

School Psychologists' Decision Making in Evaluations for Emotional Disturbance

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY
Shanna S. Sadeh

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE DOCTOR OF PHILOSOPHY

Dr. Amanda Sullivan

July, 2016

© Shanna Sadeh, 2016

Abstract

For decades, there has been a persistent national trend of public schools disproportionately qualifying more Black students relative to White students for special education under the category of serious emotional disturbance (ED). Such disproportionality suggests but does not prove racial bias in ED evaluations. I experimentally tested how much, if at all, school psychologists' racial bias impacted eligibility determinations using a vignette methodology and between-group design with three conditions that varied by level of data ambiguity: (a) low-ambiguity data that do not meet ED criteria; (b) low-ambiguity data that meet ED criteria; and (c) highly ambiguous data. The hypothetical student in each vignette was a fifth grade male who had primarily externalizing problems. Participants completed one vignette in each ambiguity condition; student race (Black versus White) was experimentally manipulated. Participants were 60 practicing school psychologists in a northeastern state that adopted the federal regulations for ED eligibility. For each vignette, participants decided whether the student qualified as ED, rated their confidence in their decision and the diagnosticity of data included in the evaluation, and had the option to describe additional data they wish had been included in the results.

Chi-square analyses indicated there were no statistically significant differences based on race between students qualified and disqualified as ED across ambiguity conditions, providing some evidence against the racial bias theory of disproportionality. Under the highly ambiguous data condition, there was no statistically significant difference between students qualified as ED and those not qualified – i.e., regardless of race, all students had a coin-toss chance of qualifying as ED. This finding makes sense in light of the numerous ambiguous key terms in the ED criteria, which allow for more than one reasonable interpretation.

Results also showed that most school psychologists were at least moderately confident in their determinations across ambiguity conditions. Their confidence in the low-ambiguity conditions makes sense because those vignettes were designed to be relatively easy. Their

confidence in the highly ambiguous data condition may illustrate the potency and frequency of confirmation bias in decision making under conditions of high uncertainty.

Across ambiguity conditions, participants frequently identified behavior rating scales and infrequently identified achievement and intelligence scores as highly diagnostic. They identified interviews, family information, and observations with varying frequency across conditions, demonstrating that the diagnosticity of data can fluctuate depending on the presenting problems and evaluation results.

Finally, school psychologists who opted to describe additional data they wish had been included in the evaluation results primarily requested more information about interventions that had been attempted and consultation with outside mental health providers. Implications for practice and further research opportunities are discussed.

Table of Contents

List of Tables	1
Chapter 1 Introduction	2
Statement of the Problem.....	2
Purpose of the Study	6
Race by Ambiguity	7
Confidence by Ambiguity.....	7
Data Diagnosticity by Ambiguity	8
Additional Data by Ambiguity.....	8
Organization of Dissertation.....	8
Chapter 2 Literature Review.....	10
Racial Disproportionality among Students with ED.....	10
Evaluation Practices.....	13
Cognitive Biases May Contribute to Disproportionality	15
ED Evaluations May Be Vulnerable to Cognitive Biases	22
Data Diagnosticity	27
Extant Research on Bias in ED Evaluations.....	29
Chapter 3 Methods.....	34
Experimental Design.....	34
Sampling Design.....	34
Participants.....	35
Instrumentation	36
Procedures.....	39
Analyses.....	39
Chapter 4 Results	41
Sample Characteristics.....	41
Race by Ambiguity	43
Confidence by Ambiguity.....	44
Data Diagnosticity by Ambiguity	45
Additional Data by Ambiguity.....	46

	iv
Chapter 5 Discussion	49
Racial Bias	49
Data Ambiguity.....	52
Confidence	54
Data Diagnosticity	55
Additional Data.....	57
Merits and Limitations.....	57
Implications for Practice.....	61
References.....	64
Appendix A – Email invitation.....	77
Appendix B – Instrumentation.....	79

List of Tables

Table	Page
1. Participant and District Characteristics	42
2. Eligibility Determinations by Ambiguity Condition and Student Race	43
3. Most Diagnostic Data by Level of Ambiguity	45
4. Most Frequently Reported Data Participants Wished Had Been Included by Level of Ambiguity	47

Chapter 1 Introduction

Statement of the Problem

Unlawful bias in special education evaluations. The education of all children is so important in a democratic society that public education is a property right protected by the 14th Amendment of the U.S. Constitution (e.g., Jacob, Decker, & Hartshorne, 2011). Notwithstanding this constitutional safeguard, in the 1970's Congress determined that millions of students with disabilities were not receiving an appropriate public education. In response, Congress enacted the Individuals with Disabilities Education Act (IDEA) to ensure "equality of opportunity, full participation, independent living, and economic self-sufficiency for individuals with disabilities" (20 U.S.C. § 1400(c)(2), 2015). IDEA is consistent with federal antidiscrimination legislation requiring government actors to treat individuals equally without discriminating on the basis of race, color, national origin, religion, or sex (Civil Rights Act of 1964). Altogether, these federal laws impose responsibilities on schools to ensure students with disabilities receive a free and appropriate public education and that schools do so without discriminating on the basis of students' race or other protected statuses.

Although one of the purposes of IDEA is to stop schools from unlawfully discriminating against students with disabilities, evidence suggests that it may perpetuate discrimination against a different protected class – i.e., ethnic/racial minorities; (hereinafter "racial"). Some argue that IDEA's special education services and ancillary supports have devolved into government-facilitated segregation based on race (Redfield & Kraft, 2012; Reid & Knight, 2006). Research shows that racial minority students are disproportionately represented in special education (e.g., Donovan & Cross, 2002). Disproportionality is problematic because it "suggests assessment procedures are not applied equally to all racial groups, which may lead to some groups being over- or under-identified for special education" (Bollmer, Bethel, Garrison-Mogren, & Brauen, 2007, p. 186). Both kinds of disproportionality suggest disparate impact such that the educational needs of some children from minority racial and ethnic backgrounds are not being met

(Bollmer et al., 2007, p. 186). Disproportionality suggests, but does not prove, that there may be bias in evaluations for special education. In discrimination literature, bias “usually implies a prejudgment or prejudice” that contributes to “error or inaccuracy in clinical judgment” regarding minority group members (Lopez, 1989, p. 184).

Specifically, practitioners may perceive minority group members “as more disturbed or as requiring more treatment” due to an overpathologizing bias rather than actual differences between minority and majority group members (Lopez, 1989, p. 184). An overpathologizing bias based on legally protected statuses like gender or race would violate the Equal Protection Clause of the 14th Amendment and civil rights legislation if the practitioners were government actors, like personnel working in public schools.

Ideally, schools prevent unlawful discrimination by using “an objective hypothesis testing procedure to assess the eligibility of a special education candidate” (O’Reilly, Northcraft, & Dabers, 1989, p. 127). “The implication of hypothesis testing is that the diagnostic process entails systematic testing of possibilities . . . when *a priori* none of these possibilities is favored over any other” (O’Reilly et al., 1989, p. 127). However, analogue research conducted decades ago suggests that special education evaluations are vulnerable to numerous cognitive biases (e.g., confirmation, anchoring, adjustment, availability, and representativeness biases) at different stages of the assessment process (Algozzine & Ysseldyke, 1981; Davidow & Levinson, 1993; Fagley, 1988; McDermott, 1981; O’Reilly et al., 1989; Ysseldyke & Algozzine, 1982; Ysseldyke, Algozzine, Regan, & McGue, 1981). Society has a strong interest in understanding whether bias influences special education evaluations because they may be unintended conduits for unlawful discrimination and result in misplacement of students in special education. Despite the need for such research, I could not locate any experimental studies published since the 1980’s investigating the influence of racial bias on special education evaluations.

Another factor that may contribute to racial disproportionality in special education is ambiguous assessment data. By law, evaluations should include evidence from

multiple sources using multiple methods for all areas of suspected disability (20 U.S.C. §§ 1414, 1415, 2015), which can yield contradictory data that are difficult to synthesize and reconcile. Some scholars argue that the uncertainty in and complexity of school evaluations make them unjustifiably susceptible to bias (e.g., Barnett, 1988; Davidow & Levinson, 1993). An unexplored area of research is whether the lack of standardization in interpreting and weighing data gives rise to differences in professional judgment that are inconsistent with the objectives of IDEA.

Additionally, ambiguous disability constructs and vague eligibility criteria may contribute to racial disproportionality in special education. When practitioners do not understand what constitutes a disability and how to detect that disability using valid and reliable measures, the conditions for decision making are highly uncertain. “Clinical uncertainty opens the door to the possible influence of stereotypes that clinicians link to observable patient characteristics such as race” (Neighbors, Trierweiler, Ford, & Muroff, 2003, p. 239). Hence, racial stereotypes of young, Black males in the U.S. may trigger an overpathologizing bias in school personnel participating in special education evaluations.

Evaluations for emotional disturbance. Nationally, in 2011, public schools provided about 389,000 students with severe social, emotional, or behavioral problems special education under the disability category of serious emotional disturbance (ED; U.S. Department of Education (DOE), 2013). Under federal law, ED 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 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; and (E) a tendency to develop physical symptoms or fears associated with personal or school problems. 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 this section.
(34 C.F.R. § 300.8(c)(4), 2015)

Scholars have criticized the federal definition of ED because it includes several ambiguous material terms (e.g. “inappropriate types of feelings,” “satisfactory interpersonal relationships,” “long period of time,” “adversely effects”) and the vague social maladjustment exclusion is unsupported by empirical research (e.g., Olympia et al., 2003; Allen & Hanchon, 2013).

Nationally representative studies provide insight how schools apply ambiguous eligibility criteria. Findings from the nationally representative Special Education Elementary Longitudinal Study (SEELS) and the National Longitudinal Transition Study-2 (NLTS-2), which collected data during the 2000-2001 school year, show that although students with ED are a heterogeneous group, most are male (80%), about one-third live in poverty or a single-parent household, and 46% live in a household with another person who has a disability (Wagner, Kutash, Duchnowski, Epstein, & Sumi, 2005). The average age when elementary/middle schools identify students as ED is 7.8 years, although these children begin showing behavior problems on average at age 4.6 years. Within this population, about 35 % received preschool special education and 24% received early intervention services. On average, students with ED tend to have difficulty with self-control, cooperation, and communication. They tend to have “maladaptive relationships with peers and adults due to acquisition and performance deficits that include limited prosocial interactions, tendencies to misinterpret neutral social cues as hostile, and behavior patterns that impede teacher’s abilities to conduct instruction effectively” (Lane & Carter, 2006, p. 66). Many students with ED struggle academically (Lane & Carter, 2006), although about 28% have high cognitive abilities (Wagner et al., 2005).

At the national level schools have disproportionately identified Black males with ED for the past 30 years (Donovan & Cross, 2002; DOE, 2013). In 2011-2012, Black students were 1.6 times more likely than White students to qualify as ED (DOE, 2013).

Such disproportionality is alarming because there is an “increasing trend in the use of exclusionary settings for students with” ED and students with ED “are more likely to receive a large portion of their services from paraprofessionals” rather than licensed teachers (Bradley, Doolittle, & Bartolotta, 2008, p.7). Furthermore, national studies show that for nearly 30 years about 55% of students with ED have dropped out of school (Bradley et al., 2008). Post-dropout employment and social outcomes tend to be unfavorable for individuals with ED. In light of these national trends, one may argue that it is questionable whether on average special education services for students with ED provide a free and appropriate public education. Accordingly, over-identification of Black males as ED raises serious questions about civil liberties.

Role of school psychologists. School psychologists working in public schools frequently participate in special education evaluations for ED as members of legally mandated multidisciplinary evaluation teams (METs). Typically, school psychologists have graduate training in special education law, psychological and educational assessment, and the operations of school systems, which can make them uniquely influential in the evaluation process (Becker, Paternite, & Evans, 2014; Gilliam, 1979; Knoff, 1983a; Mehan, 1991; Yoshida, Fenton, Maxwell, & Kaufman, 1978). Accordingly, whether school psychologists have racial biases that impact special education evaluations for ED is an important, but unexplored, area of research.

Purpose of the Study

The overarching purpose of this study was to examine how racial bias may influence school psychologists’ eligibility determinations when students have social, emotional, or behavioral problems in school. In light of prior research indicating over-identification of Black males as ED (e.g., Donovan & Cross, 2002), evaluation outcomes for Black and White male students were investigated using hypothetical vignette methodology.

Directional hypotheses were not posited because the extant research was insufficient to do so. The research questions were as follows.

Race by Ambiguity

1. How much does student race influence school psychologists' evaluation decisions when data concerning social, emotional, and behavioral problems are highly ambiguous?

H₀: There is no association between student race and evaluation outcome when data are highly ambiguous.

H_a: Student race and evaluation outcome are associated when data are highly ambiguous.

2. How much does student race influence school psychologists' evaluation decisions when data concerning social, emotional, and behavioral problems have low ambiguity and do not meet criteria for ED?

H₀: There is no association between student race and evaluation outcome when data have low ambiguity and do not meet criteria for ED.

H_a: Student race and evaluation outcome are associated when data have low ambiguity and do not meet criteria for ED.

3. How much does student race influence school psychologists' evaluation decisions when data concerning social, emotional, and behavioral problems have low ambiguity and meet criteria for ED?

H₀: There is no association between student race and evaluation outcome when data have low ambiguity and meet criteria for ED.

H_a: Student race and evaluation outcome are associated when data have low ambiguity and meet criteria for ED.

Confidence by Ambiguity

4. How confident are school psychologists in their evaluation decisions when data are highly ambiguous?
5. How confident are school psychologists in their evaluation decisions when data have low ambiguity and do not meet criteria for ED?

6. How confident are school psychologists in their evaluation decisions when data have low ambiguity and meet criteria for ED?

Data Diagnosticity by Ambiguity

7. What data do school psychologists identify as most diagnostic when data are highly ambiguous?
8. What data do school psychologists identify as most diagnostic when data have low ambiguity and do not meet criteria for ED?
9. What data do school psychologists identify as most diagnostic when data have low ambiguity and meet criteria for ED?

Additional Data by Ambiguity

10. What data do school psychologists wish were included when the extant data are highly ambiguous?
11. What data do school psychologists wish were included when the extant data have low ambiguity and do not meet criteria for ED?
12. What data do school psychologists wish had been included when the extant data have low ambiguity and meet criteria for ED?

Organization of Dissertation

In the following literature review, I explore how biased decision making in special education may contribute to racial disproportionality by: (a) briefly explaining the complexities of the disproportionality phenomena in terms of theory and extant research approaches; (b) reviewing research from cognitive science investigating cognitive biases; (c) explaining how the legal criteria for ED may elicit cognitive biases; (d) reviewing judicial case law interpreting those ambiguous legal terms; and (e) reviewing the extant experimental research on bias in special education evaluations. Following the literature review, I describe the methods used in this study, including the recruitment of participants, development of hypothetical vignettes, procedures, and analyses. Next, I present the results, which are organized by research question. I conclude by summarizing

and interpreting the results, and discussing limitations, implications for practice and future directions for research.

Chapter 2 Literature Review

Racial Disproportionality among Students with ED

Statistical patterns demonstrating racial disproportionality in special education have existed within the education system since 1968 (e.g., Dunn, 1968; Waitoller, Artiles, & Cheney, 2010). Disproportionality generally means that students' membership in a particular group (e.g., race, linguistic, gender) affects their likelihood of qualifying for special education services, which is reflected in the group's over- or under-representation in special education (e.g., Artiles, Kozleski, Trent, Osher, & Ortiz, 2010). In studies of disproportionality based on race, White students are commonly identified as the comparison group "because they are the dominant group in society who have not had systematic problems with access and opportunity issues" (Artiles, Rueda, Salazar, & Higareda, 2005, p. 289). Also, consistently using White students as the reference group serves a practical purpose of facilitating analyses across studies.

One troubling national pattern of racial disproportionality is the persistent over-identification of Black males as ED. Since 1968, the Office of Civil Rights (OCR) has collected data on the number of children served under IDEA (Heller, Holtzman, & Messick, 1982). From 1976 to 2012, the overall national prevalence of ED has ranged from 0.6% to 1.0%; from 2009 to 2012 it has been a steady 0.8%. There has been statistical evidence of disproportionality in ED since at least 1978, when the OCR's survey data showed Black students were 1.74 times more likely than White students to qualify as ED (Heller et al., 1982). Forty years later, in 2008, data from the Office of Special Education Programs showed that Black students were 2.38 times more likely than White students to qualify as ED (Wiley, Bringham, Kauffman, & Bogan, 2013). The most recent national data from 2011-2012 show that the ED rate was 1.3% for Black students and 0.8% for White students (U.S. Department of Education, 2013). Put differently, Black students were 1.63 times more likely than White students to be identified as ED.

This pattern is concerning because the benefits of special education may be outweighed by its iatrogenic effects depending upon the characteristics of the student, school, community, and other factors. Nationally, students who receive special education tend to perform worse academically than their peers who have not been identified as having a disability even though special education services are supposed to include intense, specialized interventions to support academic progress (e.g., Artiles et al., 2010; Morgan, Frisco, Farkas, & Hibel, 2010). Indeed, there is some evidence that preschoolers with delays and disabilities do better in kindergarten if they do not participate in preschool special education services (Sullivan & Field, 2013). There is some evidence that special education classes are taught by instructors with lower qualifications than instructors in general education (Bradley et al., 2008). The effectiveness of special education for students with ED is debatable when the dropout rate for this group has been about 55% for more than 20 years. After dropping out, 73% of students who had been identified as ED were arrested and more than 30% were unemployed within five years after leaving school (Wagner, 1995; Blackorby & Wagner, 1996). Findings such as these raise concerns that on average students with ED may not be receiving the free and appropriate public education to which they are entitled under constitutional and federal law.

Not only might special education systems fail to confer an appropriate education, these systems might actually inflict harm on students with ED by segregating them from the general education population. Some scholars theorize that segregated special education classrooms for students with ED inadvertently facilitate “deviant peer influence and the loss of opportunities for positive influence from well-adjusted peers” (Gifford-Smith, Dodge, Dishion, & McCord, 2005, p. 7). Furthermore, such iatrogenic effects may disparately impact Black students because there is some evidence that Black students receiving special education “are only half as likely to be placed in general education environment as their White peers” (e.g., Artiles et al., 2010, p. 285). In light of these

concerning national statistics and scholarship, it is important for stakeholders to know what perpetuates patterns of racial disproportionality in special education.

For almost forty years, scholars, educators, and policymakers have debated what causes racial disproportionality in special education. Some justify disproportionality on the grounds that education-related delays and deficits actually manifest more frequently in minority children because of their increased exposure to developmental risk factors like poverty, prenatal toxins, and other factors that are outside the control of school systems (e.g., Donovan & Cross, 2002). Others argue that disproportionality reflects bias in the special education identification process and implicate power and privilege within institutional structures and professional practices (e.g., Skiba, Poloni-Staudinger, Simmons, Feggins-Azziz, & Chung, 2005; Artiles et al., 2010). Trying to untangle the factors contributing to disproportionality, researchers have explored whether race is a proxy for poverty in the overrepresentation of Black students in special education. The overlap of poverty and poor school performance has been extensively documented (Rothstein, 2004), although the role of class in special education identification specifically continues to need further examination because schools do not qualify all poor students for special education (e.g., Donovan & Cross, 2002; Skiba et al., 2005). Addressing the “unproven links” between poverty and disproportionality, Skiba and colleagues (2005, p. 131) tested descriptive models of factors associated with special education. Their study of state-level ED data from Indiana showed that students in high-poverty school corporations were 2.08 times more likely than students in wealthier corporations to be identified as ED (Skiba et al., 2005). However, when they controlled for poverty, race significantly influenced the odds of special education identification such that Black students were 1.5 times as likely as White students to qualify as ED. In other words, poverty was not a significant predictor ED identification in this particular sample. They concluded that in multivariate analyses, poverty is “a weak and inconsistent predictor of disproportionality” across disability categories (Skiba et al., 2005, p. 141).

Other researchers using multilevel analyses that account for nesting variables have found that class and gender attenuate the race effect (Sullivan & Bal, 2013). Specifically, archival data from an urban, diverse, Midwestern school district showed that controlling for age, Black males were 2.99 times more likely than White males to have ED (Sullivan & Bal, 2013). The race effect was attenuated to an insignificant odds ratio (OR) of 1.05 when the analysis included more individual- and district-level variables. Numerous other student-level variables contributed significantly to patterns of over-representation of ED: age (OR=1.32), male (OR=2.20), free/reduced lunch status (OR=2.40), and suspensions (OR=1.42). At the school-level, students were more likely to be ED if their school had many students that received free/reduced school lunch. Overall, these results suggest that race may be a proxy for the influence of class and gender in differential rates of ED.

Descriptive studies like these have produced inconsistent findings about the occurrence and extent of disproportionality depending upon the level of analysis (e.g., grade, school, district, state, or national), characteristics of the school and community (e.g., demographic and economic diversity of context), and specific disability category (e.g., Waitoller et al., 2010). Rigorous experimental research is needed to measure the potential impact of bias against legally protected statuses like race and gender. Society has a strong interest in elucidating the role, if any, of unlawful bias in special education evaluations. If government actors like school personnel are racially biased in their implementation of IDEA, then they would be violating the Equal Protection Clause of the 14th Amendment to the U.S. Constitution as well as antidiscrimination legislation like the Civil Rights Act of 1964. Experimental studies of the role of bias in special education evaluations are needed to answer this empirical question. If the evidence demonstrates bias, then professional training and public policies can be developed to target the bias.

Evaluation Practices

It is critical for practitioners to conduct comprehensive ED evaluations because poor assessments may contribute to over-representation of minority students in special

education (Allen & Hanchon, 2013). Additionally, federal law requires nondiscriminatory, comprehensive evaluations that utilize multiple informants and multiple measures (20 U.S.C. §§ 1414, 1415, 2015). Specifically, the MET must “use a variety of assessment tools and strategies to gather relevant functional, developmental, and academic information . . . [and] not use any single measure or assessment as the sole criterion for determining whether a child” meets the disability criteria (20 U.S.C. § 1414(b)(2), 2015). Instruments need to be technically sound and used for purposes that are valid and reliable. Notably, the purpose of ED evaluations is not to diagnose students with specific types of emotional disorders (e.g., depression, anxiety, adjustment disorder) and no such diagnosis needs to be established for a child to qualify as ED. Likewise, mental health diagnoses do not automatically qualify students for special education because the ED eligibility criteria have educational and mental/health/behavioral components – i.e., the emotional/social/behavioral problem must adversely impact students’ educational performance. There are no guidelines for how practitioners should weigh the educational and mental components in determining eligibility (Becker et al., 2014).

Recent evidence from a survey of school psychologists suggests that many practitioners do not conduct legally defensible, comprehensive ED evaluations, which typically should include: classroom observation by the school psychologist; teacher, parent, and child interviews; and normative data derived from rating scales completed by the parent and at least one teacher (Allen & Hanchon, 2013; McConaughy & Ritter, 2002). About 20% of the 172 practicing school psychologists who participated in the survey reported they included only one or two sources of data in their ED evaluations; about 5% included no data. Fortunately, about 60% of participants included four or five sources of data. These findings invite future research on what school contexts, graduate training, and presenting behavior problems are associated with the comprehensiveness (and hence lawfulness) of ED evaluations.

One can understand why ED evaluations that include few or no sources of data are appealing to practitioners given the difficulty of synthesizing and reconciling the sometimes contradictory or incomplete results that can come from a comprehensive evaluation. Acknowledging this difficulty, advocates of best practices in ED evaluations encourage practitioners to construe diverging data as evidence of environmental influences on behavior: “Differing views about children’s behavior do not mean that one informant is right and the other is wrong. Instead, differing perspectives underscore the need for multiple information sources to assess children’s functioning across settings” (McConaughy & Ritter, 2002, p. 701). Additionally, best practices for ED evaluations emphasize discerning what motivates the child and his or her strengths so that this information can be used in intervention planning. In sum, the lawful and recommended practice is not to reduce the amount of data in ED evaluations; rather, practitioners need to find resources to enable them to engage in lawful and complex decision making that explains patterns within the data.

Cognitive Biases May Contribute to Disproportionality

Bias in complex decision making. Professional judgment involving complex information – like the results of comprehensive ED evaluations – often yields predictable patterns, also known as *bias*. Bias is “systematic, divergent decisions based upon comparable information (e.g., assessment data) as a function of group membership (e.g., racial, gender, or cultural differences)” (Huebner, 1991, p. 50). Some scholars argue that the uncertainty in and complexity of school evaluations make them unjustifiably susceptible to bias (e.g., Barnett, 1988; Davidow & Levinson, 1993; Redfield & Kraft, 2012; Ysseldyke & Algozzine, 1983; Ysseldyke, Algozzine, & Richey, 1982).

Although there may be some instances of overt bias, most bias among those in helping professions tends to be implicit (e.g., Sue et al., 2007) and occurs because “clinicians as well as people in general may not be fully aware of what they are thinking once a stimulus has been presented to them” (Dumont, 1993, p. 197). For instance, evidence from experimental research and social indicators shows persistent racial bias

against Blacks despite a decrease in self-reported attitudes of racism in the general public (Whaley, 1998). This paradox undergirds the theory of aversive racism, which posits that some liberal Whites unconsciously engage in veiled racism that allows them to maintain their prosocial self-image. Some argue that in clinical settings, “racial bias influences ‘well-intentioned’ White mental health professionals’ judgments about Blacks and leads to subsequent racial discrimination with severe consequences” (Whaley, 1998, p. 52). Indeed, evidence from the 1995 multisite MacArthur Violence Risk Assessment Study showed that the strongest predictor of racial disparities in the diagnosis of adults with schizophrenia was the research interviewer’s perception of the client’s honesty and trustworthiness, a construct vulnerable to cultural biases (Eack, Bahorik, Newhill, Neighbors, & Davis, 2012). Likewise, some evidence suggests that Black and White clinicians tend to attribute hallucinations and paranoid/suspicious attitudes more often to Black patients than White patients in assessing for schizophrenia (Trierweiler et al., 2000). Such racial disparities have been demonstrated repeatedly with regard to diagnosis of psychotic disorders (Schwartz & Blankenship, 2014). Some researchers have examined whether such covert biases occur between children and their service providers. Indeed, nationally representative data shows that elementary school teachers rate Black students as having more externalizing behavior symptoms than European American students and that ethnic differences in ratings increase with age, perhaps because over time “racial stereotypes and expectations become more salient” (Miner & Clarke-Stewart, 2008, p. 783). Such findings warrant further investigation of covert biases in special education evaluations.

Both overt and implicit biases diminish the quality and equity of decisions because bias impairs impartial consideration of the classification criteria and data (Garb, 2005) and objective hypothesis testing (O’Reilly et al., 1989). Scholars across various psychological fields have attempted to open the “black box” of decision making to understand better the validity and reliability of psychological diagnoses (e.g., López, 1989; McDermott, 1981; Rehder & Kim, 2009). One critical component of evaluations is

the determination of causation – i.e., what causes the presenting problems? Meehl (1973) aptly described this inquiry as a search for the “differentiating causal agent, the thing which is true of [the client] and not of the others who have remained ‘healthy’” (p. 248). Under IDEA, METs must infer the causes of problems giving rise to referrals for initial special education evaluations. Because there is no empirically valid method for determining causation in ED evaluations (Katsiyannis & Maag, 2001), causal inferences are one potential conduit for implicit biases in diagnostic reasoning.

Numerous types of biases have been investigated by cognitive science researchers. Seminal research by Kahneman and Tversky (1977) shows that: (a) “errors of judgments are often systematic rather than random, manifesting bias rather than confusion;” (b) there are common and predictable biases in the professional judgment of experts across areas of specialty; and (c) erroneous intuitions are compellingly attractive even when “a person is fully aware of their nature” (p. i-ii). In addition to these general trends, there are specific types of biases that frequently occur in human decision making. One form of bias is the cognitive heuristic, which is a rule-of-thumb or mental shortcut that “reduce[s] the complex task of assessing probabilities and predicting behavior to simpler judgmental operations” (Davidow & Levinson, 1993, p. 354). Heuristics may help people make quick, efficient judgments in difficult situations (e.g., paramedics or military personnel; for review see Klein, 2008), but they also may lead people to ignore important information and reinforce inaccurate stereotypes based on race, religion, gender or other statuses protected by law (Whaley, 1998). Prior scholars have recognized heuristics as a source of error in professional judgment generally (e.g., Kahneman & Tversky, 1977) and in special education diagnoses specifically (Davidow & Levinson, 1993; Fagley, 1988; McDermott, 1981). Below, I briefly explain four specific heuristics that can distort data analysis in special education evaluations.

Confirmation bias. Confirmation bias is the tendency to seek opinions and facts that support existing hypotheses and beliefs (O’Reilly et al., 1989). According to information processing theories, people have limited capacities to attend to information,

which causes them to search for and selectively pay more attention to evidence that supports initial impressions or hypotheses, and to the extent that contrary evidence comes to their attention, they find reasons to discredit it.

Within special education research, studies investigating the influence of confirmation bias have produced conflicting results. Ysseldyke and colleagues investigated the influence of confirmation bias in special education (e.g., Ysseldyke et al., 1981; Ysseldyke & Algozzine, 1982). Their findings have been cited frequently as evidence that referral statements unduly bias special education evaluations (Burns, 1992). In a series of analogue studies, they experimentally manipulated the type of referral (and other variables) and presented assessment data that unambiguously indicated the student's performance was within the average range. Arguably, an unbiased evaluator would have determined the students did not qualify for special education services. In one study involving educators and school psychologists, "students referred for behavior problems were significantly more often diagnosed and labeled as emotionally disturbed than students referred for academic problem;" no similar effect was found for academic referrals and learning disability (LD; Ysseldyke et al., 1981, p. 174). In another study, approximately 52% of participants found the student eligible for special education services even though the data were in the normal range (Ysseldyke & Algozzine, 1982).

In a similar study, 224 school professionals evaluated whether a hypothetical student would have academic difficulties; results showed that participants were more likely to predict math difficulty if the referral problem was academic rather than behavioral; they were more likely to predict reading difficulties for girls (but not boys) under the same circumstances (Algozzine & Ysseldyke, 1981). The participants' "decisions to classify a child as ED were influenced by that child's reported behavior in a referral statement" (Ysseldyke & Algozzine, 1981, p. 433) but the referral had no apparent influence on classifications for LD or intellectual disabilities (ID). Based on these studies, Ysseldyke and colleagues concluded, in part, "referral information is given

considerable weight in making outcome decisions” (Ysseldyke et al., 1981, p. 174; Ysseldyke & Algozzine, 1983).

O’Reilly and colleagues (1989) also have found evidence suggesting that referral information triggers a confirmation bias in special education evaluations. In their experimental study, 40 school psychologists in Arizona apparently did not utilize objective-hypothesis testing when evaluating hypothetical case files. The results showed “a significant tendency for eligibility judgments to mirror the stated reason for referral” (O’Reilly et al., 1989, p. 132-133).

Not all researchers, however, have found evidence of confirmation bias when investigating the influence of referral statements on special education evaluations (e.g., Huebner & Cummings, 1985; Huebner, 1990). Attempting to reconcile these discrepant findings, some scholars argued that referral information *and* test data typically influence diagnostic outcomes. When participants seemingly disregarded test data in studies by Ysseldyke and colleagues, they may have done so because the data were “unusually ambiguous” (Huebner, 1991, p. 56) or test observations suggested problematic test performance (Burns, 1992). Other methodological differences among the studies may also account for the contradictory results – e.g., degree of similarity between study samples, referral statements, and types and format of assessment data presented to participants (Burns, 1992; Huebner, 1991). One review of the literature in this area noted how difficult it is to reconcile discrepant findings because studies of special education “decision-making have been virtually devoid of theoretical underpinnings” (Huebner, 1991, p. 59).

Memory heuristics. Memory heuristics, such as the availability and representativeness heuristics, can “influence our subjective probability estimates of various outcomes without reference to their actual frequency” (Triplet, 1992, p. 35). For example, school psychologists that work primarily with students with behavior problems may be likely to over-estimate the prevalence of behavior problems among the general population because they can vividly and easily recall many behavior problems, leading

them to disregard the prevalence of behavioral disorders. Such biased decision making exemplifies the availability heuristic. “If instances are easily brought to mind, then the frequency is judged as high and the probability is judged as great” (Fagley, 1988, p. 312). This bias may be rooted in the structure of human memory; two factors that trigger the availability heuristic are the quantity (high frequency) and quality (vividness) of these memories. Similarly, if school psychologists decide whether a student has ED by comparing them to their mental prototype of a typical child classified as ED, then they are utilizing the representativeness heuristic (Fagley, 1988). This heuristic can be especially problematic if the prototype is based on outdated, unreliable, or incomplete information. Both the availability and representativeness heuristics endanger the validity of evaluations because school psychologists may be ignoring other salient facts unrelated to the heuristics.

These memory heuristics have received little attention among researchers of special education evaluations, but their influence on differential diagnoses has been investigated in other fields. For instance, researchers of discrimination in the medical field found in experimental studies that “homosexual patients were perceived as having AIDS even when their symptoms were not consistent with this diagnosis” (Triplet, 1992, p. 315), suggesting that the decision makers (i.e., college students in psychology courses) relied upon representativeness and availability heuristics to diagnose the patient. Some scholars implicate the availability heuristic in illusory correlations (Fagley, 1988), which are the “tendency to assume falsely that a relationship exists between two variables, such as an observed symptom and a diagnostic classification” (Gnys, Willis, & Faust, 1995, p. 60). How illusory correlations influence, if at all, inferences in ED evaluations in an area for future research.

Anchoring and adjustment biases. Anchoring and adjustment biases occur when decision makers are influenced by the order in which they consider data (Fagley, 1988). For instance, if the school psychologist’s record review reveals that a first grade student has been suspended for violent behavior, the psychologist may construe ambiguous play

observed at recess behavior as aggressive. If school psychologists allow the order in which they collect or consider data to influence them, they introduce arbitrary and capricious error into eligibility determinations. As with the memory heuristics, the influence of these heuristics on special education evaluations has received little attention.

Other sources of cognitive error. The likelihood of schools qualifying students for special education varies depending upon where students live (Singer, Palfrey, Butler, & Walker, 1989). Such inconsistencies may be due not only to cognitive heuristics, but also to important differences in diagnostic styles. For example, practitioners may weigh differently the importance of diagnostic cues (e.g., referral, various forms of formal assessment data, educational history, and family background; e.g., Davidow & Levinson, 1993; McDermott, 1981). Some practitioners may have a tendency to give much credence to informal teacher observations whereas others consider norm-referenced results to have the most authority. Such inconsistency tends to occur not only across practitioners but also across the caseload of an individual practitioner. Some practitioners “will include or ignore certain kinds of diagnostic information because of political pressures, time restraints, or parental pressure” (Davidow & Levinson, 1993, p. 353). The lack of standardization in the weighing of diagnostic data may give rise to meaningful differences in professional judgment that are inconsistent with the objectives of IDEA.

Another potential source of decision making inconsistency is theoretical orientation (Barnett, 1988; Davidow & Levinson, 1993; McDermott, 1981). Theories may frame decision making, such that practitioners with different theoretical orientations (e.g., cognitive-behavioral versus psychodynamic) may view the same facts and render contradictory – yet theoretically sound – diagnoses. Some scholars contend that such inconsistencies cannot be excused by the assurance that professional judgment undergirds the variance (Davidow & Levinson, 1993; McDermott, 1981) and advocate for an alternative solution: special education diagnosis per actuarial software (Davidow & Levinson, 1993). Their proposal harkens back the debates in the 1950’s and 60’s over clinical versus statistical diagnostic prognosis (e.g., Meehl, 1954). This debate calls into

question whether humans are capable of reliably collecting and processing information per IDEA. More recently, Youngstrom (2013) advocates for psychologists to use an evidence-based medicine approach to clinical assessment by integrating patient preferences and evaluation data with simplified Bayesian methods (i.e., a slide ruler). This hybrid approach to psychological assessment arguably enables practitioners to apply objective criteria like base rates to individual cases, which mitigates the influence of heuristics and other biases.

ED Evaluations May Be Vulnerable to Cognitive Biases

School psychologists working in public schools frequently interpret and apply special education law to help decide whether a student qualifies for special education services under IDEA. Many of the key terms in IDEA and its regulations are ambiguous. A law is ambiguous if it permits more than one reasonable meaning. The meaning of administrative rules (i.e., regulations) and statutes evolves through “three stages: drafting, implementation (including enforcement), and interpretation by the courts” (Uniform Law Commission, 2013). School psychologists and other members of METs contribute to this meaning-making process in the second stage: implementation of regulations and statutes. METs discern the meaning of IDEA to determine whether a student’s assessment results meet the express legal criteria for qualification. In this respect, MET members are similar to judges because they interpret and apply law. They differ from judges, however, because judges are trained to follow canons of construction when determining the meaning of law.

Cannons of construction are a system of rules for interpreting legal texts like statutes and regulations. These cannons can vary slightly across jurisdictions, but there is a model form called the Uniform Statute and Rule Construction Act (National Conference of Commissioners of Uniform State Laws, 1995). One particularly important and commonly adopted interpretation rule is the plain meaning canon described in Section 18 of the Act. According to this rule, judges should interpret a statute literally if the meaning is obvious from the text, but they should not apply the obvious meaning if it

leads to an unjust result or contradicts the express purposes and objectives of the statute. Likewise, if a statute has ambiguous terms, those terms should be interpreted in ways consistent with the express purposes and objectives of the statute. In other words, METs should interpret IDEA so that the meaning given to ambiguous language advances the express purposes of IDEA. IDEA has several express purposes, including “ensur[ing] that all children with disabilities have available to them a free appropriate public education [FAPE] that emphasizes special education and related services designed to meet their unique needs and prepare them for further education, employment, and independent living” (20 U.S.C. § 1400(d)(1)(a), 2015). Notably, this purpose is broad and includes skills beyond academics.

Federal and regulatory background. The federal criteria for special education disabilities generally and ED specifically have key terms that are ambiguous. IDEA defines *child with a disability* to mean a child:

- (i) with intellectual disabilities, hearing impairments (including deafness), speech or language impairments, visual impairments (including blindness), serious emotional disturbance (referred to in this chapter as “emotional disturbance”), orthopedic impairments, autism, traumatic brain injury, other health impairments, or specific learning disabilities; and (ii) who, by reason thereof, needs special education and related services. (20 U.S.C. § 1401(3)(A), 2015).

Thus, IDEA protects only those students who qualify as having at least 1 of the 14 conditions identified therein, and who need special education because of that condition. Notably, IDEA does not define or otherwise provide guidance for interpreting two key terms: *by reason thereof*, which indicates causation (e.g., May, 2009), or *need*. MET members may reasonably disagree about what constitutes a *need* for special education. How much need is enough? How do you measure the need? How long must the need exist? IDEA does not answer these questions. Further, *by reason thereof* may be interpreted to mean that the disability is the singular cause of the student’s need for

services or that the disability is one of several factors contributing to the student’s need for services –i.e., mixed causation.

Further complicating matters, IDEA prohibits METs from qualifying a student as disabled “if the determinant factor for such determination is” lack of appropriate instruction in math or reading, or limited English proficiency (20 U.S.C. § 1414(b)(5), 2013). Again, there are no guidelines for what methods or standards METs should use to differentiate a *determinant factor* from a lesser contributory factor.

Not only must recipients of IDEA funding comply with Congress’s mandates, but they also must follow the rules (i.e., regulations) of the Department of Education (DOE). Congress authorized the DOE to publish rules to help states implement IDEA (20 U.S.C. § 1406(a), 2015). The DOE has defined each of the 14 disabilities Congress specified in IDEA. It defined ED as:

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 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; and (E) a tendency to develop physical symptoms or fears associated with personal or school problems. (34 C.F.R. § 300.8(c)(4), 2015)

These five criteria include undefined terminology that arguably supports more than one reasonable interpretation. For instance, *adversely affects* indicates causation but it does not say how much or what kind of causation suffices, just like the causal language (i.e., *by reason thereof* and *determinant factor*) in IDEA.

Perhaps the most notoriously ambiguous ED regulation is the social maladjustment (SM) exclusion. The DOE ruled that ED “does not apply to children who are socially maladjusted, unless it is determined that they have an emotional disturbance” (C.F.R § 300.7(a)(9), 2015). In other words, the student’s need for special education

cannot be due solely to SM. Again, the DOE has not defined SM or methods METs should use to differentiate ED from SM. Prior authors have mistakenly attributed the SM exclusion to Congress (e.g., Olympia et al., 2004), but the DOE is responsible for this rule. This distinction is important because it clarifies that the DOE has the power to revise or eliminate the SM exclusion. Communications from the DOE, however, indicate that it has no intention to do so (71 Fed. Reg. 46, 549-50, 2006).

In sum, students qualify as ED if the MET determines that (a) they meet one of the five criteria for ED; (b) ED, rather than SM or something else, *adversely affects* their *education performance*; (c) *by reason thereof* they need special education services; and (d) that lack of appropriate instruction in math or reading, or limited English proficiency was not a *determinant factor* in this determination.

Judicial interpretations of ED criteria. The federal courts of appeals have grappled with giving meaning to undefined terms in ED criteria. In the cases summarized below, parents commenced due process proceedings to challenge the school district's determination that their child did not qualify for special education. Largely, the parties agreed on the background facts of the case – i.e., what happened, when, where, to whom, etc. Their dispute was about whether these facts constituted sufficient evidence to meet the legal definition of “child with a disability” (20 U.S.C. § 1401(3)(A), 2015; 34 C.F.R. § 300.8, 2015) and consequently, entitled the student to special education. The three cases summarized below are different, however, in how the judges interpreted the causal mandate and other key terms in the ED eligibility criteria.

In 1995, the Ninth Circuit held in the *Wartenberg* case that mixed causation did not preclude IDEA coverage. The student in that case, Jeremy, had severe conduct disorder and attention deficit hyperactivity disorder (ADHD). His eligibility for special education hinged on whether SM or LD caused his behavior problems, which in turn prevented him from learning at school. At trial, the school psychologist testified that the primary cause of Jeremy's problems was SM, which he defined as conduct that Jeremy knew was wrong and could control if he wanted to do so. The school psychologist also opined that “ascribing

Jeremy's poor performance to some neurochemical deficiency was overly speculative, because Ritalin therapy did not provide relief from symptoms he would have expected if Jeremy had a neurochemical imbalance" (*Wartenberg*, 1995, p. 888). Rejecting the school psychologist's theory, the Ninth Circuit held that the primary cause of Jeremy's learning problems was a disability, not merely SM. Relying upon an opinion letter written by the DOE, the Ninth Circuit reasoned that when a covered condition (like ED) and a non-covered condition (like SM) both caused the student's learning problems, then the student qualified for special education if the *primary* cause of the learning problem was a covered condition. The court did not explain how it distinguished a primary cause from a lesser contributory cause.

A few years later the Fourth Circuit articulated a more stringent interpretation of the causation requirement in *Springer* (1998). It ruled that Edward Springer's behavior problems (e.g., criminal activity, skipping classes, fighting) amounted to SM rather than ED. It cited for support that Edward's behavior problems did not begin until tenth grade, when he started using illegal drugs. Further, it concluded that even if Edward had ED, there was insufficient evidence that ED rather than his truancy, drug abuse, and delinquent behavior adversely affected his poor educational performance. Notably, the Fourth Circuit did not identify what kind of evidence *could* prove that Edward's school problems resulted from his alleged depression rather than SM. Other judges have also denied students special education because parents were unable to "untangl[e] cause and effect in the context of drug use, misbehavior, and depression" (*Nguyen*, 2010, p. 51).

Around the same time, the Second Circuit heard a case with relatively similar facts as *Springer* but it reached a very different outcome. In *Muller* (1998), Treena Muller, a ninth-grader, had conduct problems at school (skipping class, not completing assignments) and at home (disobeying parents, breaking curfew). Her problems escalated to an attempted suicide, which led to her admission in a psychiatric hospital for 25 days and then three weeks of day treatment. Her providers diagnosed her with conduct disorder, depression, and post-traumatic stress disorder (PTSD) related to early childhood

experiences and recommended long-term placement in a residential treatment and educational facility. On appeal, the Second Circuit held that expert psychological testimony that Treena's emotional problems adversely affected her educational performance, along with evidence that educational performance improved in the therapeutic setting, were sufficient to satisfy the burden of proof for causation. In other words, treatment responsiveness and expert testimony were *de facto* evidence of causation. Interestingly, only students able to access treatment and present evidence of improvement can avail themselves of the Second Circuit's liberal construction of the causation requirement.

In sum, these federal court of appeals cases demonstrate how different interpretations of ambiguous federal ED criteria can lead to inconsistent outcomes among students with similar characteristics. Arguably, school psychologists and other members of METs may rely on cognitive heuristics and other covert biases to discern and apply these ambiguous legal terms in special education evaluations. How much, if at all, such inferences contribute to racial disproportionality in special education is an unanswered empirical question.

Data Diagnosticity

Categorical reasoning facilitates various forms of decision making, such as diagnosing patients and predicting performance of students in school. Arguably, "categories are a central means by which old experiences guide our responses to new ones" (Rehder, 2010). Conceptualizations of categorization have drawn heavily upon theories of psychological essentialism, which posit that "things have essences that make them what they are" (Ahn & Kim, 2001, p. 29). One way of determining the essence of an object is to consider its features (e.g., parts, attributes, dimensions; Sloman, Love, & Ahn, 1998). Categorical reasoning involves weighing and sometimes inferring the features of a specific artifact or concept to determine whether it satisfies the general criteria for group membership. Classifiers tend to engage in feature-weighing, which is evaluating the importance of a feature relative to other features within the category. The

weight of a feature can be measured subjectively by how essential or central its role is within the network of features for the category or “the degree to which it lends conceptual coherence” (e.g., Sloman et al., 1998, p. 190). One method for measuring centrality is to rate how easily one can transform a mental representation of an object by eliminating or replacing the target feature without changing other aspects of the representation. For example, it is not easy to imagine a cat without a head, so head is a central feature of a cat. Thus, head has a low mutability rating. There is an inverse relationship between feature centrality and mutability.

The weighing of features has implications for diagnostic decisions in educational and psychological evaluations, especially relating to issues of unlawful bias. For instance if student race is a feature that MET members weigh heavily in determining eligibility for special education, then the evaluation outcomes would likely violate antidiscrimination legislation. Because biases among helping professionals like school psychologists probably are more implicit than covert, study participants seem unlikely to identify student race as highly diagnostic. A more plausible approach for ascertaining bias would be to ask participants to rank the diagnosticity of sources of data and less controversial student characteristics like sex and age, which may provide an indirect measure of psychometric bias, teacher bias, etc. For example, Knoff (1983a) investigated whether school psychologists and special educators prioritized intelligence test results above all other types of data when conducting evaluations generally (i.e., not just ED evaluations). These practitioners (and practitioners-in-training) ranked observations, communication skills, and interviews as most important in evaluations. Although Knoff (1983a) did not evaluate the cultural loading of these sources of data, studies like this may shed light on evaluation processes that contribute to disproportionality.

Using a different methodology, Becker, Paternite, and Evans (2014) examined the relative importance secondary special educators placed on the academic, behavioral, and mental health components of ED evaluations. They reported that aggression directed at an adult was the most important relative factor in ED evaluations, followed by aggression

directed at a peer, disruptive classroom behavior, or mental health diagnosis of bipolar and schizophrenia. Perceptions of aggression and disruption can implicate cultural and racial biases (Whaley, 1998); mental health diagnoses are also vulnerable to these biases (Schwartz & Blankenship, 2014). For these reasons, practitioners' ratings of data diagnosticity may provide insight into the cognitive processes that lead to evaluation outcomes and the impact, if any, that racial bias has on ED evaluations.

Extant Research on Bias in ED Evaluations

Society has a strong interest in understanding whether covert cognitive biases influence special education evaluations because these mental shortcuts may be unintended conduits for unlawful discrimination. Yet, few researchers have investigated the impact of student demographic characteristics on eligibility decisions. In 1981, Ysseldyke, Algozzine, Regan, and McGue found that sex, socioeconomic status (SES), and attractiveness apparently did not influence the assessment outcomes of 159 school personnel (e.g., administrators, school psychologists, teachers) when they evaluated a hypothetical student for LD. In a different study with 189 school psychologists in Illinois, race apparently did not influence assessment of intellectual disabilities (Huebner & Cummings, 1985). Additionally, results suggested that student urbanicity (rural versus suburban) did not influence 56 school psychologists' evaluation of hypothetical assessment data (Huebner & Cummings, 1985). For autism and ADHD, SES bias but not racial bias existed among evaluation outcomes from speech-language professionals, school psychologists, and private child psychiatrists in a simulation study (Cuccaro et al., 1996). The general trend of these published studies shows that MET decisions do not reflect racial bias, although some degree of SES bias may occur.

For the six reasons set forth below, the extant research on covert discrimination in special education evaluations is insufficient. First, it is outdated. It was published last century before the passage of No Child Left Behind and the 2004 amendments to IDEA, which dramatically increased accountability for academic performance in public schools. This legislation emphasizes the importance of data-based decision making, which may

have significantly changed how districts conduct special education eligibility assessments. Thus, the generalizability of prior findings to modern assessment practices is an unanswered empirical question.

Second, the research findings were based on participants evaluating hypothetical student data instead of conducting their own dynamic assessments and analyzing the resulting data. The researchers did not measure how professional judgment influenced choice of instruments, raters, and the sequence of assessment, all of which may have “profound effects on interpretations and outcomes” (Barnett, 1988, p. 663). Further, the generalizability of the findings rest upon the presumption that cognitive processes tapped in analogue research settings are the same processes that occur in real-world decision making, a presumption some scholars challenge (e.g., Galotti, 2007). For instance, some of the clinical cues available in live diagnoses are not available in analogue research (Huebner, 1991). Thus, the differences between analogue and naturalistic decision making raise doubts about how these findings relate to actual practices in the field.

Third, this body of research examined decision making by individuals, not groups. By law, special education diagnoses must result from collaborative decision making by METs, which must include parents (20 U.S.C. § 1414(d)(1)(B), 2012). None of the participants in the aforementioned samples were parents. Further, scholars hypothesize (and common sense dictates) that group dynamics impact evaluation outcomes. For instance, the power of the majority, influence of dissent, shared norms, and polarization are believed to influence MET decision making (Gutkin & Nemeth, 1997). Indeed, prior research suggests that team decision making for evaluations produces less variance in outcomes than evaluations conducted by individuals (Pfeiffer, 1982). School psychologists tend to be influential in team evaluations (Gilliam, 1979; Knoff, 1983a; Mehan, 1991; Yoshida et al., 1978), so analogue research conducted without school psychologists may not reflect outcomes in the field. This difference between individual and group decision making greatly diminishes the value of the extant research on bias in special education evaluations.

Fourth, the study participants did not have to implement their decisions or otherwise experience the consequences of the evaluation. Such insulation from practical considerations is unlikely to exist for METs. One qualitative study of MET decision making found that “resource availability was seen as having a greater influence on students identified than the degree to which students were considered as having an actual disability” (Mellard, Deshler, & Barth, 2004, p. 236). It is reasonable to speculate that special education evaluations are impacted to varying degrees by anticipated effects on available resources (Huebner, 1991) and relationships among MET members. The extant analogue research does not address these factors.

Fifth, the findings may reflect participants’ desire to provide socially acceptable eligibility decisions rather than the determinations they would make outside the research setting (e.g., Huebner, 1991). Finally, the studies that found special education evaluators tend not to be influenced by student characteristics like race and SES are contradicted by an extensive body of research in clinical settings that suggest that such characteristics frequently influence diagnoses (e.g., Lopez, 1989; Neighbors et al., 2003; for a review see Whaley, 1998). Long-standing and widespread evidence of disproportionality within special education (Donovan & Cross, 2002; Sullivan, 2011) also suggests problems with the diagnostic process. Some scholars have argued that implicit bias in professional judgment is a plausible explanation for disproportionate representation of racial minorities in particular (e.g., Redfield & Kraft, 2012). The scant special education research conducted in analogue settings is an insufficient rebuttal to these charges.

Despite these limitations, the extant research investigating discrimination in special education is commendable in light of the considerable barriers inherent in this line of inquiry. Indeed, some argue that diagnostic processes, including professional judgment, are “fuzzy, unbounded problem spaces” (Dumont, 1993, p. 197) that have not been grounded in empirical data (e.g., Davidow & Levinson, 1993). Barnett (1998) is one of the few school psychology scholars who have tried to operationalize professional judgment; he defined it as “bridging the gap between the knowledge base and actual

strategies used to solve real-life problems. It involves not only logic and problem-solving capabilities, but also an appraisal of how the hypotheses formed concerning a certain client relate to an appropriate body of research” (Barnett, 1989, p. 659). Even if one accepts this definition, professional judgment is still difficult to measure and evaluate. Unlike the diagnosis of a broken leg, mental states are latent constructs for which there is no tangible referent. Thus, in psychology

the very use of the term misdiagnosis implies a clear idea of which diagnosis is correct and which one is incorrect. Unfortunately, this assumption is unfounded; in the absence of a gold standard, if two reasonable diagnostic processes disagree, it may not be clear that one procedure is superior (more accurate or valid) to the other. (Neighbors et al., 2003, p. 239)

Some argue that sophistication or reliability of assessment methodology (e.g., structured and thorough interview versus “naturally occurring clinical impressions;” Neighbors et al., 2003, p. 239) relate to the likelihood that the diagnosis is accurate. However, the veracity of this presumption cannot be demonstrated empirically.

In sum, the extant research and theory on cognitive biases in complex decision making raises concerns about whether school psychologists ignore salient data in special education evaluations but there is only scant, outdated research examining whether covert biases result in unlawful discrimination. Notably, there are no published studies investigating the influence of cognitive biases in causal inferences. For instance, does a legal mandate for causal inferences lead METs to find illusory correlations or mistake correlation for causation? It is imperative that these empirical questions be investigated because the accuracy and equity of eligibility determinations partially indicate how effective IDEA is in protecting the educational rights of students with severe mental illnesses and other disabilities.

Accordingly, experimental research is needed to elucidate the cognitive processes that lead to special education evaluation outcomes. As key members of METs, the decision making of school psychologists is an important area of inquiry. The primary

purpose of this study is to test experimentally how much, if at all, racial bias impacts school psychologists' decision making in evaluations for ED. Results will provide some evidence about the role of bias in the persistent national trend of over-representation of Black students in the ED category. Secondary goals of this study are to examine how school psychologists weigh data in ED evaluations in order to understand better how they navigate the ambiguous ED criteria and which features of the ED category are most important in eligibility decisions. These results also may help inform understanding of causes of disproportionality.

Chapter 3 Methods

Experimental Design

The study had a between-subjects design with three conditions that varied in their degree of ambiguity. The first condition involved low-ambiguity data that do not meet criteria for ED; the second condition involved low-ambiguity data that meet criteria for ED; and the third condition involved highly ambiguous data. For each condition, participants were randomly assigned to decide whether a Black or White student qualified as ED. Race was the only independent variable for each ambiguity condition. All other aspects of the evaluation were held constant to control for confounding variables. The primary dependent variable was the evaluation decision – i.e., qualifies or does not qualify as ED. Participants were also asked to rate their confidence, rank the diagnosticity of data, and identify additional data they wished were included.

Sampling

The study had a probability sampling design. A precise power calculation could not be conducted because there was insufficient extant information about the population (e.g., number of school psychologists conducting evaluations for ED nationally, variance, standard deviation) and the impact, if any, of racial discrimination on ED evaluation outcomes (e.g., effect size). Following general guidelines for experimental social science research (e.g., Hill, 1998) and the central limit theorem (e.g., Howell, 2010), each race condition had 30 participants for a total of 60 participants.

The National Association of School Psychologists' (NASP) membership directory was used to identify potential participants. Recruitment was restricted to one state so that all participants applied the same regulations and laws in evaluations for ED.

Pennsylvania was selected because it had a large number of school psychologists who could be participants (1,652) and adopted the federal regulations for ED evaluations, potentially increasing the generalizability of the study's findings to the other states that have done the same. Other inclusion criteria for the study were that participants had to (a) provide an electronic mail (e-mail) address in their directory contact information; and (b)

be school psychologists participating in or supervising special education evaluations in Pennsylvania. Members were excluded from the recruitment process if their directory information identified them as professors, retired, graduate students, teachers, superintendents, assistant superintendents, principals, assistant principals, data analysts, self-employed, or working in the assessment industry (e.g., Pearson). Notably, special education directors, department heads, and school-based mental health professionals were not excluded as potential participants. Altogether, 857 members fit the inclusion criteria and were eligible to receive a recruitment letter.

Participants

The recruitment letter (Appendix A) was designed to maximize participation. In light of meta-analytic findings suggesting that incentives improve response rates in web-based studies (Göriz, 2010), a gift certificate of \$20 to Starbucks, iTunes, Amazon.com, or Habitat for Humanity (charitable donation) was offered for completion of the study. This incentive was described as a potential loss (i.e., “Failure to complete the study by May 30, 2014 will result in loss of the gift certificate”) consistent with loss aversion theory, which posits greater motivation to avoid a loss than to gain a reward (Tversky & Kahneman, 1991). The scarcity principle was also used to maximize participation (Fan & Yan, 2010). Specifically, members were told they were “part of a select group of school psychologists invited to participate” and a deadline for participation was set. Other strategies for increasing participation included: (a) having participants select their incentive at the beginning of the study rather than at the end (Reips, 2010); (b) offering different types of gift certificates to increase the likelihood that one would be sufficiently rewarding (Porter & Whitcomb, 2003); and (c) estimating how long it would take to complete the study (Reips, 2010). A meta-analysis of web survey research suggested that response rates increase when reminders are sent (Lozar Manfreda, Bosnjak, Berzelak, Haas, & Vehovar, 2008). Invitees received up to four reminders depending on how early in the recruitment process they were solicited.

To obtain the necessary 60 participants, 235 invitations were sent from May 15, 2014 to June 6, 2014. The overall response rate was 26%. The final sample included 42 (70%) women. The racial composition of the sample was White (95%), Black (2%), and Asian (2%). One participant (2%) did not report race.

Instrumentation

Vignettes. I developed all of the instruments used in this study (see Appendix B). I used a written vignette methodology to compare the likelihood of school psychologists identifying as ED Black versus White, male students in fifth grade. Some benefits of vignette methodology are the cost effectiveness and efficiency of study administration, as well as experimental manipulation of the variable of interest, which improves internal validity – i.e., reducing the likelihood that the observed effect is due to a confounding factor (e.g., Heverly, Fitt, & Newman, 1984). The increased internal validity, however, frequently comes at the expense of external validity – i.e., generalizability of the findings to actual practitioners in the field. Ideally, a research agenda includes both naturalistic studies with strong external validity and experimental studies with strong internal validity, and all results converge.

Little research has been published on the development of vignettes for investigating clinical judgment. Some scholars have recommended that “vignettes should be short and informative” to prevent participant fatigue (which would reduce reliability) and “reflect the mid-range level of functioning, so that sufficient variance is generated to reveal a factor’s influence” (Heverly et al., 1984, p. 47). Scholars of clinical judgment analysis recommend trying to represent the variables as if they were naturally embedded in real-world decision making contexts (Skånér, Bring, & Strender, 2004). Some experimental social psychology researchers emphasize adequate stimulus sampling, which “refers to the use of multiple instances of a stimulus category in research” to enhance the strength of construct validity (Wells & Windschitl, 1999, p. 1115). Construct validity is especially concerning when “the use of only one stimulus to represent a category can confound the unique characteristics of the selected stimulus with the

category” (Wells & Windschitl, 1999, p. 1116). For instance, if a researcher were attempting to evaluate how well individuals remembered monosyllabic versus polysyllabic names, comparing recall for *Joe* versus *Mirajoul* would confound ethnicity with the construct of interest (i.e., number of syllables). To improve construct validity, researchers would need to contrast more than one set of names from differing cultures. In sum, one needs to balance (a) the need for a sufficient number of stimuli (i.e., vignettes) to sample adequately the construct, (b) the need for sufficiently complex vignettes that reflect the ecological context for decision making, and (c) participant fatigue and boredom, which reduce reliability and participation.

In this study, three factors were important considerations in the development of vignettes: ED, race, and ambiguity. ED is broad and vague construct due to the ambiguous language of federal law and regulations (Olympia et al., 2004). Students identified as ED are a heterogeneous group (Wagner et al., 2005). For pragmatic reasons, the construct was narrowed to male students in late elementary school with primarily externalizing problems. These restrictions were based on evidence that males are more frequently identified as ED than females (Wagner et al. 2005) and the controversy about how to distinguish externalizing disorders that qualify as ED from social maladjustment (e.g., Olympia et al., 2004).

Furthermore, the vignettes were drafted in light of federal requirements and best practices for ED evaluations. Specifically, federal law requires evaluations under IDEA to be multimethod and multisource (34 C.F.R. § 300.304, 2015). Best practices in school psychology require evaluations to include a review of school records, direct observation by the school psychologist (preferably systematic), standardized rating scales, and interviews of parents, teachers, and students about factors that may contribute to the problem or potential solutions (e.g., developmental, medical, and family histories, family stressors and strengths, student’s social and academic performance, etc.; McConaughy & Ritter, 1998). Accordingly, the vignettes in this study included these evaluation components. The results for achievement and intelligence tests were held constant across

vignettes in order to focus on the social, emotional, and behavioral aspects of ED evaluations.

As for race, only two categories – Black and White – were investigated in this study. White was selected as the comparison group consistent with research on disproportionality (e.g., Artiles et al. 2005). Race was represented in the vignettes by expressly identifying the students' race as Black or White. Additionally, Black students and their family members were given names that sounded Black in U.S. culture (i.e., Antwan, Jamal, Tyrell) and White students and their family members were given names that were ambiguous or sounded White (i.e., Andy, Jay, Tom). Hereinafter, the hypothetical students are referred to as “Student T” (Tyrell/Tom), “Student J” (Jamal/Jay), and “Student A” (Antwan/Andy). Prior research on racial discrimination suggests that names alone can trigger racial biases. For instance, Beltrand and Mullainathan (2003) found in an experimental study that employers reviewing job resumes selected fewer applicants with distinctively African American names for interviews. Similarly, Milkman, Akinola, and Chugh (2012) found that professors responded less frequently and less positively to requests for meetings from hypothetical students with racially distinct names.

The degree of ambiguity in the assessment results was also considered in drafting the vignettes. Two vignettes included assessment results that were intended to have low ambiguity: Student T should not have qualified for special education and Student J should have qualified. These vignettes were designed to assess participants' competency in evaluating ED under relatively ideal conditions. Accurate responses to these vignettes helped to rule out the possibility that disproportionate identification, if any, under the ambiguous condition was due to participants misunderstanding the task presented or general assessment incompetency. The third vignette, Student A, had highly ambiguous assessment results. Prior scholars in the field of public health have posited “the effect of race/ethnicity on interpretations of behavior is exacerbated when the behavior is ambiguous or open to multiple interpretations” (van Ryn & Fu, 2003, p. 250). There have

been no published studies in school psychology investigating how ambiguity in case vignettes interacts with bias, but some scholars have called for such research (e.g., Huebner & Cummings, 1985). The vignettes were piloted with thirteen individuals (e.g., graduate students and practicing school psychologists) until the vignettes had the intended level of ambiguity.

Time constraints were also considered in designing the vignettes. To maximize the response rate and reliability of responses, the study was designed to be completed in about 15-20 minutes. Thus, the vignettes were about 1,000 words.

Questionnaire. All participants answered the same questions after each vignette about their eligibility decision, confidence in their decision, diagnosticity of data, and what data, if any, did they wish had been included but was missing in the results. After completing the vignettes, the participants answered questions about their personal characteristics (gender, race) and work experiences (e.g., years of practice, urbanicity of school).

Procedures

I obtained approval from my university institutional review board. I administered the study via the Internet using Qualtrics (2014). Participants activated a URL link in the recruitment letter to launch the study module. I obtained informed consent via the first question. Participants were unable to return to prior screens (i.e., no “back” button); they had to answer each question to proceed to the next screen. The final screen presented a unique random code and instructed participants to e-mail the code to me to receive compensation. This code verified completion of the study and prompted payment, which was made electronically.

Analyses

To answer Research Questions 1, 2, and 3, a contingency table analysis was conducted in SPSS and the chi-square statistic was used to compare the observed counts with those that would be expected if there were no association between race and special education outcome. This nonparametric statistic is appropriate “if no expected cell

frequencies are less than one and no more than 20% are less than five” (Elliott & Woodward, 2007, p.115). Responses to the remaining research questions were analyzed with frequency counts.

Chapter 4 Results

Sample Characteristics

As shown in Table 1, the majority of the school psychologists was White, held Master's degrees, and worked in suburban public schools in Pennsylvania that did not report disproportionality in special education. About a third of the participants were in the first five years of practice; and about a third had between 11-20 years of practice. About half of the participants were involved in 21-75 special education evaluations in an academic school year.

Table 1
Participant and District Characteristics

Variable	n	%
Gender		
Male	18	30
Female	42	70
Race		
Asian	1	1.67
Black	1	1.67
White (non-Hispanic)	57	95
Refused to disclose	1	1.67
Highest degree of education related to school psychology		
M.A.	32	53
Ph.D.	8	13
Psy.D.	1	1.67
Ed.D.	8	13.33
C.A.G.S.	11	18.33
Years of practice		
0-5	19	31.67
6-10	11	18.33
11-20	23	38.33
21-30	4	6.67
31-40	3	5
41 or more	0	0
Employment setting		
Public school	58	96.67
Private school	2	3.33
Urbanicity		
Rural	5	8.33
Suburban	33	55
Urban	22	36.67
Number of evaluations per year		
1-20	10	16.67
21-50	22	36.67
51-75	17	28.33
76-100	9	15
101 or more	2	3.33
District disproportionality in ED evaluations		
Yes	3	5
No	50	83.33
Don't know	7	12.67

Race by Ambiguity

Question 1. *How much does student race influence school psychologists' evaluation decisions when data concerning social, emotional, and behavioral problems are highly ambiguous?*

H_o: There is no association between student race and evaluation outcome when data have high ambiguity.

H_a: Student race and evaluation outcome are associated when data have high ambiguity.

As shown in Table 2, of the 27 students (45%) qualified as ED, 15 were Black and 12 were White. Of the 33 students (55%) who did not qualify as ED, 15 were Black and 18 were White. The chi square analysis provided insufficient evidence to reject the null hypothesis that there was no statistically significant relationship between race and evaluation outcome, $\chi^2(1) = 0.61, p = 0.44$.

Table 2

Eligibility Determinations by Ambiguity Condition and Student Race

Ambiguity condition	Qualified as ED			Not qualified as ED		
	White Student	Black Student	Total	White Student	Black Student	Total
Low: not ED	0	0	0	30	30	60
Low: ED	25	27	52	5	3	8
High	12	15	27	18	15	33

Question 2. *How much does student race influence school psychologists' evaluation decisions when data concerning social, emotional, and behavioral problems have low ambiguity and do not meet criteria for ED?*

H_o: There is no association between student race and evaluation outcome when data have low ambiguity and do not meet criteria for ED.

H_a: Student race and evaluation outcome are associated when data have low ambiguity and do not meet criteria for ED.

All 60 students were determined not to qualify as ED. A chi square analysis was not conducted because the evaluation outcome was constant. The evidence was insufficient to reject the null hypothesis that race and evaluation outcome did not have a statistically significant relationship.

Question 3. *How much does student race influence school psychologists' evaluation decisions when data concerning social, emotional, and behavioral problems have low ambiguity and meet criteria for ED?*

H₀: There is no association between student race and evaluation outcome when data have low ambiguity and meet criteria for ED.

H_a: Student race and evaluation outcome are associated when data have low ambiguity and meet criteria for ED.

Of the 52 students qualified as ED, 27 were Black and 25 were White. Of the 8 students who were not qualified as ED, 3 were Black and 5 were White. The chi square analysis provided insufficient evidence to reject the null hypothesis that there was no statistically significant relationship between race and evaluation outcome, $\chi^2(1) = 0.58$, $p = 0.45$.

Confidence by Ambiguity

Question 4. *How confident are school psychologists in their evaluation decisions when the data are highly ambiguous?*

Among the 60 participants, 14 (23%) reported high confidence, 39 (65%) reported moderate confidence, and 7 (12%) reported low confidence. There was no statistically significant relationship between race and confidence level $\chi^2(1) = 1.52$, $p = 0.47$.

Question 5. *How confident are school psychologists in their evaluation decisions when the data have low ambiguity and do not meet criteria for ED?*

Among the 60 participants, 35 (58%) reported high confidence, 25 (42%) reported moderate confidence, and none reported low confidence. There was no statistically significant relationship between race and confidence level, $\chi^2(1) = 1.71$, $p = 0.19$.

Question 6. *How confident are school psychologists in their decisions when the data have low ambiguity and meet criteria for ED?*

Among the 60 participants, 43 (72%) reported high confidence, 17 (28%) reported moderate confidence, and none reported low confidence. There was no statistically significant relationship between race and confidence level, $\chi^2(1) = 1.05$, $p = 0.31$.

Data Diagnosticity by Ambiguity

Question 7. *What data do school psychologists identify as most diagnostic when data are highly ambiguous?*

As shown in Table 3, the type of data most frequently selected as diagnostic when results were highly ambiguous were Behavior Assessment System for Children (BASC) results (72%), family history (68%), and student interview (65%). Cognitive (5%) and achievement (7%) test results as well as development history (13%) were selected infrequently.

Table 3
Most Diagnostic Data by Level of Ambiguity

	Low ambiguity not qualified		Low ambiguity qualified		High ambiguity	
	n	%	n	%	n	%
Teacher interview	30	50	33	55	26	43
Student interview	7	12	13	22	39	65
School psychologist's direct observation	48	80	42	70	33	55
Classroom grades	25	42	19	32	19	32
Discipline history	22	37	24	40	12	20
BASC results	44	73	45	75	43	72
Cognitive testing results	8	13	4	7	3	5
Standardized achievement results	28	47	3	5	4	7
Developmental history	6	10	28	47	8	13
Family information	5	8	23	38	41	68
Attendance history	0	0	0	0	0	0

Question 8. *What data do school psychologists identify as most diagnostic when data have low ambiguity and do not meet criteria for ED?*

The majority (80%) of participants selected the school psychologist's direct observation results; and 73% selected the BASC results. Few participants (8%) considered family information or developmental history as diagnostic in this condition.

Question 9. *What data do school psychologists identify as most diagnostic when data have low ambiguity and meet criteria for ED?*

BASC results (75%) and school psychologist's direct observation (70%) were identified most frequently. There was a considerable increase in ratings for developmental history (47%) and family history (38%). The lowest ratings were for cognitive (7%) and achievement (5%) test results.

Additional Data by Ambiguity

Question 10. *What data do school psychologists wish were included when data are highly ambiguous?*

In response to this optional question, 62% of participants identified additional data that they wish had been included in the assessment results. As shown in Table 4, participants sought additional information about classroom interventions and from outside mental health service providers.

Table 4
Most Frequently Reported Data Participants Wished Had Been Included by Level of Ambiguity

Data	Low ambiguity					
	Not ED		ED		High ambiguity	
	n	%	n	%	n	%
Additional data regarding behavioral interventions, including positive supports, functional behavior analysis, and fidelity	18	30%	8	13%	13	22%
Scaled for Assessing Emotional Disturbance	3	5%	0	0%	0	0%
Emotional Disturbance Decision Tree	0	0%	2	3%	0	0%
Data relating to attention deficit hyperactivity disorder	2	3%	4	7%	2	3%
Additional information from current and prior teachers (e.g., rating scales, interviews)	9	15%	3	5%	5	8%
More direct observations in different settings and times	4	7%	2	3%	0	0%
Class wide intervention data	3	5%	0	0%	0	0%
Data relating to home-school collaboration	3	5%	0	0%	2	3%
Data relating to outside mental health diagnoses and treatment	4	7%	9	15%	8	13%

Question 11. *What data do school psychologists wish were included when data have low ambiguity and do not meet criteria for ED?*

In response to this optional question, 73% of participants described additional data that they wish had been included in the assessment results. Almost 30% of participants requested more data about classroom behavior interventions, including implementation of positive supports and functional behavior assessment, fidelity, and progress monitoring. Another common theme was for more information from current and

prior teachers (e.g., BASC results, interviews). Three participants requested data from the Scales for Assessing Emotional Disturbance. Although the vignette indicated that the student had never been evaluated or treated by mental health professionals, five participants wanted to confer with such providers to discuss mental health diagnoses and treatment.

Question 12. *What data do school psychologists wish were included when data have low ambiguity and meet criteria for ED?*

In response to this optional question, 45% of participants identified additional data that they wish had been included in the assessment results. The most frequent request was for data relating to outside mental health diagnoses and services, such as updated psychiatric evaluations and therapy progress. The second most common type of data sought was additional information about interventions implemented at school and the student's responses to them. Two participants indicated they would have used the Emotional Disturbance Decision Tree to aid their determination.

Chapter 5 Discussion

Summary of Findings

The primary purpose of this study was to determine how much, if at all, student race (Black or White) impacted school psychologists' decisions in ED evaluations when the evaluation results (a) had low ambiguity and did not meet ED criteria, (b) had low ambiguity and met ED criteria; and (c) were highly ambiguous. I found that racial bias *did not* impact school psychologists' eligibility determinations for ED under any of the three conditions. These findings did not support the racial bias theory of disproportionality. One secondary purpose was to elucidate how school psychologists' confidence in their decisions related to data ambiguity and student race. I found that neither ambiguity level nor student race had a significant impact on confidence ratings, and that the majority of participants were at least moderately confident in their decisions. An additional goal was to examine which data school psychologists considered most diagnostic in order to ascertain insight into their cognitive processes. Results indicated that school psychologists weighed data differently across the three conditions, suggesting they do not prioritize one data source consistently over others. Finally, participants were given the option of describing data they wished had been included in the assessment results but were not. Results showed school psychologists wanted more data relating to intervention implementation and consultation with mental health providers.

Racial Bias

Prior scholars have conducted descriptive research studies to elucidate factors associated with racial disproportionality in special education (e.g., Artiles et al., 2010; Waitoller et al., 2010). I built upon this foundational research by experimentally testing one of these factors — racial bias — with a key member of METs — school psychologists — for a controversial disability category — ED. I found that racial bias *did not* impact school psychologists' eligibility determinations for ED. Additionally, school psychologists' confidence in their decisions was not significantly related to student race.

The context for interpreting these results is somewhat obscure because prior research on bias in special education evaluations occurred 20-30 years ago, before the reauthorization of IDEA and popularity of response to intervention (RTI) and positive behavioral interventions and supports (PBIS). Nonetheless, this study's results are consistent with results from prior experimental studies demonstrating that school psychologists were not racially biased in evaluations for LD (Ysseldyke et al., 1981), intellectual disabilities (Huebner & Cummings, 1985), ADHD, and autism (Cuccaro et al., 1996). Similar to these prior researchers, I prioritized internal validity over external validity by using a vignette methodology and isolating race as the independent variable. This approach to investigating a complex phenomenon like disproportionality, which descriptive research suggests involves individual- and school-level variables (e.g., Sullivan & Bal, 2013), requires incremental manipulation of variables of interest (e.g., SES, geographical region, variations in state ED criteria, MET members, group decision making processes) and numerous replications. Future research systematically investigating these variables is needed before one makes broad conclusions about causes of ED disproportionality.

Racial bias may not have manifested in this study because the stimulus (i.e., written vignettes) may not have evoked the implicit biases that may arise when the MET members interact personally and have a relationship history with the student. Sensory information that comes from dialect, gestures, hairstyle, and clothing, among other things, may be triggers for stereotype biases in evaluations. Written vignette methodology may detect overt cases of racial bias better than implicit biases deriving from relationship and sensory contexts. Future analogue research using different media forms (e.g., videos) should be conducted to see if these results are replicated.

It is also possible that I did not find racial bias because the participants were not racially biased. School psychologists should have received graduate training on using nondiscriminatory practices to evaluate students for special education and may have learned about theories like aversive racism (Whaley, 1998) or cultural reproduction

theory (Skiba et al., 2006) which posits that “racial and class equity are reproduced over time through institutional and individual actions and decisions that maintain the status quo at the expense of less privileged groups” (p. 1426). Such processes are believed to operate often at implicit or habitual levels in decision making inherent in socially constructed disabilities like ED, but one can mitigate them by reflecting on their personal values and biases and engaging in systematic assessment practices (McConaughy & Ritter, 2002; Miranda, 2008).

Although this study makes an important contribution to the research base, it is not sufficient evidence upon which to make broad conclusions about whether the racial bias theory (e.g., Artiles et al., 2010) or increased risk theory (e.g., Donovan & Cross, 2002) better explain disproportionality. The study did not test an important alternative, potential source of racial bias: teacher referrals (e.g., Skiba et al., 2008). Indeed, meta-analytic findings suggest that teachers tend to refer more Hispanic and Black students than White students for evaluation (Hosp & Reschly, 2003). Research from last century shows that teachers tend to refer minority students more often for behavior rather than academic problems (Gottlieb, Gottlieb, & Trongone, 1991). More recent research shows that special educators consider aggression, classroom disruption, and mental health diagnoses of bipolar disorder and schizophrenia to be the top indicators in evaluating for ED (Becker et al., 2014), all of which are vulnerable to cultural and racial biases (e.g., Whaley, 1998). Similarly, one recent study found that teachers refer for evaluation when behavior problems include subject constructs like defiance, inappropriate physical behavior, aggression, and social/relationship difficulties (Briesch, Ferguson, Volpe, & Briesch, 2012).

Teacher referrals are also an important alternative source of potential bias because evidence suggests that they may trigger confirmation biases – i.e., evaluators seek data to confirm the referral rather than testing hypotheses objectively and systematically (e.g., Algozzine & Ysseldyke, 1981; O’Reilly et al., 1989; Ysseldyke et al., 1981). Indeed, the majority of students teachers refer for evaluation are qualified by the MET for special

education (Ysseldyke & Algozzine, 1981; Ysseldyke, Vanderwood, & Shriner, 1997). One may interpret this trend as evidence that teachers, who tend to know the student's school performance best, understand the qualification requirements and refer students for evaluation that are likely to meet these requirements (e.g., see MacMillan, Gresham, Lopez, & Bocian, 1996 for support of "teachers as tests" in special education identification). Given the highly ambiguous eligibility criteria for categories like ED (Allen & Hanchon, 2013), this presumption of teachers as effective screeners for ED warrants further scrutiny. As such, it is premature to rule out the racial bias theory of disproportionality before investigating whether teachers are racially biased their referrals for ED evaluation.

Data Ambiguity

One unique contribution of the study was the manipulation of data ambiguity; a variable that has long needed to be addressed in experimental studies of special education evaluation decision making (Huebner, 1991). The two low-ambiguity conditions (i.e., qualifies and does not qualify) evaluated schools psychologists' knowledge, skills, and abilities for conducting ED evaluations under ideal conditions – i.e., when the data across multiple measures and multiple respondents converge consistently. Results showed 100% agreement among school psychologists when the low-ambiguity data did not meet ED criteria and 87% agreement when data met ED. Of the eight participants who did not qualify the student as ED when the data were intended to meet ED criteria, five of them also did not qualify the student under the highly ambiguous condition. In other words, about 8% of the sample did not qualify any student as ED under any condition, raising questions about what kind of evidence such participants would need to qualify a student as ED, if ever they would. More than 15 years ago, Kauffman (1999) discussed numerous factors that likely interfere with identification of ED (e.g., overriding concern for stigma, preferring false negatives to false positives, misconstruing the least restrictive environment, calling special educating ineffective, etc.). This study may have unveiled such resistance to identifying ED, which is consistent with epidemiological findings that

ED is under-identified in the population (Brauner & Stephens, 2006; Merikangas et al., 2010). Future research exploring further the evidentiary thresholds for qualifying students as ED is warranted to ensure that students have meaningful access to services under the ED category and that such access is equitable across legally protected classes (e.g., sex, race, religion, etc.).

Another important finding is that when the data were highly ambiguous, the likelihood of the student qualifying as ED was nearly that of a coin toss (45% versus 55%). These results bring to mind the reasoning of U.S. Supreme Court Justice Stewart, who candidly defined “hard-core pornography” as “I know it when I see it” (384 U.S. 184, 197, 1964). Such subjectivity in identifying ED may be attributable to school psychologists’ diverging interpretations of ambiguous key terms in the federal definitions of ED. When the data were unambiguous, nearly all of the school psychologists saw the same thing. But when the assessment results did not converge neatly, the precise meaning school psychologists attributed to the legal criteria for ED became more important and possibly less consistent. These results make sense in light of the “fuzzy, unbounded problem space” (Dumont, 1993, p. 197) that constitutes ED eligibility criteria. Exacerbating this ambiguity is the fact that latent constructs at issue in psychoeducational evaluations are accurate to the extent that other psychologists agree or disagree with the diagnosis, which amounts to validating an inference with another inference as opposed to a tangible referent (e.g., Neighbors et al., 2003). Further, practitioners often do not receive feedback about the accuracy of their classification decision, which makes learning from decision making difficult (Garb, 1996). Under conditions of such high uncertainty, it is not surprising that the complex process of decision making has resulted in arbitrary outcomes in this study.

Notably, there was no statistically significant evidence that school psychologists’ decisions were impacted by racial bias. Instead, all students — regardless of race — had a coin toss chance to qualify as ED. The arbitrariness of these outcomes contrasts starkly with the high-stakes nature of special education eligibility determinations and warrants

future research to determine the pervasiveness of this problem (i.e., how often do initial evaluations for special education yield ambiguous results) and how increasing the objectivity of the ED criteria may impact racial disproportionality and the overall prevalence of ED. It is possible that the arbitrariness of the outcomes was due, at least in part, to the fact that the decision making occurred individually rather than as an evaluation team. There is some evidence that team decision making for evaluations produces less variance in outcomes than evaluations conducted by individuals (Pfeiffer, 1982); future research is needed to ascertain the impact of decision making context.

Confidence

Participants were asked to rate their confidence for each evaluation decision as low, moderate, or high. Across vignettes, ratings were skewed heavily towards moderate to high confidence. It is possible that the participants were very confident in their decisions because there were no other MET members participating in the evaluation to challenge their reasoning or offer alternative hypotheses. They did not experience the complex group dynamics involved in building consensus and managing dissension (e.g., Gutkin & Nemeth, 1997).

Results also showed that on average, participants were more confident ruling out ED than qualifying the student as ED under the low-ambiguity condition. This slight difference may reflect a general reluctance or discomfort for qualify students as ED, which is consistent with the persistently low national ED rate of about 1% (Kauffman, 1999). Additionally, participants were more confident in both of their decisions under the low-ambiguity condition than their single decision under the high-ambiguity condition. Interestingly, about 88% of school psychologists had at least moderate confidence in their decisions even when assessment results were highly ambiguous. Presumably, participants would not have felt confident determining ED eligibility if they had appreciated and considered carefully more than one reasonable interpretation of the ED criteria and evaluation results. High confidence under conditions of high uncertainty suggests participants may have utilized a cognitive heuristic to simplify the complex task of

applying ambiguous criteria to ambiguous evaluation results. One such heuristic that prior research shows has influenced special education evaluations is the confirmation bias (e.g., Ysseldyke et al., 1981; Ysseldyke & Algozzine, 1982). Participants may not have considered alternative interpretations because they searched for and selectively paid more attention to data that supported their initial impressions about eligibility and disregarded diverging interpretations (O'Reilly et al., 1989). Additional research utilizing qualitative methods like think-aloud would shed light on what cognitive processes enabled participants to feel confident about decisions for which there were no apparently correct answers.

Data Diagnosticity

Cognitive psychologists have explored how people consider various features of an object when classifying them into defined categories (e.g., Sloman et al., 1998). One measure of feature centrality is diagnosticity, “the probability than an object belongs to a certain category given that it has a certain feature; e.g., the probability that a patient has major depressive episode given that the patient has feelings of worthlessness” (Kim & Ahn, 2002, 472). In prior studies, school psychologists and special educators have ranked the types of data they generally consider most influential in determining eligibility (e.g., Allen & Hanchon, 2013; Knoff, 1983b). Such findings provide some insight about which factors school psychologists considered to be the “differentiating causal agent, the thing which is true of [the client] and not of the others who have remained ‘healthy’” (Meehl, 1973, p. 248). They also shed light on how school psychologists prioritize relatively objective measures (e.g., norm-referenced scores) and subjective measures (e.g., interviews, observations). The prior studies, however, presumed some constancy in data diagnosticity across evaluations by asking about practitioner’s *general* practices. Here, I explored how much the value of data was dependent upon the unique constellation of evaluation results and presenting problems. Participants were asked to identify which top three sources of information were most important in their decision making for each eligibility determination.

Across vignettes, many school psychologists (72-75%) ranked BASC scores as important in ED evaluations, which is consistent with prior findings (Allen & Hanchon, 2013). In light of disproportionality trends, school psychologists should consider whether the BASC and other rating scales are nondiscriminatory measures in terms of race by examining the sufficiency of racial minorities in the norm-reference groups. Some reliance on norm-referenced rating scales is appropriate to help evaluators discern how discrepant the behaviors and emotions at issue are from typically developing children. METs should not, however, use BASC scores “as the sole criterion for determining whether a child” meets the disability criteria (20 U.S.C. § 1414(b)(2), 2015).

Another trend consistent across vignettes was the low diagnosticity of cognitive and academic information. These ratings are interesting in light of the legal requirement that ED adversely impact the student’s educational performance. It appears as though participants did not interpret “educational performance” to mean only academic skills; rather, they interpreted the term broadly to include social and functional domains consistent with IDEA’s express purposes of ensuring “equality of opportunity, full participation, independent living, and economic self-sufficiency for individuals with disabilities” (20 U.S.C. § 1400(c)(2), 2015). Alternatively, school psychologists may have a narrow interpretation of educational performance, but they emphasize the mental health component more than the educational component in determining ED eligibility. As prior researchers have noted (Becker et al., 2014), more research is needed to understand how practitioners integrate the educational and mental health components in evaluations.

The diagnosticity of the more qualitative data tended to shift across vignettes. School psychologists rated the student interview and family background as highly diagnostic in the highly ambiguous condition when the student’s problems seemed to arise from changes within the family, and less diagnostic when the vignette suggested the student’s problems could be addressed by different approaches to classroom management. These results suggest that the participants considered the relevancy of school and family factors on a case-by-case basis.

In sum, school psychologists do not seem to be engaging in formulaic or unidimensional decision making when they evaluate for ED. These results suggest that considerable professional judgment is used to navigate the ED criteria. Additional research is needed to understand better how school psychologists reconcile data that do not converge well, especially because the lack of standardization in weighing the diagnosticity of data may explain trends in outcomes like disproportionality.

Additional Data

Participants had the option of identifying what additional data they wished had been included in the assessment results for each vignette. There was little consensus among participants about data that should have been included but was omitted, suggesting that the multimethod, multisource evaluation data provided were sufficient for determining ED. The most frequently sought additional data were RTI, PBIS, and fidelity data, which highlights school psychologists' awareness of multitier systems of support for social, emotional, and behavior problems. They also suggest that the participants were familiar with best practices in ED evaluations (McConaughy & Ritter, 2002) because they sought information about student strengths and the ecological aspects of student behavior. Additionally, 13-15% of participants wanted consultation with outside mental health providers in the low-ambiguity-qualifies condition and highly ambiguous condition. These results may be an underestimate of school psychologists' motivation to consult with outside providers because two vignettes (i.e., low-ambiguity-does-not qualify condition and highly ambiguous condition) indicated that there were no such providers with whom they could consult. These results are consistent with prior findings showing that special educators consider diagnoses from mental health providers as highly relevant in ED evaluations (Becker et al., 2014). Nonetheless, future research should help clarify how and when school psychologists engage in consultation with outside mental health providers during ED evaluations, and the demographic characteristics of students who receive such services. This information may help inform disproportionality trends.

Merits and Limitations

There are numerous merits and limitations that need to be considered when interpreting these results. One great strength of the study is its internal validity. The experimental design isolated racial bias as a causal mechanism in determining eligibility for ED. I am not aware of any other study within the past 30 years experimentally testing the impact of race on school psychologists' evaluations of ED. Thus, the results make an important contribution to the study of disproportionality and decision making in school psychology.

Another strength of the study is that it examined how data ambiguity relates to decision making in ED evaluations. Prior research has suggested that decisions are more vulnerable to biases when decision making conditions are uncertain and/or complex (e.g., Kahneman & Tversky, 1977), which can be the case in evaluations for ED given the ambiguous material terms in the legal criteria and requirement for multimethod and multisource data collection. Accordingly, this study differentiated between low- and high-ambiguity assessment results. The low-ambiguity condition evaluated participants' general competency in ED evaluations so that differences, if any, between the groups in the high-ambiguity condition could be attributed to racial bias. It also provided useful descriptive data on how well practitioners understand and apply the ED criteria under ideal conditions – i.e., unambiguous data. A final strength of this study is that the vignettes followed best practices for evaluations by including information from record review, interview of family, student, and teachers, as well as results from rating scales, direct observation, and standardized achievement and cognitive tests.

An important drawback to this study is that the strong internal validity came at the expense of external validity. The vignette methodology differs from real practice in many ways. For instance, federal law requires that METs – not school psychologists by themselves – determine whether students qualify as ED, yet the social and cognitive dynamics of group decision making were not addressed in this study. Additionally, in practice the evaluation process can be iterative, such that data analysis begets questions to explore with further data collection –i.e., the parent interview raises issues to explore in

teacher interview and vice versa. Here, the assessment results were static and presented all at once; participants were not allowed to follow up on unanswered questions. Indeed, 45-73% of participants described additional data they wish had been included in the results. This limitation is particularly relevant for the vignette where the data were intended to be ambiguous because it is possible that practitioners in real life would not have made a determination without collecting further data. Because the vignette methodology did not simulate an interactive data collection process, the outcomes may not represent decisions that would have been made in practice.

Another potential limitation is that the sample size may have been inadequate to demonstrate the discriminatory effect under investigation. A precise power calculation could not be conducted because there was insufficient extant information about the population (e.g., number of school psychologists conducting evaluations for ED nationally, variance, standard deviation) and the impact, if any, of racial discrimination on ED evaluation outcomes (e.g., effect size). Although the total sample size of 60 participants was justifiable in the field of experimental social science research (e.g., Hill, 1998) and the central limit theorem (e.g., Howell, 2010), additional research with larger samples is needed to understand better the impact of power on effect size.

With regards to the sampling frame, there are four limitations. First, the sampling frame excluded from the pool of potential participants individuals who were unlikely to be involved in public school evaluations for ED (e.g., academic faculty, Pearson employees). Because it is unknown how involved this subpopulation of school psychologists is in ED evaluations, the impact of excluding them from the sample is also unknown. Second, the sampling frame excluded an estimated 25-33% school psychologists who were not members of NASP (Hosp & Reschly, 2002, 2003). Although the possibility exists that school psychologists who are not members of NASP differ significantly from NASP members, prior research using similar methodology has yielded representative findings (e.g., Hosp & Reschly, 2002; Hutton & Dubes, 1992). Third, the sampling frame excluded school psychologists who are members of NASP but did not

provide in the membership directory an e-mail address or sufficient employment information so that an e-mail address could be located via Internet searches. Fourth, the sampling frame excluded school psychologists who provided information to the membership directory but they did not have access to the Internet and thus, could not receive the invitation to participate.

Another important limitation of the study was that only 26% of individuals invited to participate actually did so. Potential differences between the nonresponders and participants may distort the results such that they are not representative of the intended population – i.e., school psychologists practicing in Pennsylvania. Likewise, two individuals started the study but they did not complete it. Although the number of break-offs was low, their impact the generalizability of the results is also worth considering.

The representativeness of the sample that did complete the study may also be a limitation. The sample was primarily White, female, and working in suburban districts not identified as having problems with disproportionality. According to relatively recent study on the demographics of school psychology practitioners (Curtis, Costello, & Gelley, 2012), the field is predominantly female (77%), White (91%), and working in public schools (84%) that are suburban (43%). These statistics suggest that the sample was roughly comparable to the national population of school psychologists. However, it is difficult to evaluate how comparable this sample is to the population of school psychologists who work in districts where racial disproportionality manifests because there is a lack of data on this subpopulation of professionals.

The use of incentives provides another potential source of error in the study. It is possible that participants rushed through the vignettes without performing the desired analysis in order to receive the incentive, resulting in data that do not reflect practices in the field. The duration of how long the participants had the study module open was recorded. Values ranged from about 6 to 61 minutes, with an average of 23 minutes and mode of 15 minutes. About 25% of the participants completed the survey in less than 15

minutes, which raises concerns about carefully they deliberated their evaluation decisions. The impact, if any, of duration on the quality of the results is unknown.

Additionally, the materials used in the study raise some concerns about the generalizability of the results. Students with ED are a highly heterogeneous group (Wagner et al., 2005), yet participants only completed three hypothetical evaluations, raising questions about whether the study had adequate stimuli sampling (Wells & Windschitl, 1999) to elicit the patterns of disproportionality that have occurred at the national level for more than 30 years (Waitoller et al., 2010). Likewise, it is possible that written vignette methodology is not a sufficiently potent forum for triggering racial bias in school psychologists, although prior researchers have found evidence of racial discrimination in experimental studies using written materials and only manipulating the name of hypothetical individuals (e.g., Beltrand, M., & Mullainathan, 2003; Milkman et al., 2012).

For all of these reasons, the results of this study make meaningful progress in evaluating racial bias as a causal mechanism contributing to disproportionality in special education, but they cannot by themselves be determinative of the issue because of their numerous limitations. Additional experimental research is needed to replicate or refute these findings.

Implications for Practice

By enacting ambiguous ED laws, the federal government has shifted to local education authorities, district administrators, and individual MET members the responsibility for giving meaning to the ED eligibility criteria. This delegation of authority presents an opportunity and a challenge for school practitioners. The opportunity is to tailor the precise meaning of vague terms like “over a long period of time” (34 C.F.R. § 300.8(c)(4), 2015) to the culture and resources of the local region. The challenge is to ensure consistency across individual cases so that ambiguous terms have the same meaning for students regardless of their legally protected statuses – e.g., race, gender. To prevent disparate treatment, school psychologists and other regular members

of METS (e.g., special education teachers, administrators) should reflect regularly and as a group about how they are interpreting ambiguous terminology. Such on-going collaboration among school professionals may help mitigate cognitive biases that frequently distort complex decision making and facilitate equitable access to the services and supports intended by IDEA.

In addition to formulating the meaning of ED collectively, school professionals may find it useful to approach ED evaluations from a biopsychosocial perspective. The biopsychosocial model of development reflects many of the principles of developmental systems theory and ecological theory (e.g., Burns, 2011; Christenson, 2004). Dodge and Pettit (2003) posited a biopsychosocial model of the development of conduct problems that provides a useful framework for understanding the etiology of social, emotional, and behavioral problems involved in ED evaluations. Based upon their review of the vast empirical and theoretical literature investigating the development of conduct disorder, they asserted that conduct problems result from interactions among multiple factors, including biological dispositions, sociocultural contexts, life experiences, and “the child’s acquired pattern of processing social information” (Dodge & Pettit, 2003, p. 350; see also Raine, 2002). To illustrate this multi-level complexity, they compared conduct disorder to heart disease: both conditions are “vaguely defined constructs with questionable validity, but both also have clearly identifiable referents, such as a myocardial infarction or an act of homicide” and “a single causal agent will never be found” for either condition (Dodge & Pettit, 2003, p. 350). Perhaps they likened conduct disorder to heart disease because heart disease does not have the social stigma of mental illness (for review of stigma of childhood mental disorders, see Mukolo, Heflinger, & Wallston, 2010) and people may be more inclined to accept the complexity of its development rather than blame the patient for causing its manifestation. Indeed, the biopsychosocial model of development implicates factors beyond an individual’s control (e.g., genetics, prenatal environments) in the manifestation of social, emotional, or behavioral problems, hindering a simple allocation of blame (i.e., causation). Instead, correlational research suggests that multiple,

diverse, distal risk factors and proximal causes create “unique developmental pathways” from which conditions like heart disease or antisocial conduct emerge (Dodge & Pettit, 2003, p. 350).

These unique developmental pathways also may be conceptualized as developmental cascades, which are the “cumulative consequences for development of the many interactions and transactions occurring in developing systems that result in spreading effects across levels, among domains at the same level, and across different systems or generations” (Masten & Cicchetti, 2010). In other words, social, emotional, or behavioral problems in adolescence or adulthood may be the cumulative effect of interactions among multiple levels (molecular, person, family, school, neighborhoods, society) in multiple domains (e.g., genetic, biological, cognitive, social, interpersonal) across development (e.g. prenatal, early childhood, transitioning to school, puberty). Because of this complexity, the scientific method does not yet allow for causal inferences about students’ educational performance.

Importantly, one should not interpret this research as evidence that some children are destined to develop ED; rather, the evidence is probabilistic and describes risks. Some students with high cumulative risk for developing social, emotional, or behavioral problems may be resilient and not have problems whereas students with fewer cumulative risks manifest severe problems. The aforementioned correlational research is useful in ED evaluations to rebut common assumptions that the cause of social, emotional, or behavioral problems is identifiable through special education evaluations. METs also should use information about risk factors and cascade effects to conduct comprehensive evaluations that include developmental histories and interpret assessment results ethically.

References

- Ahn, W., & Kim, N. S. (2001). The causal status effect in categorization: An overview. *The Psychology of Learning and Motivation, 40*, 23-65.
- Algozzine, B., & Ysseldyke, J. E. (1981). Decision makers' prediction of students' academic difficulties as a function of referral information. *The Journal of Educational Research, 73*, 145–150.
- Allen, R. A., & Hanchon, T. A. (2013). What can we learn from school-based emotional disturbance assessment practices? Implications for practice and preparation in school psychology. *Psychology in the Schools, 50*, 290-299. doi: 10.1002/pits.21671
- Artiles, A. J., Kozleski, E. B., Trent, S. C., Osher, D., & Ortiz, A. (2010). Justifying and explaining disproportionality, 1968-2008: A critique of underlying views of culture. *Exceptional Children, 76*, 279-299.
- Artiles, A. J., Rueda, R., Salazar, J. J., & Higareda, I. (2005). Within-group diversity in minority disproportionate representation: English language learners in urban school districts. *Exceptional Children, 71*, 283-300.
- Barnett, D. (1988). Professional judgment: A critical appraisal. *School Psychology Review, 17*, 658–672.
- Becker, S. P., Paternite, C. E., & Evans, S. W. (2014). Special educators' conceptualizations of emotional disturbance and educational placement decision making for middle and high school students. *School Mental Health, 6*, 163-174. doi: 10.1007/s12310-014-9119-7
- Beltrand, M., & Mullainathan, S. (2003). *Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination*. (No. w9873). National Bureau of Economic Research.
- Blackorby, J., & Wagner, M. (1996). Longitudinal postschool outcomes of youth with disabilities: Findings from the National Longitudinal Transition Study. *Exceptional Children, 62*, 399- 413.

- Bollmer, J., Bethel, J., Garrison-Mogren, R., & Brauen, M. (2007). Using the risk ratio to assess racial/ethnic disproportionality in special education at the school-district level. *The Journal of Special Education, 41*, 186-198.
- Bradley, R., Doolittle, J., & Bartolotta, R. (2008). Building on the data and adding to the discussion: The experiences and outcomes of students with emotional disturbance. *Journal of Behavioral Education, 17*, 4-23. doi: 10.1007/s10864-007-9058-6
- Brannon, L. A., & Carson, K. L. (2003). The representativeness heuristic: Influence on nurses' decision making. *Applied Nursing Research, 16*, 201-204. doi: 10.1016/S0897-1897(03)00043-0
- Brauner, C. B., & Stephens, C. B. (2006). Estimating the prevalence of early childhood serious emotional/behavioral disorders: Challenges and recommendations. *Public Health Reports, 121*, 303-310.
- Briesch, A. M., Ferguson, T. D., Volpe, R. J., & Briesch, J. M. (2012). Examining teacher's perceptions of social-emotional and behavioral referral concerns. *Remedial and Special Education, 34*, 249-256. doi: 10.1177/0741932512464579
- Burns, C. W. (1990). Judgment theory and school psychology. *Journal of School Psychology, 28*, 343-349. doi:10.1016/0022-4405(90)90023-Z
- Burns, C. W. (1992). Psychoeducational decision making, test scores, and descriptive data: Selected methodological issues. *Journal of School Psychology, 30*, 1-16. doi:10.1016/0022-4405(92)90016-X
- Burns, M. K. (2011). School psychology research: Combining ecological theory and prevention science. *School Psychology Review, 40*, 132-139.
- Capistrano Unified Sch. Dist. v. Wartenberg*, 59 F.3d 884 (9th Cir. 1995).
- Chavez, L. M., Shrout, P. E., Alegria, M., Lapatin, S., & Canino, G. (2010). Ethnic differences in perceived impairment and need for care. *Journal of Abnormal Child Psychology, 38*, 1165-1177. doi: 10.1007/s10802-010-9428-8
- Christenson, S. L. (2004). The family-school partnership: An opportunity to promote the learning competence of all students. *School Psychology Review, 33*, 83-104.

- Cuccaro, M. L., Wright, H. H., Rownd, C. V., Abramson, R. K., Waller, J., & Fender, D. (1996). Professional perceptions of children with developmental difficulties: The influence of race and socioeconomic status. *Journal of Autism and Developmental Disorders*, *26*, 461–469. doi:10.1007/BF02172830
- Curtis, J. M., Costello, M. J., & Gelley, C. (2012). School psychology 2010: Demographics, employment, and the context for professional practice – Part 1. *NASP Communiqué*, *40*(7), xx-xx.
- Davidow, J., & Levinson, E. M. (1993). Heuristic principles and cognitive bias in decision making: Implications for assessment in school psychology. *Psychology in the Schools*, *30*, 351–361. doi:10.1002/1520-6807(199310)30:4<351::AID-PITS2310300410>3.0.CO;2-X
- Dodge, K. A., & Pettit, G. S. (2003). A biopsychosocial model of the development of chronic conduct problems in adolescence. *Developmental Psychology*, *39*, 349-371. doi: 10.1037/0012-1649.39.2.349
- Donovan, S., & Cross, C. T. (2002). *Minority students in special and gifted education*. Washington, D.C.: National Academy Press. Retrieved from <http://site.ebrary.com/id/10032383>
- Dumont, F. (1993). Inferential heuristics in clinical problem formulation: Selective review of their strengths and weaknesses. *Professional Psychology: Research and Practice*, *24*, 196–205. doi:10.1037//0735-7028.24.2.196
- Dunn, L. M. (1968). Special education for the mildly retarded: Is much of it justifiable? *Exceptional Children*, *35*, 5–22.
- Eack, S. M., Bahorik, A. L., & Newhill, C. E. (2012). Interviewer-perceived honesty as a mediator of racial disparities in the diagnosis of schizophrenia. *Psychiatric Services*, *63*, 875- 880. doi:
- Elliot, A. C., & Woodward, W. A. (2007). Analysis of Categorical Data. In A. C. Elliot & W. A. Woodward (Eds.), *Statistical Analysis Quick Reference Guide*, (pp. 113-152). Thousand Oaks: SAGE Publications, Inc.

- Fagley, N. S. (1988). Judgmental heuristics: Implications for the decision making of school psychologists. *School Psychology Review, 17*, 311–321.
- Fan, W., & Yan, Z. (2010). Factors affecting response rates of web survey: A systematic review. *Computers in Human Behavior, 26*, 132-139. doi: 10.1016/j.chb.2009.10.015
- Galotti, K. M. (2007). Decision structuring in important real-life choices. *Psychological Science, 18*, 320–325. doi:10.1111/j.1467-9280.2007.01898.x
- Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1*, 67–89. doi:10.1146/annurev.clinpsy.1.102803.143810
- Gifford-Smith, M., Dodge, K. A., Dishion, T. J., & McCord, J. (2005). Peer influence in children and adolescents: Crossing the bridge from developmental to intervention science. *Journal of Abnormal Child Psychology, 33*, 255-256. doi: 10.1007/s10802-005-3563-7
- Gilliam, J. E. (1979). Contributions and status rankings of educational planning committee participants. *Exceptional Children, 45*, 466-468.
- Gnys, J. A., Willis, W. G., & Faust, D. (1995). School psychologists' diagnoses of learning disabilities: A study of illusory correlation. *Journal of School Psychology, 33*, 59–73. doi:10.1016/0022-4405(94)00023-2
- Görizt, A. S. (2010). Using lotteries, loyalty points, and other incentives to increase participant response and completion. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced Methods for Behavioral Research on the Internet* (pp. 219-233). Washington, D.C.: American Psychological Association.
- Gottlieb, J., Gottlieb, B. W., & Trongone, S. (1991). Parent and teacher referrals for a psychoeducational evaluation. *The Journal of Special Education, 25*, 155-167.
- Gutkin, T. B., & Nemeth, C. (1997). Selected factors impacting decision making in prereferral intervention and other school-based teams: Exploring the intersection between school and social psychology. *Journal of School Psychology, 35*, 195–216. doi:10.1016/S0022-4405(97)00005-8

- Heller, K. A., Holtzman, W., & Messick, S. (Eds.) (1982). *Placing Children in Special Education: A Strategy for Equity*. Washington, D. C.: National Academy Press.
- Heverly, M. A., Fitt, D. X., & Newman, F. L. (1984). Constructing case vignettes for evaluating clinical judgment: An empirical model. *Evaluation and Program Planning*, 7, 45-55 (1984).
- Hill, R. (1998). What sample size is “enough” in internet survey research? *Interpersonal Computing and Technology: An Electronic Journal for the 21st Century*, 6 (3-4), 1-10.
- Hosp, J. L., & Reschly, D. J. (2002). Regional differences in school psychology practice. *School Psychology Review*, 31, 11-29.
- Hosp, J. L., & Reschly, D. J. (2003). Referral rates for intervention or assessment: A meta-analysis of racial differences. *The Journal of Special Education*, 37, 67-80.
- Howell, D. C. (2010). *Statistical Methods for Psychology* (ed. 7). Belmont, CA: Duxbury Press.
- Huebner, E. S. (1990). The generalizability of the confirmation bias among school psychologists. *School Psychology International*, 11, 281–286.
doi:10.1177/0143034390114005
- Huebner, E. S. (1991). Bias in special education decisions: The contribution of analogue research. *School Psychology Quarterly*, 6, 50–65. doi:10.1037/h0088240
- Huebner, E. S., & Cummings, J. A. (1985). The impact of sociocultural background and assessment data upon school psychologists’ decisions. *Journal of School Psychology*, 23, 157–166. doi:10.1016/0022-4405(85)90006-8
- Hutton, J. B., & Dubes, R. (1992). Assessment practices of school psychologists: Ten years later. *School Psychology Review*, 21, 271-284.
- Jacobellis v. Ohio*, 378 U.S. 184, 1964 (Steward, J., concurring)
- Kahneman, D., & Tversky, A. (1977). *Intuitive prediction: Biases and corrective procedures* (Technical Report No. PTR-1042-77-6) (p. 44). Defense Advanced Research Projects Agency.

- Katsiyannis, A., & Maag, J. W. (2001). Manifestation determination as a golden fleece. *Exceptional Children, 68*, 85-96.
- Kauffman, J. M. (1999). How we prevent the prevention of emotional and behavioral disorders. *Exceptional Children, 65*, 448-468.
- Kim, N. S., & Ahn, W. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology: General, 131*, 451-476. doi: 10.1037//0096-3445.131.4.451
- Klein, G. (2008). Naturalistic decision making. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*, 456-460. doi:10.1518/001872008X288385
- Knoff, H. M. (1983a). Investigating the disproportionate influence and status of multidisciplinary child study teams. *Exceptional Children, 49*, 367-370.
- Knoff, H. M. (1983b). Effect of diagnostic information on special education placement decisions. *Exceptional Children, 49*, 440-443..
- Lane, K. L., & Carter, E. W. (2006). Supporting transition-age youth with and at risk for emotional and behavioral disorders at the secondary level: A need for further inquiry. *Journal of Emotional and Behavioral Disorders, 14*, 66-70.
- López, S. R. (1989). Patient variable biases in clinical judgment: Conceptual overview and methodological considerations. *Psychological Bulletin, 106*, 184-203. doi:10.1037//0033-2909.106.2.184
- Lozar Manfreda, K., Bosnjak, M., Berzelak, J., Haas, I., & Vehovar, V. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. *International Journal of Market Research, 50*, 79-104.
- MacMillan, D. L., Gresham, F. M., Lopez, M. F., & Bocian, K. M. (1996). Comparison of students nominated for prereferral interventions by ethnicity and gender. *The Journal of Special Education, 30*, 133-151.

- Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology*, *22*, 491-495. doi: 10.1017/S0954579410000222
- May, K. (2009). By reason thereof: Causation and eligibility under the Individuals with Disabilities Education Act. 2009 *Brigham University Education and Law Journal* 173.
- McConaughy, S. H., & Ritter, D. R. (2002). Best practices in multidimensional assessment of emotional or behavioral disorders. In A. Thomas, & J. Grimes (Eds.), *Best Practices in School Psychology V* (pp.1303-1320). Bethesda, MD: NASP Publications.
- McDermott, P. A. (1981). Sources of error in the psychoeducational diagnosis of children. *Journal of School Psychology*, *19*, 31-44. doi:10.1016/0022-4405(81)90005-4
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- Meehl, P. E. (1973). Why I do not attend case conferences. In P. E. Meehl (Ed.) *Psychodiagnosis: Selected Papers*, (pp. 225-302). Minneapolis: University of Minnesota Press.
- Mehan, H. (1991). The schools' work of sorting students. In D. Zimmerman & D. Boden (Eds.), *Talk and social structure* (pp. 71-90). Cambridge, England: Polity Press.
- Mellard, D. F., Deshler, D. D., & Barth, A. (2004). LD identification: It's not simply a matter of building a better mousetrap. *Learning Disability Quarterly*, *27*, 229. doi:10.2307/1593675
- Merikangas, K. R., He, J.-P., Brody, D., Fisher, P. W., Bourdon, K., & Koretz, D. S. (2010). Prevalence and treatment of mental disorders among US children in the 2001-2004 NHANES. *Pediatrics*, *125*, 75-81. doi: 10.1542/peds.2008-2598
- Milkman, K. L., Akinola, M., & Chugh, D. (2012). Temporal distance and discrimination: An audit study in academia. *Psychological Science*, *23*, 710-717. doi: 10.1177/0956797611434539

- Miner, J. L., & Clarke-Stewart, A. (2008). Trajectories of externalizing behavior from age 2 to age 9: Relations with gender, temperament, ethnicity, parenting, and rater. *Developmental Psychology, 44*, 771-786. doi: 10.1037/0012-1649.44.3.771
- Miranda, A. H. (2008). Best practices in increasing cross-cultural competence. In A. Thomas, & J. Grimes (Eds.), *Best Practices in School Psychology V* (pp.1739-1349). Bethesda, MD: NASP Publications.
- Morgan, P. L., Frisco, M. L., Farkas, G., & Hibel, J. (2010). A propensity score matching analysis of the effects of special education services. *The Journal of Special Education, 43*, 236-254. doi: 10.1177/0022466908323007
- Mukolo, A., Heflinger, C. A., & Wallston, K. A. (2010). The stigma of childhood mental disorders: A conceptual framework. *Journal of the American Academy of Child & Adolescent Psychiatry, 49*, 92-103.
- Muller v. Committee on Special Educ.*, 145 F.3d 95 (2nd Cir. 1998).
- National Conference of Commissioners of Uniform State Laws. (1995). *Uniform Statute and Rule Construction Act*. Retrieved from <http://uniformlaws.org/Act.aspx?title=Statute%20and%20Rule%20Construction%20Act>
- Neighbors, H. W., Trierweiler, S. J., Ford, B. C., & Muroff, J. R. (2003). Racial differences in DSM diagnosis using a semi-structured instrument: The importance of clinical judgment in the diagnosis of African Americans. *Journal of Health and Social Behavior, 44*, 237. doi:10.2307/1519777
- Nguyen v. D.C.*, 681 F. Supp. 2d 49 (Dist. Col. 2010).
- O'Reilly, C., Northcraft, G. B., & Dabers, D. (1989). The confirmation bias in special education eligibility decisions. *School Psychology Review, 18*, 126-135.
- Olympia, D., Farley, M., Christiansen, E., Pettersson, H., Jenson, W., & Clark, E. (2004). Social maladjustment and students with behavioral and emotional disorders: Revisiting basic assumptions and assessment issues. *Psychology in the Schools, 41*, 835-847. doi:10.1002/pits.20040

- Pfeiffer, S. I. (1982). The superiority of team decision making. *Exceptional Children, 49*, 68-69.
- Porter, S. R., & Whitcomb, M. E. (2003). The impact of lottery incentives on student survey response rates. *Research in Higher Education, 44*, 389-407.
- Raine, A. (2002). Biological studies of antisocial and violent behavior in children and adults: A review. *Journal of Abnormal Child Psychology, 30*, 311-326.
- Redfield, S. E., & Kraft, T. (2012). What color is special education? *Journal of Law & Education, 41*, 129-200.
- Rehder, B. (2010). Causal-based categorization: A review. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 52, pp. 39-116). San Diego, CA: Academic Press.
- Rehder, B., & Kim, S. (2009). Classification as diagnostic reasoning. *Memory & Cognition, 37*, 715-729.
- Reid, D. K., & Knight, M. G. (2006). Disability justifies exclusion of minority students – A critical history grounded in disability studies. *Educational Researcher, 35*(6), 18-23.
- Reips, U. D. (2010). Design and formatting in Internet-based research. *Advanced Internet Methods in the Behavioral Sciences, 29-43*.
- Rothstein, R. (2004). *Class and schools: Using social, economic, and educational reform to close the black-white achievement gap*. New York: Teachers College Press.
- Schwartz, R. C., & Blankenship, D. M. (2014). Racial disparities in psychotic disorder diagnosis: A review of empirical literature. *World Journal of Psychiatry, 4*, 133-140. doi: 10.5498/wjp.v4.i4.133
- Skånér, Y, Bring, J., & Strender, L.-E. (2004). Selecting representative case vignettes for clinical judgment studies: Examples from two heart failure studies. *Quality & Quantity, 38*, 621-635.

- Skiba, R. J., Poloni-Staudinger, L., Simmons, A. B., Feggins-Azziz, L. R., & Chung, C.-G. (2005). Can poverty explain ethnic disproportionality in special education? *The Journal of Special Education, 39*, 130-144. doi:
- Skiba, R. J., Simmons, A., Ritter, S., Gibb, A. C., Rausch, M. K., Cuarado, J., & Chung, C.-G. (2008). Achieving equity in special education: History, status, and current challenges. *Exceptional Children, 74*, 264-288.
- Skiba, R. J., Simmons, A., Ritter, S., Kohler, K., Henderson, M., & Wu, T. (2006). The context of minority disproportionality: Practitioner perspectives on special education referral. *Teachers College Record, 108*, 1424-1459.
- Singer, J. D., Palfrey, J. S., Butler, J. A., & Walker, D. K. (1989). Variation in special education classification across school districts: How does where you live affect what you are labeled? *American Educational Research Journal, 26*, 261. doi:10.2307/1163033
- Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science, 22*, 189-228.
- Springer v. Fairfax County Schoolboard*, 134 F. 3d 659 (4th Cir. 1998).
- Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A. M. B., Nadal, K. L., & Esquilin, M. (2007). Racial microaggressions in everyday life: Implications for clinical practice. *American Psychologist, 62*, 271-286. doi:10.1037/0003-066X.62.4.271
- Sullivan, A. L. (2011). Disproportionality in special education identification. *Exceptional Children, 77*, 317-334.
- Sullivan, A. L., & Bal., A. (2013). Disproportionality in special education: Effects of individual and school variables on disability risk. *Exceptional Children, 79*, 475-494.
- Sullivan, A. L., & Field, S. (2013). Do preschool special education services make a difference in kindergarten reading and math skills?: A propensity score weighting

- analysis. *Journal of School Psychology*, 51, 243-260. doi: [10.1016/j.jsp.2012.12.004](https://doi.org/10.1016/j.jsp.2012.12.004)
- Tourangeau, R., Conrad, F. G., & Couper, M. P. (2013). *The Science of Web Surveys*. New York, NY: Oxford University Press.
- Trierweiler, S. J., Neighbors, H. W., Munday, C., Thompson, E. E., Binion, V. J., & Gomez, J. P. (2000). Clinician attributions associated with the diagnosis of schizophrenia in African American and non-African American patients. *Journal of Consulting and Clinical Psychology*, 68, 171-175. doi: 10.1037//0022-006x.68.1.171
- Triplet, R. G. (1992). Discriminatory biases in the perception of illness: The application of availability and representativeness heuristics to the AIDS crisis. *Basic and Applied Social Psychology*, 13, 302-322. doi: 10.1207/s15324834basp1303_3
- Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106, 1039-1061. doi: 10.2307/2937956
- Uniform Law Commission. (2013). *Statute and Rule Construction Act Summary*. Retrieved from <http://uniformlaws.org/ActSummary.aspx?title=Statute%20and%20Rule%20Construction%20Act>
- U.S. Department of Education, Office of Special Education Programs, Individuals with Disabilities Education Act (IDEA) database, retrieved June 10, 2013, from <http://tadnet.public.tadnet.org/pages/712>.
- van Ryn, M., & Fu, S. (2003). Paved with good intentions: Do public health and human service providers contribute to racial/ethnic disparities in health? *American Journal of Public Health*, 93, 248-255. doi: 10.2105/AJPH.93.2.248
- Wagner, M. (1995). Outcomes for youths with serious emotional disturbance in secondary school and early adulthood. *The Future of Children*, 5(2), 90-112. doi: 10.2307/1602359

- Wagner, M., Kutash, K., Duchnowski, A. J., Epstein, M. H., & Sumi, W. C. (2005). The children and youth we serve: A national picture of the characteristics of students with emotional disturbances receiving special education. *Journal of Emotional and Behavioral Disorders, 13*, 79-96. doi: 10.1177/10634266050130020201
- Waitoller, F. R., Artiles, A. J., & Cheney, D. A. (2010). The miner's canary: A review of overrepresentation research and explanations. *The Journal of Special Education, 44*, 29-49. doi: 10.1177/0022466908329226
- Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25*, 1115 doi: 10.1177/01461672992512005
- Whaley, A. L. (1998). Racism in the provision of mental health services: A social-cognitive analysis. *American Journal of Orthopsychiatry, 68*, 47-57. doi:10.1037/h0080269
- Wiley, A. L., Brigham, F. J., Kauffman, J. M., & Bogan, J. E. (2013). Disproportionate poverty, conservatism, and the disproportionate identification of minority students with emotional and behavioral disorders. *Education and Treatment of Children, 36*, 29-50.
- Yoshida, R. K., Fenton, K. S., Maxwell, J. P., & Kaufman, M. J. (1978). Group decision making in the planning team process: Myth or reality? *Journal of School Psychology, 16*, 237-244.
- Youngstrom, E. A. (2013). Future directions in psychological assessment: Combining evidence-based medicine innovations with psychology's historical strengths to enhance utility. *Journal of Clinical & Adolescent Psychology, 42*, 139-159. doi: 10.1080/15374416.2012.736358
- Ysseldyke, J.E., & Algozzine, B. (1981). Diagnostic classification decisions as a function of referral information. *Journal of Special Education, 15*, 429-435.

- Ysseldyke, J. E., & Algozzine, B. (1982). Bias among professionals who erroneously declare students eligible for special services. *The Journal of Experimental Education, 50*, 223–228.
- Ysseldyke, J. E., & Algozzine, B. (1983). On making psychoeducational decisions. *The Journal of Special Education, 15*, 429–435. doi:10.1177/002246698101500404
- Ysseldyke, J. E., Algozzine, B., Regan, R., & McGue, M. (1981). The influence of test scores and naturally-occurring pupil characteristics on psychoeducational decision making with children. *Journal of School Psychology, 19*, 167–177.
doi:10.1016/0022-4405(81)90058-3
- Ysseldyke, J. E., Algozzine, B., & Richey, L. (1982). Judgment under uncertainty: How many children are handicapped? *Exceptional Children, 48*, 531-534.
- Ysseldyke, J. E., Vanderwood, M. L., & Shriner, J. (1997). Changes over the past decade in special education referral to placement probability: An incredibly reliable practice. *Diagnostique, 23*, 193-201.

Appendix A – Email invitation

Dear [Name],

You are part of a select group of school psychologists invited to participate in a study investigating school psychologists' decision making in special education evaluations. For completing the study, you will receive a \$20 gift certificate via e-mail for iTunes, Amazon.com, or Starbucks, or you may use the gift certificate to donate \$20 to Habitat for Humanity. Failure to complete the study by May 30, 2014 will result in loss of the gift certificate.

Only school psychologists currently involved in special education evaluations may participate - i.e., assessors or supervisors of evaluations. This study requires you to spend approximately 15-20 minutes evaluating whether 3 students qualify for special education based on hypothetical student data.

Shanna Sadeh is conducting this research as part of her dissertation for the School Psychology Program at the University of Minnesota; her academic adviser, Dr. Amanda Sullivan, Ph.D., L.P., is supervising the study. Declining or accepting this invitation will not affect your relations with the University of Minnesota in any way.

Please click on the following link to access the study:

https://umn.qualtrics.com/SE/?SID=SV_4OBPYUWhR9zU401

Sincerely,

Shanna Sadeh, M.A.

Appendix B – Instrumentation

Consent

Eligible school psychologists who **complete** this study may choose to receive a \$20 gift certificate via e-mail. If this applies to you, which gift certificate would you like to receive?

- Amazon.com
- iTunes
- Starbucks Coffee
- Charitable donation to Habitat for Humanity
- I will complete the study without receiving compensation

Consent Form

School Psychologists' Decision Making in Special Education Evaluations

You are invited to be in a research study of school psychologists' decision making in evaluations for special education. Shanna Sadeh is conducting this study as part of her dissertation for the School Psychology Program at the University of Minnesota. Please read this form and ask any questions you may have before agreeing to be in the study.

Compensation: Eligible participants who complete the study will receive a \$20 gift card via e-mail for the organization they selected on the previous screen, if any.

Background Information: The purpose of this study is to investigate school psychologists' decision making in initial evaluations of students for special education according to federal law.

Procedures: If you agree to be in this study, we would ask you to do the following: (a) provide demographic information about yourself; (b) read three hypothetical vignettes summarizing assessment data and then determine whether students qualify for special education; and (c) answer questions about your decision making process. We estimate that it will take you about 15-20 minutes to complete these tasks.

Confidentiality: The records of this study will be kept private. In any report we might publish, we will not include any information that will make it possible to identify a subject. Research records will be stored securely and only researchers will have access to the records. Study data will be encrypted according to current University policy for protection of confidentiality.

Risks and Benefits of Being in the Study: There are no known risks if you decide to participate in this research study. There are no costs to you for participating in the study. The information you give will provide guidance regarding factors that influence special education eligibility determinations. The information collected may not benefit you directly, but it should provide more general benefits to the practice of school psychology and special education policy decisions.

Voluntary Nature of the Study: Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University of Minnesota. If you decide to participate, you are free to withdraw at any time without affecting those relationships.

Contacts and Questions: If you have questions now or later, **you are encouraged** to contact Shanna Sadeh by e-mail at sade0041@umn.edu, or her academic adviser, Dr. Amanda Sullivan, whose contact information is: Educational Psychology, 56 East River Parkway, (612) 626-7221, asulliva@umn.edu .

If you have any questions or concerns regarding this study and would like to talk to someone other than the researchers, **you are encouraged** to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 625-1650.

- I have read the above information. I voluntarily consent to participate in the study.
- I will not participate.

Eligibility

Are you a school psychologist who participates in special education evaluations in schools or supervises special education evaluations?

- Yes
- No

Directions**DIRECTIONS**

In the following sections, you will be asked to read 3 hypothetical vignettes about students who have social, emotional, or behavioral problems. You will be presented the qualification criteria for the category of serious emotional disturbance as defined in the Individuals with Disabilities Education Act of 2004 and its implementing regulations. You will then be asked some questions about eligibility and your decision making.

Tyrell/Tom**Assessment Summary**

Tom Simpson is an eleven-year-old, White male in fifth grade who was referred by Ms. Sanders, his classroom teacher, on March 7, 2014 for evaluation. She reported concerns about Tom's disruptive and disrespectful behaviors. The multidisciplinary evaluation team planned an assessment of Tom's functioning at school. The assessment results are summarized below. Please assume this assessment is sufficiently comprehensive to satisfy district evaluation procedures.

Teacher Interview

Ms. Sanders reported that Tom is one of four boys in her class that have "made it impossible for [her] to teach this year." Together, these boys disrupt the class by humming during instruction, throwing wadded paper at her when her back is turned to the class, pretending they are blind when it is time to line up at the door, clapping when students get an answer wrong, and drawing cartoons of her in embarrassing situations – e.g., passing gas in a movie theater. She considers Tom the ringleader of the group and believes the other students are trying to impress him with their antics.

Three months ago, Ms. Sanders implemented a behavior plan for Tom whereby he loses a minute of recess each time he violates a classroom expectation. Consequently, Tom and the other boys have not had recess for six weeks. They are supposed to sit at their desks with their heads down during recess, but instead they play in the classroom. Ms. Sanders also has been calling home with weekly progress reports for the past two months. She has not seen any improvement in Tom's behavior. Academically, Tom has received low scores on many class assignments because he turns them in late or incomplete. His performance on unit tests, however, indicates average achievement relative to his peers.

Developmental/Medical History

Tom's developmental and medical history were unremarkable. Ms. Simpson had a healthy pregnancy with no prenatal exposure to drugs or alcohol. Tom's birth was planned and uncomplicated. Tom met normal developmental milestones for talking, walking, and toilet training. He has no known medical issues that may be affecting his functioning at school.

Family Information

Tom lives with his biological mother and father. Ms. Simpson did not work outside the home until Tom was in first grade. Mr. Simpson is a high school teacher and spends summers at home with Tom. The family has stable housing, employment, and no reported trauma history or current stressors.

Student Interview

Tom hates math and wishes he could read or write instead of doing math. He enjoys reading popular fiction books and writing stories about battles in space. He described his relationship with Ms. Sanders as "okay." He feels comfortable asking her for help when he needs it.

School Psychologist Direct Observations

Over the course of a month, the assessor conducted five direct observations of Tom. These observations typically occurred between 11 a.m. and noon when the class had large group instruction for math. In each session, the assessor conducted a 21 minute Systematic Direct Observation (SDO) of Tom and two of his male peers. The results for SDO are described below.

Academic Engagement

Across five observations, Tom was observed to be academically engaged between 50% and 70% of the total time observed. His comparison peers were observed to be academically engaged between 48% and 90% of the total time observed. Tom was frequently looking at items from his desk or out the window instead of attending to instruction or completing assigned tasks.

Disruptive Behavior

Across five observations, Tom was observed to be disruptive between 40% and 68% of the total time observed. His comparison peers were observed to be disruptive between 75% and 90% of the time. Tom's disruptive behaviors usually were blurting out answers to questions the teacher posed to the group without raising his hand, laughing excessively when the teacher made a mistake, and talking to peers during instruction.

Disrespectful Behavior

Across five observations, Tom engaged in disrespectful behavior between 30% and 55% of the total

observed time. His comparison peers engaged in disrespectful behavior between 45% and 65% of the time. Tom's disrespectful behavior was ignoring instructions to complete tasks.

Test Scores

Behavior Assessment System for Children (BASC) (Results had sufficient validity)

Domain	Teacher	Parent	Student
Behavior Symptoms Index	Significant	Average	Average
Externalizing Problems	Significant	Average	Average
-Hyperactivity	Significant	Average	Average
-Aggression	At-Risk	Average	Average
-Conduct Problems	Significant	Average	Average
Internalizing Problems	Average	Average	Average
-Anxiety	Average	Average	Average
-Depression	Average	Average	Average
-Somatization	Average	Average	Average
-Withdrawal	Average	Average	Average
School Problems	Significant	Average	Average
Adaptive Skills	Average	Average	Average

Woodcock Johnson Test of Cognitive Abilities

General Intellectual Ability – Average (95% CI 89-99)

No significant discrepancies among cluster scores.

Woodcock Johnson Test of Academic Achievement

Broad Reading – Average

Broad Math – Low Average

Broad Written Language – Average

Summary of Record Review

Satisfactory=S

Unsatisfactory=U

Office Referral=OR

Days of Suspension=#X

	Kinder.	1st	2nd	3rd	4th	5th
Classroom performance						
-Math	S	S	S	S	S	S
-Reading	S	S	S	S	S	S
-Writing	S	S	S	S	S	S
Discipline infractions						5 OR
Attendance rate	100%	98%	100%	93%	100%	100%

Assessment Summary

Tyrell Simpson is an eleven-year-old, Black male in fifth grade who was referred by Ms. Sanders, his classroom teacher, on March 7, 2014 for evaluation. She reported concerns about Tyrell's disruptive and disrespectful behaviors. The multidisciplinary evaluation team planned an assessment of Tyrell's functioning at school. The assessment results are summarized below. Please assume this assessment is sufficiently comprehensive to satisfy district evaluation procedures.

Teacher Interview

Ms. Sanders reported that Tyrell is one of four boys in her class that have "made it impossible for [her] to teach this year." Together, these boys disrupt the class by humming during instruction, throwing wadded paper at her when her back is turned to the class, pretending they are blind when it is time to line up at the door, clapping when students get an answer wrong, and drawing cartoons of her in embarrassing situations – e.g., passing gas in a movie theater. She considers Tyrell the ringleader of the group and believes the other students are trying to impress him with their antics. Three months ago, Ms. Sanders implemented a behavior plan for Tyrell whereby he loses a minute of recess each time he violates a classroom expectation. Consequently, Tyrell and the other boys have not had recess for six weeks. They are supposed to sit at their desks with their heads down during recess, but instead they play in the classroom. Ms. Sanders also has been calling home with weekly progress reports for the past two months. She has not seen any improvement in Tyrell's behavior. Academically, Tyrell has received low scores on many class assignments because he turns them in late or incomplete. His performance on unit tests, however, indicates average achievement relative to his peers.

Developmental/Medical History

Tyrell's developmental and medical history were unremarkable. Ms. Simpson had a healthy pregnancy with no prenatal exposure to drugs or alcohol. Tyrell's birth was planned and

uncomplicated. Tyrell met normal developmental milestones for talking, walking, and toilet training. He has no known medical issues that may be affecting his functioning at school.

Family Information

Tyrell lives with his biological mother and father. Ms. Simpson did not work outside the home until Tyrell was in first grade. Mr. Simpson is a high school teacher and spends summers at home with Tyrell. The family has stable housing, employment, and no reported trauma history or current stressors.

Student Interview

Tyrell hates math and wishes he could read or write instead of doing math. He enjoys reading popular fiction books and writing stories about battles in space. He described his relationship with Ms. Sanders as "okay." He feels comfortable asking her for help when he needs it.

School Psychologist Direct Observations

Over the course of a month, the assessor conducted five direct observations of Tyrell. These observations typically occurred between 11 a.m. and noon when the class had large group instruction for math. In each session, the assessor conducted a 21 minute Systematic Direct Observation (SDO) of Tom and two of his male peers. The results for SDO are described below.

Academic Engagement

Across five observations, Tyrell was observed to be academically engaged between 50% and 70% of the total time observed. His comparison peers were observed to be academically engaged between 48% and 90% of the total time observed. Tyrell was frequently looking at items from his desk or out the window instead of attending to instruction or completing assigned tasks.

Disruptive Behavior

Across five observations, Tyrell was observed to be disruptive between 40% and 68% of the total time observed. His comparison peers were observed to be disruptive between 75% and 90% of the time. Tyrell's disruptive behaviors usually were blurting out answers to questions the teacher posed to the group without raising his hand, laughing excessively when the teacher made a mistake, and talking to peers during instruction.

Disrespectful Behavior

Across five observations, Tyrell engaged in disrespectful behavior between 30% and 55% of the total observed time. His comparison peers engaged in disrespectful behavior between 45% and 65% of the time. Tyrell's disrespectful behavior was ignoring instructions to complete tasks.

Test Scores

Behavior Assessment System for Children (BASC) (Results had sufficient validity)

Domain	Teacher	Parent	Student
Behavior Symptoms Index	Significant	Average	Average
Externalizing Problems	Significant	Average	Average

-Hyperactivity	Significant	Average	Average
-Aggression	At-Risk	Average	Average
-Conduct Problems	Significant	Average	Average
Internalizing Problems	Average	Average	Average
-Anxiety	Average	Average	Average
-Depression	Average	Average	Average
-Somatization	Average	Average	Average
-Withdrawal	Average	Average	Average
School Problems	Significant	Average	Average
Adaptive Skills	Average	Average	Average

Woodcock Johnson Test of Cognitive Abilities

General Intellectual Ability – Average (95% CI 89-99)

No significant discrepancies among cluster scores.

Woodcock Johnson Test of Academic Achievement

Broad Reading – Average

Broad Math – Low Average

Broad Written Language – Average

Summary of Record Review

Satisfactory=S

Unsatisfactory=U

Office Referral=OR

Days of Suspension=#X

	Kinder.	1st	2nd	3rd	4th	5th
Classroom performance						
-Math	S	S	S	S	S	S
-Reading	S	S	S	S	S	S
-Writing	S	S	S	S	S	S
Discipline infractions						5 OR

Attendance rate	100%	98%	100%	93%	100%	100%
-----------------	------	-----	------	-----	------	------

Law Governing Emotional Disturbance

A child with a disability means a child:

(i) with intellectual disabilities, hearing impairments (including deafness), speech or language impairments, visual impairments (including blindness), serious emotional disturbance (referred to as "emotional disturbance"), orthopedic impairments, autism, traumatic brain injury, other health impairments, or specific learning disabilities; and (ii) who, by reason thereof, needs special education and related services.

Emotional disturbance is:

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

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 this section.

Does the student qualify for special education for emotional disturbance under federal law?

- Yes, he qualifies for special education for emotional disturbance.
- No, he does not qualify for special education for emotional disturbance.

What data were most helpful in making this decision? Please select 3-4 types of data.

- Teacher interview
- Student interview
- School psychologist's direct observation
- Classroom grades
- Discipline history
- BASC results
- Cognitive testing results
- Standardized achievement results
- Developmental history
- Family information
- Attendance history

What additional data, if any, would have helped you make this decision?

How confident are you in your decision?

- High confidence
- Moderate confidence
- Low confidence

Jamal/Jay

Assessment Summary

Jamal Jones is an eleven-year-old, Black male in fifth grade who was referred by Ms. Sullivan, his classroom teacher, on March 3, 2014 for evaluation. She reported concerns about Jamal's unprovoked aggression towards peers, outbursts that disrupt class instruction, refusal to comply with adult directives, and disrespectful attitude towards authority figures. The multidisciplinary evaluation team planned an assessment of Jamal's functioning at school. The assessment results are summarized below. Please assume this assessment is sufficiently comprehensive to satisfy district evaluation procedures.

Development/Medical History

Jamal was born premature at 27 weeks and spent 3 months in the hospital. He met all of his developmental milestones for walking, talking, and toilet training within normal time ranges. Jamal was a fussy baby that was difficult to soothe. Ms. Jones had concerns about his long and frequent temper tantrums and sensitive emotions during his early childhood. Jamal was diagnosed by mental health professionals with attention deficit hyperactivity disorder (ADHD) in kindergarten and oppositional defiance disorder in second grade. He manages his ADHD with medication. Children's Psychological Services, an outside provider, has conducted individual and family therapy with Jamal and his family on a weekly basis for the past two years.

Family History

Jamal lives with Ms. Jones, his sister Leticia (age 13), sister Shondra (age 11), and his maternal grandmother, Aikira. The family has stable housing and experiences some stress due to concerns about finances. Aikira usually cares for the children after school and on the weekends because of Ms. Jones' work schedule. Aikira reported that Jamal has been "running wild" this year and refuses to obey her instructions to stay in the apartment, complete his homework, or help with chores. She reported he has always had a very active body and difficulty paying attention. Since he was 2 or 3 years old, his general mood has been irritable and he has had tantrums that can last 20-30 minutes, during which time he will throw anything around him, punch holes in the wall, scream, curse, and sometimes cry. These episodes happen 3 to 4 times a week. The Jones expected Jamal to grow out of these tantrums, but the problems have gotten worse over time. This year is the first year that Jamal has had trouble at school. Previously, he was able to "hold it together" at school and then he seemed to "unravel" once he got home. As Jamal has grown bigger, his outbursts have become more destructive and frighten family.

Teacher Interview

Ms. Sullivan reported that Jamal has consistently had behavior problems in her class since the school year began. Sometimes he appears to act out in order to get attention from peers; other times he seems to act out because he is overcome with emotion and then hits, yells, or throws things impulsively. He has a low-frustration tolerance for trivial setbacks like a pencil point breaking or not being called on to answer a question. He has no apparent friends in the class. His interactions with peers usually involve teasing or bossing others around. His interactions with adults are usually conflictual even if the adult attempts to engage Jamal in small talk about non-academic subjects like sports.

Ms. Sullivan has implemented several interventions to support Jamal. For instance, in September, she moved Jamal's desk to the front row near her desk in order to help him pay attention and to give him verbal reminders to stay on task. His behavior improved once he had preferential seating, but the improvement lasted for only two or three days. In October, Ms. Sullivan implemented a

behavior plan using positive reinforcement to elicit appropriate classroom behaviors. This intervention did not result in any changes in Jamal's problematic behaviors.

Ms. Sullivan reported that Jamal's strengths are that he is a bright student and could achieve academic success if he attended to instruction and completed academic tasks. Jamal's grandmother frequently communicates with her about his school progress and tries to implement consequences at home when Jamal acts out in school.

Student Interview

Jamal acknowledged that he has had trouble following rules in school this year. He identified the following triggers: students making too much noise or getting too close to him, teachers not allowing him to read his comic books, and people telling him what to do. He reported that he tends to "shut down" or "explode" when he gets angry and that he is trying to practice in school the skills he is learning in therapy, but that it is hard for him to do so. His favorite activities in school are science experiments and gym, especially basketball. He denied having any close friendships. At home, he enjoys playing video games, watching basketball, and reading comic books.

School Psychologist Direct Observations

Over the course of a month, the assessor conducted five direct observations of Jamal. These observations typically occurred between 11 a.m. and noon when the class had large group instruction for reading or writing. In each session, the assessor conducted a 21 minute Systematic Direct Observation (SDO) of Jamal and two of his male peers. The results for SDO are described below.

Academic Engagement

Across five observations, Jamal was observed to be academically engaged between 20% and 30% of the total time observed. His comparison peers were observed to be academically engaged between 47% and 89% of the total time observed. When Jamal did engage with his assignments, he seemed invested in his work product and motivated to do well because he often proudly shared his assignments with other students and participated in large group discussion.

Disruptive Behavior

Across five observations, Jamal was observed to be disruptive between 21% and 46% of the total time observed. His comparison peers were observed to be disruptive between 0% and 10% of the time.

The SDO assessment method did not capture the magnitude or form of Jamal's disruptive behavior. Unlike his comparison peers, Jamal danced next to his desk during large-group instruction or independent work time. He also wandered around the room disturbing other students and rapped aloud. His comparison peers were disruptive by tapping a pencil or whispering to their neighbors, which interfered less with instruction. Because of this difference in the magnitude of disruptive behavior Jamal displayed, the SDO ratings probably under-represent the degree to which his behavior impacted his learning, his peers' learning, and his teacher's ability to lead the class.

Disrespectful Behavior

Across five observations, Jamal engaged in disrespectful behavior between 60% and 80% of the total observed time. His comparison peers engaged in disrespectful behavior between 0% and 11%.

Test Scores**Behavior Assessment System for Children (BASC)**

(Results had sufficient validity)

Domain	Teacher	Parent	Student
Behavioral Symptoms Index	Significant	Significant	At-Risk
Externalizing Problems	Significant	Significant	Average
- Hyperactivity	Significant	Significant	At-Risk
- Aggression	Significant	Significant	Average
- Conduct Problems	Significant	Significant	Average
Internalizing Problems	Average	Significant	Significant
- Anxiety	Average	At-Risk	Significant
-Depression	Average	Significant	At-Risk
- Somatization	Average	Average	Significant
-Withdrawal	Average	Significant	At-Risk
School Problems	Significant	Significant	At-Risk
Adaptive Skills	At-Risk	Average	Average

Woodcock Johnson Test of Cognitive Abilities

General Intellectual Ability – Average (95% CI 89-99)

No significant discrepancies among cluster scores.

Woodcock Johnson Test of Academic Achievement

Broad Reading – Average

Broad Math – Low Average

Broad Written Language – Average

Summary of Record Review

Satisfactory=S

Unsatisfactory=U

Office Referral=OR

Days of Suspension=#X

	Kinder.	1st Grade	2nd Grade	3rd Grade	4th Grade	5th Grade

Classroom performance						
Math	S	U	S	S	U	U
Reading	S	U	S	U	S	U
Writing	S	S	S	S	S	U
Discipline infractions	5 OR	3 OR, 1x	6 OR	4 OR, 3X	9 OR, 1X	8 OR, 2X
Attendance rate	100%	98%	100%	93%	95%	99%

Assessment Summary

Jay Jones is an eleven-year-old, White male in fifth grade who was referred by Ms. Sullivan, his classroom teacher, on March 3, 2014 for evaluation. She reported concerns about Jay's unprovoked aggression towards peers, outbursts that disrupt class instruction, refusal to comply with adult directives, and disrespectful attitude towards authority figures. The multidisciplinary evaluation team planned an assessment of Jay's functioning at school. The assessment results are summarized below. Please assume this assessment is sufficiently comprehensive to satisfy district evaluation procedures.

Development/Medical History

Jay was born premature at 27 weeks and spent 3 months in the hospital. He met all of his developmental milestones for walking, talking, and toilet training within normal time ranges. Jay was a fussy baby that was difficult to soothe. Ms. Jones had concerns about his long and frequent temper tantrums and sensitive emotions during his early childhood. Jay was diagnosed by mental health professionals with attention deficit hyperactivity disorder (ADHD) in kindergarten and oppositional defiance disorder in second grade. He manages his ADHD with medication. Children's Psychological Services, an outside provider, has conducted individual and family therapy with Jay and his family on a weekly basis for the past two years.

Family History

Jay lives with Ms. Jones, his sister Leah (age 13), sister Sara (age 11), and his maternal grandmother, Ashley. The family has stable housing and experiences some stress due to concerns about finances. Ashley usually cares for the children after school and on the weekends because of Ms. Jones' work schedule. Ashley reported that Jay has been "running wild" this year and refuses to obey her instructions to stay in the apartment, complete his homework, or help with chores. She reported he has always had a very active body and difficulty paying attention. Since he was 2 or 3 years old, his general mood has been irritable and he has had tantrums that can last 20-30 minutes, during which time he will throw anything around him, punch holes in the wall, scream, curse, and sometimes cry. These episodes happen 3 to 4 times a week. The Jones expected Jay to grow out of

these tantrums, but the problems have gotten worse over time. This year is the first year that Jay has had trouble at school. Previously, he was able to “hold it together” at school and then he seemed to “unravel” once he got home. As Jay has grown bigger, his outbursts have become more destructive and frighten family.

Teacher Interview

Ms. Sullivan reported that Jay has consistently had behavior problems in her class since the school year began. Sometimes he appears to act out in order to get attention from peers; other times he seems to act out because he is overcome with emotion and then hits, yells, or throws things impulsively. He has a low-frustration tolerance for trivial setbacks like a pencil point breaking or not being called on to answer a question. He has no apparent friends in the class. His interactions with peers usually involve teasing or bossing others around. His interactions with adults are usually conflictual even if the adult attempts to engage Jay in small talk about non-academic subjects like sports.

Ms. Sullivan has implemented several interventions to support Jay. For instance, in September, she moved Jay’s desk to the front row near her desk in order to help him pay attention and to give him verbal reminders to stay on task. His behavior improved once he had preferential seating, but the improvement lasted for only two or three days. In October, Ms. Sullivan implemented a behavior plan using positive reinforcement to elicit appropriate classroom behaviors. This intervention did not result in any changes in Jay’s problematic behaviors.

Ms. Sullivan reported that Jay’s strengths are that he is a bright student and could achieve academic success if he attended to instruction and completed academic tasks. Jay’s grandmother frequently communicates with her about his school progress and tries to implement consequences at home when Jay acts out in school.

Student Interview

Jay acknowledged that he has had trouble following rules in school this year. He identified the following triggers: students making too much noise or getting too close to him, teachers not allowing him to read his comic books, and people telling him what to do. He reported that he tends to “shut down” or “explode” when he gets angry and that he is trying to practice in school the skills he is learning in therapy, but that it is hard for him to do so. His favorite activities in school are science experiments and gym, especially basketball. He denied having any close friendships. At home, he enjoys playing video games, watching basketball, and reading comic books.

School Psychologist Direct Observations

Over the course of a month, the assessor conducted five direct observations of Jay. These observations typically occurred between 11 a.m. and noon when the class had large group instruction for reading or writing. In each session, the assessor conducted a 21 minute Systematic Direct Observation (SDO) of Jay and two of his male peers. The results for SDO are described below.

Academic Engagement

Across five observations, Jay was observed to be academically engaged between 20% and 30% of the total time observed. His comparison peers were observed to be academically engaged between 47% and 89% of the total time observed. When Jay did engage with his assignments, he seemed

invested in his work product and motivated to do well because he often proudly shared his assignments with other students and participated in large group discussion.

Disruptive Behavior

Across five observations, Jay was observed to be disruptive between 21% and 46% of the total time observed. His comparison peers were observed to be disruptive between 0% and 10% of the time.

The SDO assessment method did not capture the magnitude or form of Jay's disruptive behavior. Unlike his comparison peers, Jay danced next to his desk during large-group instruction or independent work time. He also wandered around the room disturbing other students and rapped aloud. His comparison peers were disruptive by tapping a pencil or whispering to their neighbors, which interfered less with instruction. Because of this difference in the magnitude of disruptive behavior Jay displayed, the SDO ratings probably under-represent the degree to which his behavior impacted his learning, his peers' learning, and his teacher's ability to lead the class.

Disrespectful Behavior

Across five observations, Jay engaged in disrespectful behavior between 60% and 80% of the total observed time. His comparison peers engaged in disrespectful behavior between 0% and 11%.

Test Scores

Behavior Assessment System for Children (BASC)

(Results had sufficient validity)

Domain	Teacher	Parent	Student
Behavioral Symptoms Index	Significant	Significant	At-Risk
Externalizing Problems	Significant	Significant	Average
- Hyperactivity	Significant	Significant	At-Risk
- Aggression	Significant	Significant	Average
- Conduct Problems	Significant	Significant	Average
Internalizing Problems	Average	Significant	Significant
- Anxiety	Average	At-Risk	Significant
- Depression	Average	Significant	At-Risk
- Somatization	Average	Average	Significant
- Withdrawal	Average	Significant	At-Risk
School Problems	Significant	Significant	At-Risk

Adaptive Skills	At-Risk	Average	Average
-----------------	---------	---------	---------

Woodcock Johnson Test of Cognitive Abilities

General Intellectual Ability – Average (95% CI 89-99)

No significant discrepancies among cluster scores.

Woodcock Johnson Test of Academic Achievement

Broad Reading – Average

Broad Math – Low Average

Broad Written Language – Average

Summary of Record Review

Satisfactory=S

Unsatisfactory=U

Office Referral=OR

Days of Suspension=#X

	Kinder.	1st Grade	2nd Grade	3rd Grade	4th Grade	5th Grade
Classroom performance						
Math	S	U	S	S	U	U
Reading	S	U	S	U	S	U
Writing	S	S	S	S	S	U
Discipline infractions	5 OR	3 OR, 1x	6 OR	4 OR, 3X	9 OR, 1X	8 OR, 2X
Attendance rate	100%	98%	100%	93%	95%	99%

Law Governing Emotional Disturbance

A *child with a disability* means a child:

(i) with intellectual disabilities, hearing impairments (including deafness), speech or language impairments, visual impairments (including blindness), serious emotional disturbance (referred to as "emotional disturbance"), orthopedic impairments, autism, traumatic brain injury, other health impairments, or specific learning disabilities; and (ii) who, by reason thereof, needs special education and related services.

Emotional disturbance is:

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

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 this section.

Does the student qualify for special education for emotional disturbance under federal law?

- Yes, he qualifies for special education for emotional disturbance.
- No, he does not qualify for special education for emotional disturbance.

What data were most helpful in making this decision? Please select 3-4 types of data.

- Teacher interview
- Student interview
- School psychologist's direct observation
- Classroom grades
- Discipline history
- BASC results
- Cognitive testing results
- Standardized achievement results
- Developmental history
- Family information
- Attendance history

What additional data, if any, would have helped you make this decision?

How confident are you in your decision?

- High confidence
- Moderate confidence
- Low confidence

Antwan/Andy

Assessment Summary

Antwan Williams is an eleven-year-old, Black male in fifth grade who was referred by Ms. Thomas, his classroom teacher, on April 7, 2014 for evaluation. She reported concerns about Antwan's outbursts that disrupt class instruction, refusal to comply with adult directives, and disrespectful attitude towards authority figures. The multidisciplinary evaluation team planned an assessment of Antwan's functioning at school. The assessment results are summarized below. Please assume this assessment is sufficiently comprehensive to satisfy district evaluation procedures.

Developmental/Medical History

Antwan's developmental and medical history were unremarkable. Ms. Williams had a healthy pregnancy with no prenatal exposure to drugs or alcohol. Antwan's birth was planned and uncomplicated. Antwan met normal developmental milestones for talking, walking, and toilet training. He has no known medical issues that may be affecting his functioning at school.

Family Information

Antwan lives with his biological mother and father and two younger sisters (Tysha, age 9 and Rayna, age 7). Mr. Williams has been out of work for approximately a year and a half. Ms. Williams reported some family stress around her husband's job loss, but that the family can meet its financial obligations with her salary. The family has stable housing and no reported trauma history. Ms. Williams noticed a change in Antwan last summer, about nine months ago. He showed less interest in activities that used to excite him (e.g. basketball, swimming, going to the park) and spent a lot more time inside by himself. She reported that he frequently appears sad or tired although he has an early bedtime and seems to sleep through the night. She attributed his moody and sullen attitude to his age. Ms. Williams wants to understand why Antwan is acting out in school and how she can help.

Teacher Interview

Ms. Thomas reported that Antwan disrupts the class by speaking without permission, swearing, and refusing to follow the classroom routine and procedures – e.g., not asking permission before taking the bathroom pass, sharpening his pencil during instruction, refusing to pass papers to the students behind him, and laughing excessively when Ms. Thomas makes an error. Ms. Thomas described Antwan as moody, irritable, and disrespectful. Ms. Thomas has tried to get to know Antwan in order to improve their relationship, but Antwan has not responded positively to her attempts. For example, Ms. Thomas invited Antwan to have lunch with her, which students usually consider a special reward. Antwan declined this invitation using curse words.

Student Interview

Within the past month, Antwan's father moved out of the family home. Antwan reported that this change upset him and that he often cried at home in the afternoon because he missed his father. He also reported that he felt close to his mom and appreciated his father's attendance at his basketball events. Antwan admitted that some of his outbursts at school may be related to his feelings about recent events at home.

Antwan reported that his academic performance last semester was "bad" and that he intended to do better this semester. Antwan disclosed that his classroom teacher, Ms. Thomas, wanted him to do well in school and had confidence in Antwan. He stated that he understood what was expected of him in class and that he liked being in Ms. Thomas's class. Overall, Antwan expressed a clear understanding of how he could be successful and did not try to excuse his poor academic

performance.

School Psychologist Direct Observations

Over the course of a month, the assessor conducted five direct observations of Antwan. These observations typically occurred between 11 a.m. and noon when the class had large group instruction for reading or writing. In each session, the assessor conducted a 21 minute Systematic Direct Observation (SDO) of Antwan and two of his male peers. The results for SDO are described below.

Academic Engagement

Across five observations, Antwan was observed to be academically engaged between 50% and 75% of the total time observed. His comparison peers were observed to be academically engaged between 50% and 89% of the total time observed. Antwan did not appear to struggle with the difficulty of the academic tasks presented to him. Sometimes he was off-task because he was providing unsolicited advice to a peer on how to do the task.

Disruptive Behavior

Across five observations, Antwan was observed to be disruptive between 10% and 30% of the total time observed. His comparison peers were observed to be disruptive between 5% and 15% of the time. Antwan's disruptive behaviors usually were blurting out answers to questions the teacher posed to the group without raising his hand, talking to peers during instruction, and tapping rhythms with his hands on his desk during independent quiet work time. He responded positively to redirection from Ms. Thomas but then usually engaged in the undesirable behavior again after temporary compliance.

Disrespectful Behavior

Across five observations, Antwan engaged in disrespectful behavior between 25% and 50% of the total observed time. His comparison peers engaged in disrespectful behavior between 5% and 15%. Antwan's disrespectful behaviors included swearing at his peers and Ms. Thomas, and refusing to comply with routine directions like putting away folders, sitting on carpet, etc.

Test Scores

Behavior Assessment System for Children (BASC)

(Results had sufficient validity)

Domain	Teacher	Parent	Student
Behavioral Symptoms Index	Significant	At-Risk	At-Risk
Externalizing Problems	Significant	Average	Average
- Hyperactivity	Significant	Significant	At-Risk
- Aggression	At-Risk	Average	Average
- Conduct Problems	Significant	Average	Average
Internalizing Problems	Average	Significant	Significant

- Anxiety	Average	At-Risk	Significant
-Depression	Average	Significant	At-Risk
- Somatization	Average	Average	Significant
-Withdrawal	Average	Significant	At-Risk
School Problems	Significant	At-Risk	At-Risk
Adaptive Skills	At-Risk	Average	Average

Woodcock Johnson Test of Cognitive Abilities

General Intellectual Ability – Average (95% CI 89-99)

No significant discrepancies among cluster scores.

Woodcock Johnson Test of Academic Achievement

Broad Reading – Average

Broad Math – Low Average

Broad Written Language – Average

Summary of Record Review

Satisfactory=S

Unsatisfactory=U

Office Referral=OR

Days of Suspension=#X

	Kinder.	1st Grade	2nd Grade	3rd Grade	4th Grade	5th Grade
Classroom performance						
Math	S	S	S	S	U	U
Reading	S	U	S	U	S	S
Writing	S	S	S	U	S	U
Discipline infractions			3 OR	6 OR		5 OR, 3X
Attendance rate	100%	98%	100%	93%	95%	99%

Assessment Summary

Andy Williams is an eleven-year-old, White male in fifth grade who was referred by Ms. Thomas, his classroom teacher, on April 7, 2014 for evaluation. She reported concerns about Andy's outbursts that disrupt class instruction, refusal to comply with adult directives, and disrespectful attitude towards authority figures. The multidisciplinary evaluation team met and planned an assessment of Andy's functioning at school. The assessment results are summarized below. Please assume this assessment is sufficiently comprehensive to satisfy district evaluation procedures.

Developmental/Medical History

Andy's developmental and medical history were unremarkable. Ms. Williams had a healthy pregnancy with no prenatal exposure to drugs or alcohol. Andy's birth was planned and uncomplicated. Andy met normal developmental milestones for talking, walking, and toilet training. He has no known medical issues that may be affecting his functioning at school.

Family Information

Andy lives with his biological mother and father and two younger sisters (Tara, age 9 and Rebecca, age 7). Mr. Williams has been out of work for approximately a year and a half. Ms. Williams reported some family stress around her husband's job loss, but that the family can meet its financial obligations with her salary. The family has stable housing and no reported trauma history. Ms. Williams noticed a change in Andy last summer, about nine months ago. He showed less interest in activities that used to excite him (e.g. basketball, swimming, going to the park) and spent a lot more time inside by himself. She reported that he frequently appears sad or tired although he has an early bedtime and seems to sleep through the night. She attributed his moody and sullen attitude to his age. Ms. Williams wants to understand why Andy is acting out in school and how she can help.

Teacher Interview

Ms. Thomas reported that Andy disrupts the class by speaking without permission, swearing, and refusing to follow the classroom routine and procedures – e.g., not asking permission before taking the bathroom pass, sharpening his pencil during instruction, refusing to pass papers to the students behind him, and laughing excessively when Ms. Thomas makes an error. Ms. Thomas described Andy as moody, irritable, and disrespectful. Ms. Thomas has tried to get to know Andy in order to improve their relationship, but Andy has not responded positively to her attempts. For example, Ms. Thomas invited Andy to have lunch with her, which students usually consider a special reward. Andy declined this invitation using curse words.

Student Interview

Within the past month, Andy's father moved out of the family home. Andy reported that this change upset him and that he often cried at home in the afternoon because he missed his father. He also reported that he felt close to his mom and appreciated his father's attendance at his basketball events. Andy admitted that some of his outbursts at school may be related to his feelings about recent events at home.

Andy reported that his academic performance last semester was "bad" and that he intended to do better this semester. Andy disclosed that his classroom teacher, Ms. Thomas, wanted him to do well in school and had confidence in Andy. He stated that he understood what was expected of him in class and that he liked being in Ms. Thomas's class. Overall, Andy expressed a clear understanding

of how he could be successful and did not try to excuse his poor academic performance.

School Psychologist Direct Observations

Over the course of a month, the assessor conducted five direct observations of Andy. These observations typically occurred between 11 a.m. and noon when the class had large group instruction for reading or writing. In each session, the assessor conducted a 21 minute Systematic Direct Observation (SDO) of Andy and two of his male peers. The results for SDO are described below.

Academic Engagement

Across five observations, Andy was observed to be academically engaged between 50% and 75% of the total time observed. His comparison peers were observed to be academically engaged between 50% and 89% of the total time observed. Andy did not appear to struggle with the difficulty of the academic tasks presented to him. Sometimes he was off-task because he was providing unsolicited advice to a peer on how to do the task.

Disruptive Behavior

Across five observations, Andy was observed to be disruptive between 10% and 30% of the total time observed. His comparison peers were observed to be disruptive between 5% and 15% of the time. Andy's disruptive behaviors usually were blurting out answers to questions the teacher posed to the group without raising his hand, talking to peers during instruction, and tapping rhythms with his hands on his desk during independent quiet work time. He responded positively to redirection from Ms. Thomas but then usually engaged in the undesirable behavior again after temporary compliance.

Disrespectful Behavior

Across five observations, Andy engaged in disrespectful behavior between 25% and 50% of the total observed time. His comparison peers engaged in disrespectful behavior between 5% and 15%. Andy's disrespectful behaviors included swearing at his peers and Ms. Thomas, and refusing to comply with routine directions like putting away folders, sitting on carpet, etc.

Test Scores

Behavior Assessment System for Children (BASC)

(Results had sufficient validity)

Domain	Teacher	Parent	Student
Behavioral Symptoms Index	Significant	At-Risk	At-Risk
Externalizing Problems	Significant	Average	Average
- Hyperactivity	Significant	Significant	At-Risk
- Aggression	At-Risk	Average	Average
- Conduct Problems	Significant	Average	Average
Internalizing Problems	Average	Significant	Significant

- Anxiety	Average	At-Risk	Significant
-Depression	Average	Significant	At-Risk
- Somatization	Average	Average	Significant
-Withdrawal	Average	Significant	At-Risk
School Problems	Significant	At-Risk	At-Risk
Adaptive Skills	At-Risk	Average	Average

Woodcock Johnson Test of Cognitive Abilities

General Intellectual Ability – Average (95% CI 89-99)

No significant discrepancies among cluster scores.

Woodcock Johnson Test of Academic Achievement

Broad Reading – Average

Broad Math – Low Average

Broad Written Language – Average

Summary of Record Review

Satisfactory=S

Unsatisfactory=U

Office Referral=OR

Days of Suspension=#X

	Kinder.	1st Grade	2nd Grade	3rd Grade	4th Grade	5th Grade
Classroom performance						
Math	S	S	S	S	U	U
Reading	S	U	S	U	S	S
Writing	S	S	S	U	S	U
Discipline infractions			3 OR	6 OR		5 OR, 3X
Attendance rate	100%	98%	100%	93%	95%	99%

Law Governing Emotional Disturbance

A *child with a disability* means a child:

(i) with intellectual disabilities, hearing impairments (including deafness), speech or language impairments, visual impairments (including blindness), serious emotional disturbance (referred to as "emotional disturbance"), orthopedic impairments, autism, traumatic brain injury, other health impairments, or specific learning disabilities; and (ii) who, by reason thereof, needs special education and related services.

Emotional disturbance is:

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

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 this section.

Does the student qualify for special education for emotional disturbance under federal law?

- Yes, he qualifies for special education for emotional disturbance.
- No, he does not qualify for special education for emotional disturbance.

What data were most helpful in making this decision? Please select 3-4 types of data.

- Teacher interview
- Student interview
- School psychologist's direct observation
- Classroom grades
- Discipline history
- BASC results
- Cognitive testing results
- Standardized achievement results
- Developmental history
- Family information
- Attendance history

What additional data, if any, would have helped you make this decision?

How confident are you in your decision?

- High confidence
- Moderate confidence
- Low confidence

Demographics

To help us interpret our results, we would like to know some information about you and your school psychology practice.

What is your gender?

- Male
- Female
- Other

What is your race/ethnicity?

- American Indian or Alaska Native
- Asian
- Black
- Hispanic
- Native Hawaiian and Other Pacific Islander
- White (non-Hispanic)
- More than 1 race/ethnicity
- I do not wish to disclose my race/ethnicity

In which state do you work?

What is your highest degree related to the field of school psychology?

- M.A.
- Ph.D.
- Psy.D.
- Ed.D.
- C.A.G.S.

How many years have you been practicing school psychology?

- 0-5 years
- 6-10 years
- 11-20 years
- 21-30 years
- 31-40 years
- 41 or more years years

In what type of setting do you practice most often?

- Public school
- Private school
- Clinical setting
- Other setting

How would you best describe the region where you work?

- Urban
- Suburban
- Rural

Approximately how many initial special education evaluations are you involved in during a typical academic school year?

- 1-20
- 21-50
- 51-75
- 76-100
- 101 or more

Does your school district disproportionately identify more Black students as having an emotional disturbance than White students?

- Yes
- No
- I do not know