Synthesis, Object-Number Sequencing, Pair Cancellation, and Memory for Words. These subtests produce scores for the seven broad areas of intellectual functioning. These areas are Crystalized Intelligence ($\alpha = .93$), Fluid Intelligence ($\alpha = .94$), Auditory Processing ($\alpha = .92$), Processing Speed ($\alpha = .94$), Short-Term Working Memory ($\alpha = .91$), Long-Term Storage and Retrieval ($\alpha = .97$), and Visual Processing ($\alpha = .86$). The WJ-IV Cog also produces an overall intellectual functioning score ($\alpha = .97$).

The Woodcock-Johnson IV Tests of Achievement (Schrank et al., 2014) contains 11 subtests in the standard battery (nine additional in the extended battery) and can be used to obtain 22 cluster scores. The subtests are Letter-Word Identification, Applied Problems, Spelling, Passage Comprehension, Calculation, Writing Samples, Word Attack, Oral Reading, Sentence Reading Fluency, Math Facts Fluency, and Writing Fluency. Cluster scores in the areas of Reading ($\alpha = .95$), Broad Reading ($\alpha = .97$), Basic Reading Skills ($\alpha = .95$), Reading Comprehension ($\alpha = .93$), Reading Fluency ($\alpha = .96$), Mathematics ($\alpha = .96$), Broad Mathematics ($\alpha = .97$), Math Calculation Skills ($\alpha = .97$), Math Problem Solving ($\alpha = .95$), Written Language ($\alpha = .94$), Written Expression ($\alpha = .92$), and Broad Achievement ($\alpha = .99$) can be obtained in addition to several others.

Testing using the reliable scales listed above is completed to determine eligibility for special education services (Lander, 2010). Given that only the FSIQ or overall intelligence
scores and the subject area composite scores from the achievement tests have consistent reliability coefficients over the recommended .90, only FSIQ and composite achievement scores are included in this study. The index scores did not meet the stringent .90 cutoffs, thus, could not be included. In order to be useful, these psychological tests must demonstrate that in addition to being reliable, they must also be valid. The assessment of a test scores’ validity is more complex than evaluating its reliability. The psychoeducational tools discussed in the previous section have evidence presented for various types of validity in their manuals. Although this is very important information, studies completed by researchers not affiliated with the tests are able to provide more compelling validity evidence (e.g., Canivez, 2014; Mortimore, 2012; Reynolds, Keith, Fine, Fisher, & Low, 2007).

Types of Validity

Validity, measuring what is meant to be measured, is of vital importance in the context of psychoeducational testing. Without validity, it would be inappropriate to draw inferences based on test scores (Sattler, 2008). In other words, determining a student has an educational disability using a test with invalid scores would be incongruous and would likely lead to poor decision making in the schools. Validity is necessary to ensure the high quality services that children deserve (Kamphaus et al., 2000). Since the 1930s, the validity of testing instruments has been a focus (Haney, 1981). In the context of psychoeducational testing; content, construct, and criterion-related validity are most important and necessary to obtain useful data to be used in decision making.

Content validity addresses whether items on a test are representative of what is being measured (Cronbach & Meehl, 1955; Sattler, 2008) and is often determined deductively (Cronbach & Meehl, 1955). To establish content validity, experts are often implored to
determine whether the items on a test appear to be measuring the proper content. A test that demonstrates content validity will show no item difficulty difference between groups. For example, studies have shown that content bias generally accounts for a small and insignificant percentage of the variance in scores (Reynolds, 2000). Overwhelmingly, the conclusion is that the achievement and IQ tests discussed have demonstrated content validity (Sattler, 2008). Many of the more popular rating scales also have research indicating content validity (e.g., BDI; Beck, Steer, & Brown, 2012).

Construct validity has been a concern in the psychological community since the inception of intellectual testing (Reynolds, 2000) and can be the most difficult to measure and maintain (Cronbach & Meehl, 1955). It refers to how well a test measures a predetermined construct (Sattler, 2008) and relies on causal inferences of measurement during evaluation (Teglasi, Nebbergall, & Newman, 2012). For example, on a test of intelligence, a test with good construct validity would in fact measure intelligence. The construct of intelligence is present and seemingly non-changing; the measure must be a valid enough instrument to accurately be measuring intelligence (Teglasi et al., 2012). All other types of validity arguably feed into construct validity; making construct validity the most important and complicated type of validity (Cronbach & Meehl, 1955; Haney, 1981).

Construct validity is further divided into convergent and divergent/discriminant validity. Convergent validity is evaluated by how well two tests measuring the same construct correlate. Strong convergent validity will be evidenced by a strong positive correlation. Divergent or discriminant validity examines the relation between two tests that purport to measure different constructs (Sattler, 2008). The two tests should have a near zero correlation. This type of validity is important to testing in that it is necessary to be able to distinguish between individuals
in different classifications. Construct validity is measured most commonly by the use of factor analysis and looks at how items load on factors. Construct validity is threatened when items load on different factors for different groups of individuals (Skiba et al., 2002). One of the primary ways that construct validity is evaluated is through comparative factor analysis of an instrument across groups. Again, the achievement and IQ tests that have been discussed have demonstrated excellent construct validity.

Criterion-related validity is used to measure how well scores on a test correlate with a preset criterion (Sattler, 2008). This type of validity becomes important when the task of the psychologist is to predict an important criterion (Hunt, 1996). Scores from a test of intelligence should be used to predict a relevant criterion and that criterion should be fair to all groups who are being measured by it. Individuals taking the test should have had the opportunity to learn the skills that are being assessed (Hunt, 1996). As an example, English language learners may not have had the opportunity to learn the skill of speaking English, and it would not be fair to measure and compare them to others using a test score that depends on English language proficiency. In this case, it would also be unfair to use IQs that are also highly dependent on English language proficiency to predict achievement since these students do not yet possess the necessary prerequisite English language skills.

Criterion-related validity also includes the concepts of concurrent and predictive validity. Concurrent validity concerns the relationships between the scores on two related measures at the same point in time (Cronbach & Meehl, 1955; Sattler, 2008) and is necessary for validating new measures (Hale et al., 2001). Predictive validity examines correlations between a score on a test (the predictor) measured at the current time and the criterion measured later in time (Cronbach & Meehl, 1955; Sattler, 2008). Traditionally, intelligence tests (IQs) have been used to predict
academic achievement, income, and job performance (Lander, 2010). For school psychologists, the relationship between IQ and achievement is vital in determining underachievement.

Underachievement is one of the necessary components for the diagnosis of learning disabilities. Learning disabilities is one of the most frequent classifications given to children referred for educational difficulties. Without predictive validity, a classification of a specific learning disability could not be made. The predictive power that comes from predicative validity is used to evaluate decisions made. For instance, is the prediction about a student's success the following year valid? In terms of psychoeducational testing, it is imperative that the measures used have predictive power. Fortunately, the research literature indicates that most intelligence tests have strong predictive power when academic achievement is the criterion. Many rating scales, such as the CARS, have demonstrated evidence for criterion-related validity (Ozonoff, Boodlin-Jones, & Solomon, 2005)

A test that is determined to have construct, criterion-related, and content validity can still yield uninterruptable results. The validity of any test can be destroyed by numerous factors external to the test. For example, a student’s refusal to provide timely answers, lack of sleep, medical disorders, an examiner’s bias, breaking from standardized procedures, or other factors can render that particular test result invalid (Hambleton & Kanjee, 1995; Lopez, Lamar, & Scully-Demartini, 1997; Sattler, 2008). Test scores that are still interpreted without proper validity evidence or when that particular score’s validity has been destroyed by an external factor can lead to errors in placement in diagnostic categories.

**Group Representation Bias**

The assessment of children in educational situations almost always includes a test of intelligence (Brooks, 2010). It should be noted that the practice of always including a test of
intelligence in an assessment battery has been a matter of debate and discontent within the U.S. since the early 1900s (Hunt, 1996). While these tests themselves have been shown to be valid, their use in diagnostic decision making has led to disadvantaged groups being overrepresented (when compared to their numbers in the population) in special education, being denied academically vital services, and underrepresented in programs for gifted students (McBee, Shaunessy, & Matthews, 2012).

No bias. Kaufman (1976, as cited in Brooks, 2010) reported that the validity of the WISC-R (one of the most frequently given tests of intelligence for children between the ages of 6 and 16) was not affected by extraneous factors such as sex, age, SES, or ethnicity. Research with the WISC-III and -IV (later revisions of the WISC-R) has also replicated this finding. There is evidence to support the same construct, criterion-related, and content validity regardless of sex, age, SES, or ethnicity (Skiba, Knesting, & Bush, 2002). Thus research has been unable to consistently or conclusively show that items on the primary intelligence tests are different for different racial groups (Skiba et al., 2002).

Inequalities in educational placement. Although research indicates that intelligence tests are valid regardless of sex, age, SES, or ethnicity, it is also widely agreed that there are consistent mean differences on IQs between different groups of individuals (Hunt, 1996). It is vital to determine why these differences exist in order to determine whether information gleaned from intelligence tests is valid for use. Are these differences real and meaningful, or are the variables of SES, ethnicity, and age external variables that can or should automatically make the variables of intelligence and achievement uninterpretable? Variables such as SES, educational exposure, ethnic or cultural differences, or a health disadvantage could play a role in poor performance on intelligence and achievement tests (Hale et al., 2001; Hambleton & Kanjee,
1995). When IQs yield lower scores due to one of the previous variables, it is left to the examiner to determine the appropriate use of those results (Skiba et al., 2002). The examiner must know when the administration of a test or the resulting scores are unfair.

**Racial and cultural inequalities.** Reynolds (2000) stated that there are four possible explanations to why the four major ethnic groups (Asian, Black, Hispanic, and White) differ on their scores of intellectual ability: (a) the differences are genetically based, (b) the differences are environmentally based, (c) the differences are genetically and environmentally based, and (d) the differences in scores are due to the lack of specificity of the tests to examine minority group abilities (Reynolds, 2000). Even after the control of SES, African-American children are more likely than Caucasian children to experience long-term poverty, which leads to a significant difference in their IQs (Graham-Bergmann et al., 2010). Others argue that much of the variation between the races on measures of intellectual abilities can be accounted for by SES (McBee et al., 2012), while others maintain that SES is not a full explanation of the varying scores (Holliman, 2001). In either instance, there are systematic differences accounting for the disparate scores among minority and majority populations.

The average black student in the U.S. will score lower than 2/3 of his or her white peers on standardized tests (Hallinan, 2001). Minority students enter school in kindergarten at a disadvantage, learn slower, and do not catch up to their peers over time (Mitchell, Croy, Spicer, Frankel, & Emde, 2011). Black, Hispanic, and American Indian children tend to enter school at a lower level of intellectual functioning (Mitchell et al, 2011) and are considered to be at-risk students. Students from low SES backgrounds, regardless of race, often enter school with cognitive delays due to deprived educational environments (Mitchell et al, 2011). Students who are considered to be at-risk (i.e., low SES, racial minorities) achieve lower academically than
their not at-risk peers and have poorer overall educational prognoses (Graham-Bermann et al., 2010).

Implications of group racial differences in standardized testing reflect numerous problems within the psychological and educational systems (Frisby, 1999). Given the outcry for the reconciliation of the racial differences on intelligence tests, the courts have been forced to step in. In many of these cases, the validity of testing instruments has been called into question (Haney, 1981). For instance, in California, Larry P. v. Riles (1975) determined that African American children could not be assigned to special education programs using intelligence testing (Hunt, 1996). The U.S. Department of Education (1993) also reported that students from minority and disadvantage backgrounds were systematically overlooked for inclusion in gifted programs (McBee et al., 2012).

Students whose first language is not English are overly referred for special education services and overrepresented in classrooms reserved for students with learning disabilities, emotional or behavioral disturbances, and intellectual impairments. These students are also underrepresented in gifted classrooms (Lopez et al., 1997; McBee et al., 2012). Furthermore, these students are often administered psychological tests in a language that is not their own or via an interpreter. In both cases, the validity of the results and the consequent decisions made are affected (Hambleton & Kanjee, 1995; Lopez et al., 1997).

Reasons for differences among racial or social classes on achievement and aptitude vary, but include explanations such as: (a) school personnel provide different learning opportunities based on their outward characteristics (Hallinan, 2001), such as the under nomination and underrepresentation of African Americans and Hispanics in gifted classrooms (Jordan et al.,...
2012) and (b) schools in various areas of the U.S. have different resources and are racially and ethnically different (Hallinan, 2001).

**Gender inequalities.** Research has been sparse on the gender differences in IQ testing. Of the research that has found gender differences, the directions of those differences are not consistent (Graham-Bermann et al., 2010). Hedges and Nowell (1995) indicated that males are overrepresented at both the higher and lower ends of intellectual functioning. Females, on the other hand, do not show a difference from the normal distribution (as cited in Hunt, 1996).

Goldbeck, Daseking, Hellwig-Brida, Waldmann, and Petermann (2010) investigated sex differences on the German WISC-IV. The authors found that there was only a one point difference in FSIQ scores between males and females. This difference had a marginal effect size. Boys, however, did outperform girls on the Verbal Comprehension Index and the Perceptual Reasoning Index. Females outperformed males on the Processing Speed Index. The Working Memory Index showed no sex differences. The authors indicate that their research does not indicate separate sex norms are necessary (Goldbeck et al., 2010).

Given the scarcity and inconsistency of research on gender differences, there appears to be no legitimacy to the claim that intellectual tests are biased toward either gender (Hunt, 1996). If current research does not support the differences, how then are males overrepresented in special education?

**Medical issues.** Individuals who suffer from chronic illnesses are likely to have lower cognitive functioning when compared to their medically healthy peers (Stewart, Kennard, Waller, & Fixler, 1994). They are also far more likely to suffer academically (Bassin, Schatz, Posey, & Tapor, 2010). Kaufman (1976, as cited in Brooks, 2010) found that health status of children could influence WISC-R scores. Various medications and illnesses can masquerade as
attentional deficits, academic deficits, and cognitive deficits. Children who suffer from neurological injuries tend to have lower scores than would be predicted given their particular cognitive injury. Children with congenital heart disease can demonstrate poor performance in the areas of productivity, leisure, and self-care (Imms, 2004) which can in turn decrease academic achievement and affect ability testing. Furthermore, children who have undergone organ transplants also suffer from IQ deficits (Stewart et al., 1994). Students diagnosed with a brain tumor are likely to experience difficulties with information processing speed and executive functioning, such as organization and decision making (Barkon, 2009). It can be difficult to determine if lower scores are in fact an outcome of an injury, disease, or medication or if an outside factor has influence. These outside factors could include less of an opportunity to learn than healthy peers due to isolation, hospitalizations, peer rejection, lack of independence, and lower expectations (Stewart et al., 1994).

Given the debate regarding the fairness of standardized tests, it is necessary to ensure that a thorough assessment is conducted with cultural sensitivity (Skiba et al., 2002). Administering a culturally unbiased test is only one step in an unbiased assessment. To ensure the best and most reliable outcomes for students, the entire assessment must be unbiased and culturally competent (Skiba et al., 2002). This means that those assessing a student must be culturally competent and free of prejudices that interfere with best practice (Skiba et al., 2002). In doing so, outcomes, such as a diagnostic label, will be more likely to be accurate and fair. Given the rise in concern about the fairness of psychological testing for certain populations, professional groups, such as the American Psychological Association (APA) and the National Association of School Psychologists (NASP) have put forth guidelines for providing services to these populations (e.g., APA, 2000; APA, 2007; NASP, 2010). These guidelines lay out, among other
things, who is qualified to give and interpret psychological tests, how tests should be selected for use, how to work with special populations, and whether use in certain circumstances are valid for interpretation. So, if research has shown that there is no bias inherent in the measures of intellectual ability and achievement themselves, from where is the disproportionate representation of disadvantaged students in special education coming? Could it be that the variables of race, sex, or SES are simply being used to make these placements?

**Decision Making**

Group versus individual decision making has been a hot topic in social psychology for many years. Factors such as group think (desire of the group to maintain harmony which can lead to poor decision making), conformity (an individual’s or group’s desire to fit in or stay with social norms), or foot in the door phenomena (upon compliance with a small request, an individual is more likely to comply with a larger request without thinking critically) can affect decisions made. Decision making is a key factor in the realm of psychoeducational evaluations.

**Group decision making.** Group decision making is relevant to school psychology and psychoeducational testing in that decisions about students are made in Individual Education Plan (IEP) meetings or Multidisciplinary Team (MDT) meetings. In fact, many of the decisions made by school psychologists occur within these team meetings (Gutkin & Nemeth, 1997). School psychologists, various school personnel, parents, and sometimes students collaborate to identify problems and solutions for academic and behavioral problems (Lasser & Laurie, 2007). In IEP or MDT meetings, school psychologists and other team members make decisions that can have serious and long-term effects for students’ futures (Lasser & Laurie, 2007).

Team decision making, as opposed to individual decision making, is legally mandated by Public Law 94-142 for all decisions regarding special education placements and planning for
students (Gutkin & Nemeth, 1997; Huebner & Gould, 1991; Kaiser & Woodman, 1985). The purpose of this law was to limit the decision making power of any individual professional in the MDT process. Although these mandated MDT decisions were controversial at the outset (Huebner & Gould, 1991), they are now thought of as valued for their fair and collaborative nature (Lesser & Laurie, 2007). School psychologists are faced with balancing the thoughts and beliefs of other members, balancing dual roles, and dealing with ethical dilemmas (Lesser & Laurie, 2007). The added social work load of these MDT meetings is not always facilitative of good decision making (Kaiser & Woodman, 1985).

Group decisions can be affected by a number of outside sources. Socioemotional strife among members can lead to poor decision making (Kaiser & Woodman, 1985). Even when all group members are highly knowledgeable and invested in the outcome of a student, poor decisions can be made while the members are in agreement (Gutkin & Nemeth, 1997). Groupthink can lead members of the group to not consider all viewpoints and risks associated with a decision (Gutkin & Nemeth, 1997). Members are likely to go along with the majority viewpoint in order to avoid confrontation.

Groups can unknowingly or knowingly put social pressure on individuals which can lead to incorrect, misguided, or unethical decisions being made. Even school psychologists who adhere strictly to set ethical standards can be left without guidance due to a lack of social psychology being included in the ethical codes themselves (Lesser & Laurie, 2007). Group pressure can cause an individual to go against sound judgment to conform (Lesser & Laurie, 2007). Conforming is a way for an individual to feel less stress and discomfort, to cause fewer arguments, and to seem more likeable (Lesser & Laurie, 2007). Foot in the door phenomena, complying with a larger request following smaller ones, is another example of how one’s
judgment and decision making can be impaired. Framing of different decisions is another factor that affects what decision is made (Fagley, Miller, & Jones, 1999). Simple word choice can sway groups to make different decisions. For instance, a school psychologist could frame alternatives in a more positive light to decrease the amount of risk associated with other alternatives (Fagley et al., 1999).

School psychologists, and team members alike, should be aware that school based decision making teams could easily make more risky decisions than would be made individually (Lesser & Laurie, 2007). These risky decisions can lead to invalid diagnostic placements and ultimately consequences that adversely affect students. Therefore, it is important to take each of these things into consideration when team collaboration is being used (Fagley et al., 1999).

**Individual decision making.** Opposite group decision making is individual decision making. In the case of this study, individual decision making refers to decisions made by school psychologists (or school psychology doctoral students) alone. No MDT or IEP meetings took place in the decision making process. With individual decision making, many of the social psychological phenomena are not applicable.

It would be easy to criticize and invalidate the diagnostic decision making made by doctoral students and their supervisors in a school psychology training clinic. However, research has shown that experience is not related to the ability to make sound judgments. In fact, experienced clinicians did not make better judgments than novice graduate students (Shanteau & Stewart, 1992).

Training for many school psychologists first begins as an individual decision making model. Decisions are made in self-sufficient training clinics with little to no input from school systems and MDT meetings as a whole. As training progresses, decisions begin to be made in
group settings. It is important to determine if the variables used to arrive at diagnostic classifications are significantly different between individual and group decision making models.

**Types of Decision Making in School Psychology**

For all decision making in the field of school psychology, there are various ways to approach each diagnosis. For example, for a diagnosis of autism, some clinicians believe that the use of third party rating scales is useful (e.g., Autism Spectrum Rating Scale). Others would rather choose a more direct measure over which the clinician has more control (e.g., Childhood Autism Rating Scale, Autism Diagnostic Observation Schedule). In some counties or districts, a diagnosis of ADHD or autism must be first made by a medical professional prior to being identified in the school. This can be a point of contention if the medical professional and the school psychologist do not agree with one another. The most contentious of decision making models is the decision making process to arrive at a specific learning disability (SLD). The reauthorization of IDEIA (2004) references each of the following models either explicitly (discrepancy model and response to intervention model) or in its general language (psychological processes).

**Discrepancy model.** The discrepancy model has been used the longest of the three methods of SLD identification. In the discrepancy model, an overall intelligence score is used to predict the student’s academic achievement. The actual achievement scores are then compared to this predicted score. In order to determine a significant discrepancy, several methods have been used. The most common is the use of a two standard deviation rule. This means that if the actual score is two standard deviations below the predicted score, a discrepancy exists. Other methods look at whether the difference is statistically rare in the population. Some local
agencies determine their own formulas or use their own point difference (e.g., 20 points). In any of these methods, if a severe discrepancy is identified, a SLD can be diagnosed.

**Response to intervention model.** The response to intervention (RTI) model is reflective of a medical prevention model. It was created as a means to decrease the number of students that were being referred for special education services as SLD; however, its utility has greatly magnified since inception. The process is divided into three tiers. Tier one is designed to provide evidence based instruction to the whole student body. Thus, every student receives tier one intervention in his/her classrooms. Tier one should be effective for 80-85% of students. Tier two, designed to be effective for 10-15% of students, targets weak areas of understanding in small group instruction. Composition of these groups should be determined based on similar need of the students in the group. Finally, tier three instruction targets 5-10% of students. In tier three, students often received one on one intensive instruction.

Movement between the tiers is determined by frequent progress monitoring. For instance, if a student is not making progress in tier one, he or she should be moved up to tier two. Conversely, if a student is making strong progress in tier three, he or she should be moved down to tier two. Movement between the three tiers should be fluid, moving both up and down the tiers. If a student is in tier three and is still not making adequate progress, he or she goes through the referral process and can be identified as having a SLD.

**Psychological process – Patterns of strengths and weaknesses.** The third way to identify a SLD is through the use of the psychological process model. This model implores the use of the Cattell-Horn-Carroll (CHC) theory of intelligence. In the CHC model, there are seven broad abilities. These seven broad abilities are fluid reasoning, crystalized intelligence, auditory
processing, processing speed, short-term working memory, long-term memory, and visual processing. Subsumed under the seven broad abilities are a number of narrow abilities.

In order to identify a SLD, there must be a processing disorder present. A concurrent identification of a weakness in an area of achievement must also be noted. The processing disorder explains why the student is underachieving in that area. For the present study, this model was the only one of the three not used in the identification process for specific learning disabilities.

**Sources of Bias in Decision Making**

Once bias toward racial minorities in special education became a legal matter, psychologists and educators have been invested in attempting to eliminate these and any other biases (Huebner, 1983). Huebner (1991) defines bias as it relates to school psychology as “divergent decisions based upon comparable information (e.g., assessment data) as a function of group membership (e.g., racial, gender, or cultural differences)” (p. 50). Research on many types of bias has since been done to discover the sources and manifestations of bias in psychoeducational testing and decision making. There are two major types of research that attempt to do such: fictitious case studies and naturalistic studies.

Fictitious case studies or analogue studies have been the most frequently used method of assessing bias in decision making (Huebner, 1983; Huebner, 1991). They have also been thoroughly criticized for being ecologically invalid (Huebner, 1983). Through fictitious case studies, researchers have not been able to find support that there is bias in interpreting assessment results, diagnostic classification of students, or in recommendations for placement (Huebner, 1983). Therefore, it was concluded that the race of a student does not unduly influence decision making (Huebner, 1983).
Fictitious case studies have yielded other noteworthy results. When presented with falsified background information on patients, participants were less likely to label the patients as healthy. Rather, the majority of patients, regardless of mental health background were labeled as unwell (Rosenhan, 1973). Furthermore, school psychologists were more likely to qualify students as emotionally disturbed if they had a psychiatric diagnosis. This was the case whether the student met definitional criteria or not. In sum, school psychologists relied heavily on background information instead of actual eligibility when making special education recommendations (Toffalo & Pedersen, 2005).

Naturalistic studies utilize real educational records to examine racial, sex based, and socioeconomic bias (Huebner, 1983). Both males and minority students have been found to be more frequently referred for evaluations and more often found eligible for special education services (Huebner, 1983). Unfortunately, many of the early naturalistic studies were statistically limited to frequency counts. Statements linking race as a cause of more frequent placement in special education services could not be made. More advanced analysis of bias in educational placement is needed. In sum, both fictitious case studies and naturalistic studies have failed to demonstrate a link between bias and decision making in educational settings.

**Key Studies**

In a study done by Ysseldyke and Algozzine (1981), students’ sex, socioeconomic status, attractiveness, and referral information (academic or behavioral) were manipulated via computer simulation to determine which factors went into the decision making of the professional involved (special education teachers, regular education teachers, administrators, and school psychologists). All test data presented were in the average range of functioning. It was found that referral information was the only statistically significant factor in deciding whether to place
students in special education programs. Specifically, if the referral was behavioral in nature, the student was more likely to be labeled as emotionally disturbed. Fifty one percent of the participants in the study relied solely on the referral data and disregarded the test results (Ysseldyke & Algozzine, 1981).

Huebner and Cummings (1985) used Ysseldyke and Algozzine’s (1981) study as a base. The researchers maintained the two types of referral information (academic and behavioral). They then added different test data and geographic location as independent variables. The test results modeled a student with a learning disability or a student of average achievement. The options for geographic location were rural or suburban. Geographic location and referral were not used by the school psychologists to make decisions regarding student’s special education qualification. Referral questions were used in the Ysseldyke and Algozzine (1981) study, but not here. The school psychologists in this study used testing data as the impetus of their decision making. As expected, the school psychologists were less likely to label a student as learning disabled and place him/her in special education if he/she received test scores that indicated average achievement (Huebner & Cummings, 1985).

Cummings, Huebner, and McLeskey (1986) used referral and test data as their independent variables. The type of referral, academic or behavioral, had no influence over diagnostic or placement decisions. The practicing school psychologists in the study were less likely to diagnose a student with a learning disability regardless of the test results, meaning they were conservative in their diagnostic practices. Results from Cummings and colleagues are supportive of Huebner and Cummings’ (1985) findings. Recommendations taken from this study indicated that an initial hypothesis should be formed based around referral information, and then data should confirm or disconfirm that hypothesis (Cummings et al., 1986).
Huebner (1991) outlined several further studies done that illustrate the historical deficits in special education decision making. First, Neer and colleagues (1973) found that professionals are more likely to diagnose mild mental retardation in children who come from low socioeconomic status (Neer, Foster, Jones, & Reynolds, 1973). In large part, these low SES students are of minority decent (McBee et al., 2012). Frame and colleagues (1982) found racial differences for learning disability eligibility decisions. These authors found that low SES black children were less likely to be found eligible for special education services (Frame, Clarizio, Porter, & Vinsonhaler, 1982). Finally, Javel and Greenspan (1983) found that negatively oriented referral data led to school psychologists recommending more restrictive special education placements. Less restrictive placements were recommended for students with the exact same data, but a better-rounded referral.

McDermott, Watkins, and Rhoad (2014) examined the educational records of students who had been given the WISC-IV. The students were kindergarteners through twelfth graders that were tested with the purpose of determining special education eligibility. Age, sex, ethnicity, and language status were controlled. Nearly 58% were found to have a specific learning disability, nearly 12% emotionally disabled, 8% had no diagnosis, and the remaining percentages were either, ADHD, mental retardation, or speech impairments. The authors found that the FSIQ differences on the WISC-IV were due to examiner differences rather than differences in the students. This was illustrated with a 12% variance in FSIQ scores based on examiner. The authors purport that this variance negates the importance of high stakes testing when making eligibility decisions (McDermott et al., 2014).
Past research has demonstrated what occurs when school psychologists are presented with fabricated and naturalistic situations regarding educational testing, classification, and placement. Findings, however, are inconsistent. For instance, in Ysseldyke and Algozzine’s (1981) study, referral was the only information used to arrive at a decision; however, Huebner and Cummings (1985) and Cummings et al. (1986) found that referral was not used at all. This difference could have potentially been influenced by the use of different team members in the first study and only school psychologists in the second. Confirmation bias, or confirming the referral question, is not a universal phenomenon, as indicated by Huebner and Cummings (1985). The training of school psychologists could potentially play a great role in whether they approach cases in a confirmatory manner or in a manner where they consider multiple variables, not only the referral. Referral and test data are important in determining special education decisions (Huebner, 1991).

The current study aims to assess the current state of naturalistic diagnosis in a school psychology training clinic with use of the following variables: sex, age group, race, medical conditions, developmental delays, referral, evaluation type, overall IQ, math achievement, writing achievement, and reading achievement. In addition to the above mentioned variables, rating scale data from emotional and behavioral, adaptive, and disorder specific scales would have been included if possible. They would have been included due to the fact that it is suspected that they were used in the decision making process for determining disability categories. However, they were not able to be included in this data analysis given the array of different rating scales that were administered. There was not enough of any one rating scale to meet stringent data analysis requirements.
Research Questions and Hypotheses

The purpose of this study is to examine the decision-making trends for determining the educational disability classification of referred students in a school psychology training clinic. The major research question for this study is: Are there significant differences between variables used in determining eligibility decisions by students in training in a school psychology program? If there are significant differences, then further analyses can be conducted. The statistical analysis will indicate what determinant variables for diagnostic classifications are used in a school psychology training clinic and how those variables are used. It is expected that variables such as race, sex, age, and type of decision made are not the primary determinant factors for any diagnostic classification. It is also expected that IQ, achievement scores, reason for referral, developmental milestones, and medical conditions will be determinant variables for some diagnostic classifications.
Chapter III: METHOD

Participants

Data for this study were collected between the 2005 and 2013 in a school psychology training clinic at a large university in the North Eastern U.S. Data used are drawn from preexisting evaluation reports or psychoeducational reports. Clients ($N = 254$; 160 male, 94 female) range in age from four to 18 years. For the purposes of data analysis, clients were split according to age into four categories: primary (4-7; $n = 68$), intermediate (8-10; $n = 104$), middle (11-13; $n = 49$), and high (14-18; $n = 33$). The majority of the clients are Caucasian but participants from minority backgrounds are represented. For data analysis, clients were grouped into either Caucasian ($n = 197$), minority background ($n = 23$), or unknown race ($n = 34$). Developmental delays or the presence of a late developmental milestone and medical conditions were split either into present or not present. Referral question was divided into four categories: academic, behavioral, academic and behavioral, or speech and language. Finally, type of evaluation was broken down into either a clinic or field evaluation.

Procedure

Participants were referred by their parents for a comprehensive evaluation in an American Psychological Association (APA) approved training clinic or seen by said clinic through contract by local school districts. Throughout assessment, doctoral level school psychology students administered a variety of psychoeducational tests to participants, parents, and teachers under the supervision of a school psychology faculty member or a community school psychologist. These tests were scored and then double scored by the supervisors. Finally, results were written up as either an evaluation report or a psychoeducational report and shared,
when appropriate, with the clients, client’s parents, or the client's school. The reports are then filed by the school psychology program. Those filed reports are what were used for this research.

**Data Analysis**

Discriminant analysis was conducted using SPSS version 20 in order to: (a) determine if there is any significant differences between the classified groups and (b) if there are significant differences found, differentiate what variables are important in determining how group membership into diagnostic classifications was decided and how those variables were used.

Discriminant analysis begins by conducting a Multivariate Analysis of Variance to determine if any of the dependent groupings (classifications) are significantly different using any of the independent variables. If there are significant differences between the groups, discriminate functions can be formed. These discriminant functions utilize the independent variables to maximize the differences between the groups. Each independent variable has a weight or loading (a measure of its contribution) calculated for each of the formed discriminant functions. The discriminant loadings for each of the independent variables indicate which independent variables were used to separate the groups on that particular discriminant function. By looking at the absolute value and the sign (positive or negative) of that loading it is possible to make statements about how the clinician is using that variable in the classification process.

Independent variables in this study include race/ethnicity (Caucasian, minority, unknown), age group (primary, intermediate, middle, high), sex, type of evaluation (clinic, field), referral question (behavioral, academic, speech-language, behavioral and academic), developmental milestones (met or not met within normal limits), medical conditions (present, absent), writing achievement, math achievement, reading achievement, and overall IQ. Included with these variables, data from rating scales would have been important. However, rating scale
data could not be included because not all individuals were administered rating scales. Even of those who were administered rating scales, the variation in the type of rating scale was so large that there was not enough of any single scale to include in the present study. Because the rating scales given were so diverse, there was no way one could even derive overall scores like withdrawal or conduct scores.

It is expected that race, age, and sex by themselves will not contribute to any of the diagnoses and will not be the exclusive variable to load on any discriminant function. However, achievement (math, reading, and writing), IQ, referral question, developmental milestones, medical conditions, and type of evaluation are expected to be used in forming many of the classifications and should load on one or more significant functions. Loadings of .4 or larger will be used as a cutoff to determining if a variable will contribute to a discriminant function and is therefore used to make classification decisions. The dependent variable in this research is the diagnostic classification rendered by the school psychologist or team in training.¹ Categories of dependent variables are attention deficit/hyperactivity disorder (ADHD), specific learning disability (SLD), emotional disturbance (ED), intellectual disability (ID; previously mental retardation), gifted, speech and language disability, other medical conditions, comorbid ADHD and SLD, and no diagnosis. With the exception of comorbid ADHD and SLD being used, only the participant’s primary diagnosis was used. Participants who were identified as gifted (n = 9) were maintained in the sample despite the fact that there were not enough cases to satisfy the discriminant analysis assumption. However, discriminant analysis is robust to the violation of this assumptions and when the gifted cases were included a clearer output was yielded.

¹ A pilot study was conducted to explore discriminant analysis and to assure there was adequate data to complete the study.
Chapter IV: RESULTS

Discriminant analysis was conducted to determine which variables (age group, sex, referral question, type of evaluation, medical conditions, developmental milestones, IQ score, reading achievement, math achievement, and writing achievement) are used to arrive at particular psychoeducational diagnoses. Descriptive information (means, standard deviations, frequencies) can be found in Table 1. Although it would have been preferable to include variables such adaptive skill ratings, behavior ratings, and other rating scale scores in the results, it was not statistically possible. There was such a variety of instruments given, that there were not enough of any single one to be included in the study.

Out of eight possible discriminant functions, three were statistically significant. A summary of the discriminant analysis is presented in Table 2. The first function’s significance is indicated by an eigenvalue of 0.887. Wilks’ Lambda of 0.277 which, when converted to a chi-square is 311.70 with 88 degrees of freedom. While the statistical output indicates that the probability of these statistics equals .000. Zero probabilities do not exist, so it may be better understood to have a probability less than 0.0005. This discriminant function is the most important of the three and accounts for 55.6% of the variance between the groups. Table 3 shows the variables loading on this discriminant function. The .4 thumb rule for determining which variables should be considered in naming the discriminant function was used. IQ loads highest (.842) followed closely by math achievement (.823), reading achievement (.799), and writing achievement (.797). All of the other variables have small negative loadings leading to a clear understanding that this discriminant function is measuring academics, as measured by intelligence and achievement.
The second function’s significance is indicated by an eigenvalue of 0.274. Wilks’ Lambda of is 0.523 which, when converted to a chi-square is 157.37 with 70 degrees of freedom. Similarly to the first function, the second function has a probability less than 0.0005. This discriminant function accounts for an additional 17.2% of the variance. Using the .4 thumb rule as with function one, type of referral was the only variable to load on the function (.836). The remaining variables were mixed with small negative and positive loadings. So after academics (IQ and measured achievement) the reason the student was referred to the clinic or school contributed most to his or her primary classification.

The third function’s significance is indicated by an eigenvalue of 0.207. A Wilks’ Lambda of 0.667 can be converted to a chi-square of 98.50 with 54 degrees of freedom. Identical to functions one and two, function three has a probability less than 0.0005. This discriminant function accounts for 13% of the variance. For this, the third discriminant function, overall IQ was the only variable to load (-.448). There were several small and midsized negative and positive variable loadings. For the third function, type of referral approached the .4 cutoff (.316).

Each of the three significant functions is presented in Table 3. In sum, the variables of overall IQ and math, writing, and reading achievement load on function 1. Referral question loads on function 2 and overall IQ loads on function 3. The three significant discriminant functions accounted for a combined 85.8% of the variance. Therefore, overall IQ, referral question, reading achievement, writing achievement, and math achievement are the variables that contribute significantly to decision making for diagnoses and account for the majority of the variance in decision making. Overall IQ is the most important variable used by school
psychologists in making decisions about diagnostic classifications. This is evidenced by its loading on two discriminant functions, one of which where it was the only variable that loaded.

This indicates that age, sex, race, type of evaluation, medical conditions, and developmental milestones were not used in the decision making process. It is a positive that sex, race, age, and type of evaluation were not used in to make decisions. If they were, there would have been evidence bias in the decision making process. Type of evaluation not being included indicates that whether a decision is made independently or part of a multidisciplinary team, a student’s ultimate classification category was the same. Developmental milestones and medical conditions did not play a role in the decision making process. This finding was surprising in that failure to meet certain developmental milestones and presence of certain medical conditions are expected to play a role in a student’s educational condition. However, that was not shown through this study.

The two diagnoses that were best differentiated between using the significant discriminant functions were gifted and intellectual disability. Figure 1 illustrates the clear distinction between students who are considered gifted and those who are considered intellectually disabled. This distinction can be seen by the graphical representation of the centroids of each diagnosis. Centroids 5 and 6 are representative of participants with intellectual disabilities and those who are gifted. Given that IQ was the most important variable to load on the significant discriminant functions, this finding is not surprising. IQ is the main variable that differentiates these two categories in that there are four standard deviations between the rough cutoffs for gifted and an intellectual disability.

The classification cross-tabulation results are presented in Table 4. Table 4 is created by using the original classifications made by the clinician and then reclassifying the student by
putting them in the group with the highest probability as determined by the discriminant analysis. One can then create a cross-tabulation table as shown in Table 4. Table 5 is a simplified version of Table 4 and shows the frequency and percent correct for each classification category. The diagnosis groups intellectual disabilities, gifted, and no diagnosis had the best hit rates with percentages all being above 75%. Medical conditions, emotional disturbance, and ADHD had the worst hit rates with percentages all being below 15%. The low percentages for emotional disturbance and ADHD are not surprising in that rating scale information could not be included in the data analysis. Much of the information gathered to make these diagnostic decisions would be provided by rating scale information. In all, only 51.2% of cases were reclassified correctly. The two groups that were the most differentiated were intellectual disability and gifted. This finding is logical given IQ was the most important variable used by school psychologists in making diagnostic decisions and IQ is inherent in the definitions of these educational disabilities.
Chapter V: DISCUSSION

The purpose of the present study was to examine the decision-making trends for determining diagnoses of referred students in a school psychology training clinic. It was hypothesized that variables such as race, sex, age, and type of decision made are not the primary determinate factors for any diagnostic classification. It was also hypothesized that IQ and achievement scores will be determinate variables for some diagnostic classifications.

Results indicated that a combination of overall IQ, referral question, reading achievement, math achievement, and writing achievement were variables that contributed significantly to decision making for diagnoses. Those five variables accounted for over 85% of the variance in decision making. Both hypotheses were supported by the data. Neither sex, race, age, nor type of decision made contributed to the decision making. The presence of medical conditions or developmental delays also did not play a role. Overall IQ and the three areas of achievement were significant determinants in decision making. It should be noted that although the editions of intelligence and achievement test updated over the dates of the collected data, results should not have been impacted by the edition changes. The scores used (e.g., overall IQ, composite achievement scores) are comparable between editions, in that they measure the same constructs. In kind, the Diagnostic and Statistical Manual – Fifth Edition (DSM-5, 2013) was updated from the Fourth Edition during the data collection period. However, the update to the Fifth Edition does not nullify the diagnoses from the Fourth Edition.

Although referral question was not anticipated as a major contributor to decision making, perhaps it should have been based on previous research (e.g., Huebner & Cummings, 1985). When minimal evidence is known about a student, only poor hypotheses can be made about the origins of that student’s problem. Bayes Theorem states that the belief about something should
rationally change when new evidence is taken into account (Bayes & Price, 1763). Therefore, an initial hypothesis about a student should rationally change when more information, a referral question, is brought to light. Bayes Theorem also explains why overall IQ and the three achievement areas contributed to decision making. This is so because the evidence or testing results should have colored the hypotheses and therefore the decision of the clinicians.

Limitations and Future Directions

Given the limited number of data, concessions had to be made for the purposes of data analysis. First and foremost, it would have been preferable to include rating scales. However, given the constrictions of discriminant analysis, these rating scales could not be used because of the many different scales that were used. Additionally, there was a different number of scales used per subject, and no way to combine the scores. Ratings of adaptive skills, behavior (internalizing and externalizing), ADHD, and autism are all used in practice to make decisions. In fact, the absence of these scales would be nearly unheard of for an educational disability. Not being able to include rating scales in data analysis may likely have influenced the results in that it is know that the rating scales should have gone into the decision making process. Unfortunately, there was no way to investigate this in the present study. Future research on this topic should aim to obtain a larger sample or a more consistent sample, with regard to measurement, in order to include rating scales as a part of analysis.

Only five of the thirteen disability categories were able to be included in the analysis. These were speech and language impairment, emotional disability, intellectual impairment, specific learning disability, and other health impairment (which was split into ADHD and medical conditions). Although it was positive to have the other categories of gifted, no disabilities, and concurrent ADHD and SLD, it would have been preferable to include many
more of the thirteen disability categories. Paramount among these would be autism. There were not enough cases of students with autism or the other disability categories to include them in data analysis.

More of the present variables may have played a role had all areas of classification would have been available for analysis. It may be assumptive or presumptive, but for the more medically based educational disabilities (e.g., deaf-blindness, orthopedic impairment, traumatic brain injury), the presence of medical conditions or presence of developmental delay would have certainly played a determinant role in decision making. Given data restrictions, medical conditions and developmental milestones had to be broken down into either present or not present. This condensation of information for coding purposes, could have impacted the results. If they were able to be separated in to more components, the results could have revealed that they were used in decision making.

Additionally, the coding of race may have played a role in the outcome and limits the supposition that race was not included in decision making in the current study. At most, it can be said that decisions made in the school psychology training clinic were not differentiated using those of Caucasian decent and those of minority decent. Conclusions based off of further racial breakdown cannot be made at this time.

It is possible to extend the validity of this study by knowing the base rates for each of the diagnostic categories. It is known that the base rates are small for nearly all of the categories, with the exception of specific learning disabilities and the no diagnosis category. Even with a psychometrically sound, non-bias test, the prediction or hit rate will not be as high as desired due to the small base rates. In the case of small base rates, both sensitivity (ability of a test to correctly identify a disorder) and specificity (ability of a test to not identify a disorder when one
is not present) must both be high. Therefore, if the hit rates are high and there are larger base rates, the scores and decision made will be more valid or if there are high hit rates and small base rates with high specificity and sensitivity, the decisions made will be more valid.
Table 1

Descriptive Statistics of Discriminant Analysis Independent Variables and Diagnosis (N = 254)

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Means</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>94</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Male</td>
<td>160</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><strong>Age Group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>68</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Intermediate</td>
<td>104</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Middle</td>
<td>49</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>High</td>
<td>33</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>197</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Minority</td>
<td>23</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Unknown</td>
<td>34</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><strong>Evaluation Type</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinic</td>
<td>112</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Field</td>
<td>142</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><strong>Referral Question</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic</td>
<td>110</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Behavioral</td>
<td>20</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Speech/Language</td>
<td>19</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>A and B</td>
<td>105</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><strong>Developmental Delays</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>70</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Not Present</td>
<td>184</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><strong>Medical Conditions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>140</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Not Present</td>
<td>114</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><strong>Overall IQ</strong></td>
<td>--</td>
<td>93.04</td>
<td>16.73</td>
</tr>
</tbody>
</table>

**Note.**  A = Academic; B = Behavioral; SL = Speech/Language; ED = Emotional Disturbance; ID = Intellectual Disability; SLD = Specific Learning Disability; ADHD = Attention Deficit Hyperactivity Disorder.
Table 2

*Discriminant Analysis Summary Table*

<table>
<thead>
<tr>
<th>Discriminant Function</th>
<th>Eigenvalue</th>
<th>Percentage of Variance</th>
<th>Canonical Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.887</td>
<td>55.6</td>
<td>.686</td>
</tr>
<tr>
<td>2</td>
<td>0.274</td>
<td>17.2</td>
<td>.464</td>
</tr>
<tr>
<td>3</td>
<td>0.207</td>
<td>13.0</td>
<td>.414</td>
</tr>
<tr>
<td>4</td>
<td>0.119</td>
<td>7.4</td>
<td>.326</td>
</tr>
<tr>
<td>5</td>
<td>0.066</td>
<td>4.1</td>
<td>.248</td>
</tr>
<tr>
<td>6</td>
<td>0.021</td>
<td>1.3</td>
<td>.142</td>
</tr>
<tr>
<td>7</td>
<td>0.013</td>
<td>0.8</td>
<td>.115</td>
</tr>
<tr>
<td>8</td>
<td>0.008</td>
<td>0.5</td>
<td>.087</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Derived Function</th>
<th>Wilks’ Lambda</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.277</td>
<td>311.70</td>
<td>88</td>
<td>.000</td>
</tr>
<tr>
<td>2</td>
<td>.523</td>
<td>157.37</td>
<td>70</td>
<td>.000</td>
</tr>
<tr>
<td>3</td>
<td>.667</td>
<td>98.50</td>
<td>54</td>
<td>.000</td>
</tr>
<tr>
<td>4</td>
<td>.805</td>
<td>52.75</td>
<td>40</td>
<td>.085</td>
</tr>
<tr>
<td>5</td>
<td>.900</td>
<td>25.51</td>
<td>28</td>
<td>.600</td>
</tr>
<tr>
<td>6</td>
<td>.960</td>
<td>10.04</td>
<td>18</td>
<td>.931</td>
</tr>
<tr>
<td>7</td>
<td>.979</td>
<td>5.10</td>
<td>10</td>
<td>.885</td>
</tr>
<tr>
<td>8</td>
<td>.992</td>
<td>1.85</td>
<td>4</td>
<td>.764</td>
</tr>
</tbody>
</table>
Table 3

*Significant Function Structure Matrix*

<table>
<thead>
<tr>
<th></th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Sex</td>
<td>-.021</td>
</tr>
<tr>
<td>Age</td>
<td>-.151</td>
</tr>
<tr>
<td>Race</td>
<td>-.096</td>
</tr>
<tr>
<td>Evaluation Type</td>
<td>-.116</td>
</tr>
<tr>
<td>Referral</td>
<td>-.120</td>
</tr>
<tr>
<td>Developmental Delays</td>
<td>-.081</td>
</tr>
<tr>
<td>Medical Conditions</td>
<td>-.103</td>
</tr>
<tr>
<td>Overall IQ</td>
<td>.842</td>
</tr>
<tr>
<td>Reading Achievement</td>
<td>.799</td>
</tr>
<tr>
<td>Writing Achievement</td>
<td>.797</td>
</tr>
<tr>
<td>Math Achievement</td>
<td>.823</td>
</tr>
</tbody>
</table>
Table 4  

Cross-Tabulation of Classification Results

<table>
<thead>
<tr>
<th>Count</th>
<th>Predicted Group Membership</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech/Language</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Medical Conditions</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ED</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>17</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>No Diagnosis</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>79</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Gifted</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>SLD</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>13</td>
<td>1</td>
<td>0</td>
<td>18</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ADHD</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>SLD and ADHD</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Percent

<table>
<thead>
<tr>
<th>Count</th>
<th>Predicted Group Membership</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech/Language</td>
<td>28.6</td>
<td>0.0</td>
<td>0.0</td>
<td>52.4</td>
<td>0.0</td>
<td>0.0</td>
<td>14.3</td>
<td>4.8</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>Medical Conditions</td>
<td>7.7</td>
<td>7.7</td>
<td>15.4</td>
<td>23.1</td>
<td>7.7</td>
<td>0.0</td>
<td>23.1</td>
<td>7.7</td>
<td>7.7</td>
<td></td>
</tr>
<tr>
<td>ED</td>
<td>0.0</td>
<td>0.0</td>
<td>14.8</td>
<td>63.0</td>
<td>0.0</td>
<td>3.7</td>
<td>11.1</td>
<td>3.7</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td>No Diagnosis</td>
<td>3.9</td>
<td>1.0</td>
<td>5.9</td>
<td>77.5</td>
<td>1.0</td>
<td>1.0</td>
<td>5.9</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>8.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>83.3</td>
<td>0.0</td>
<td>0.0</td>
<td>8.3</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>Gifted</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>22.2</td>
<td>0.0</td>
<td>77.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>SLD</td>
<td>7.7</td>
<td>2.6</td>
<td>0.0</td>
<td>33.3</td>
<td>2.6</td>
<td>0.0</td>
<td>46.2</td>
<td>5.1</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>ADHD</td>
<td>10.0</td>
<td>0.0</td>
<td>10.0</td>
<td>55.0</td>
<td>0.0</td>
<td>0.0</td>
<td>10.0</td>
<td>15.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>SLD and ADHD</td>
<td>0.0</td>
<td>0.0</td>
<td>18.2</td>
<td>36.4</td>
<td>0.0</td>
<td>0.0</td>
<td>27.3</td>
<td>0.0</td>
<td>18.2</td>
<td></td>
</tr>
</tbody>
</table>

Note. ED = Emotional Disturbance; ID = Intellectual Disability; SLD = Specific Learning Disability; ADHD = Attention Deficit Hyperactivity Disorder.
Table 5

Discriminant Analysis Classification Results

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Total Possible Hits or Misses</th>
<th>Frequency Correct</th>
<th>Percent Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>12</td>
<td>10</td>
<td>83.3</td>
</tr>
<tr>
<td>Gifted</td>
<td>9</td>
<td>7</td>
<td>77.8</td>
</tr>
<tr>
<td>No Diagnosis</td>
<td>102</td>
<td>79</td>
<td>77.5</td>
</tr>
<tr>
<td>SLD</td>
<td>39</td>
<td>18</td>
<td>46.2</td>
</tr>
<tr>
<td>Speech/Language</td>
<td>21</td>
<td>6</td>
<td>28.6</td>
</tr>
<tr>
<td>SLD and ADHD</td>
<td>11</td>
<td>2</td>
<td>18.2</td>
</tr>
<tr>
<td>ADHD</td>
<td>20</td>
<td>3</td>
<td>15.0</td>
</tr>
<tr>
<td>ED</td>
<td>27</td>
<td>4</td>
<td>14.8</td>
</tr>
<tr>
<td>Medical Conditions</td>
<td>13</td>
<td>1</td>
<td>7.7</td>
</tr>
</tbody>
</table>

*Note. ED = Emotional Disturbance; ID = Intellectual Disability; SLD = Specific Learning Disability; ADHD = Attention Deficit Hyperactivity Disorder.*
Figure 1. Canonical Discriminant Functions of Diagnostic Categories

Figure 1. Canonical discriminant functions depicted by case. Number five, yellow in color, signifies intellectual disability (ID) and number six, red in color, signifies gifted.
References


Megan Marie Gammelmo
42515 Hollyhock Terrace Ashburn, Virginia 20148
703-597-8592 mmg5255@psu.edu

Curriculum Vitæ

EDUCATION
Ph.D., School Psychology (August 2015)
The Pennsylvania State University, University Park, Pennsylvania
Advisor: Robert Hale, Ph.D.

M.Ed., School Psychology May 2012
The Pennsylvania State University, University Park, Pennsylvania
Advisor: Beverly Vandiver Ph.D.

B.S., Psychology, magna cum laude May 2010
George Mason University, Fairfax, Virginia

Advanced High School Diploma June 2006
Hayfield Secondary School, Alexandria, Virginia

EXPERIENCE
Loudoun County Public Schools
Supervisor: Heather Applegate, Ph.D.

Practicum in School Psychology 2013-2014
State College Area School District
Supervisor: Shirley Woika, Ph.D.

Graduate Assistant Counselor, Multicultural Resource Center 2011-2014
Supervisors: Tara Scales-Williams, Ph.D., Moses Davis, Ph.D.

CERTIFICATES
Pennsylvania Educational Specialist I - School Psychology PK-12 2014-present
Virginia Pupil Personnel Service License – School Psychology 2014-present

PRESENTATIONS