

Teacher Decision-Making about Student Mental Health: The Role of Race and Gender

by

Shelley Alonso-Marsden

Department of Psychology and Neuroscience  
Duke University

Date: \_\_\_\_\_

Approved:

\_\_\_\_\_  
Kenneth Dodge, Supervisor

\_\_\_\_\_  
Gary Bennett

\_\_\_\_\_  
Laura Richman

\_\_\_\_\_  
Kathryn Whetten

Dissertation submitted in partial fulfillment of  
the requirements for the degree of Doctor  
of Philosophy in the Department of  
Chemistry in the Graduate School  
of Duke University

2017

ABSTRACT

Teacher Decision-Making about Student Mental Health: The Role of Race and Gender

by

Shelley Alonso-Marsden

Department of Psychology and Neuroscience  
Duke University

Date: \_\_\_\_\_

Approved:

\_\_\_\_\_  
Kenneth Dodge, Supervisor

\_\_\_\_\_  
Gary Bennett

\_\_\_\_\_  
Laura Richman

\_\_\_\_\_  
Kathryn Whetten

An abstract of a dissertation submitted in partial  
fulfillment of the requirements for the degree  
of Doctor of Philosophy in the Department of  
Chemistry in the Graduate School of  
Duke University

2017

Copyright by  
Shelley Alonso-Marsden  
2017

## **Abstract**

Mental health issues are prevalent in childhood, but the majority of problems go untreated. Black children and girls are less likely than their White male peers to receive mental health treatment when needed. As a primary gatekeeper, teachers may be partially responsible for these disparities. The goal of the current study was to measure implicit race and gender bias at several steps in the decision-making process leading up to, and including, referral for services. Participants were 1,106 public middle school teachers in the state of North Carolina. Teachers were presented a series of vignettes that described complex behavior, including both internalizing and externalizing problems. Perceived child race and gender was manipulated via normed given names; vignettes were otherwise identical within each condition. Controlling for teacher demographics and randomized condition, vignettes with a male name were attributed more externalizing problems, rated marginally more severe, and more likely to be referred for academic and behavioral interventions than those with female names. Vignettes with Black names were less likely to be referred for mental health treatment and marginally more likely to be referred for academic services. Results of this analysis suggest that implicit racial and gender bias may play an important role in teacher assessments of problematic child behavior and referral for services. Future research is needed to further explicate these processes and how they correlate with in vivo decision-making.

## **Dedication**

To my mother, Sara Marsden, my best friend and my inspiration. I would never have made it through my doctoral program (or 2016, or 2015) without your love and encouragement.

To all the little girls...never doubt that you are valuable and powerful and deserving of every chance and opportunity in the world to pursue and achieve your own dreams.

– Hillary Clinton

# Contents

Abstract .....	iv
List of Tables .....	ix
List of Figures .....	xi
Acknowledgements .....	xii
1. Introduction .....	1
1.1 Behavioral Presentations of Mental Illness in Childhood .....	2
1.2 Disparities in Childhood Mental Health Treatment.....	4
1.3 Critique of Previous Research Methods.....	6
1.4 Current Study.....	7
1.4.1 Social cognitive model .....	8
1.4.2 Bias in teacher implicit cognition of mental health symptoms.....	12
1.5. Study Aims and Hypotheses .....	14
2. Method .....	18
2.1 Study Design .....	18
2.2 Power Considerations and Sample Size Analysis.....	19
2.3 Recruitment .....	21
2.4 Procedure.....	22
2.5 Vignette Development.....	23
2.6 Materials .....	25
2.7 Pilot Study .....	29

2.8 Analysis Plan.....	30
3. Results.....	33
3.1 Sample Characteristics.....	33
3.2 Effect of Perceived Student Race on Problem Type.....	36
3.2.1 Multilevel Analysis .....	42
3.3 Effect of Perceived Student Race on Problem Type.....	44
3.3.1 Multilevel Analysis .....	50
3.4 Predictors of Ratings of Problem Severity, Referral, and Confidence .....	51
3.4.1 Predictors of Teacher Ratings of Problem Severity .....	52
3.4.2 Multilevel Analysis .....	54
3.4.3 Predictors of Referral for Services.....	55
3.4.4 Predictors of Referral Confidence .....	56
3.4.5 Multilevel Analysis .....	59
3.5 Predictors of Ratings of Service Referral Type.....	61
3.5.1 Multilevel Analysis Predicting Number of Academic Service Referrals.....	64
3.5.2 Multilevel Analysis Predicting Number of Behavioral Service Referrals .....	66
3.5.3 Multilevel Analysis Predicting Number of Mental Health Service Referrals...	67
3.5.4 Mediators of Child Vignette Demographics on Service Referral Decision- Making, Exploratory Analyses .....	69
3.6. Measures of Implicit Bias (Implicit Association Test) .....	72
3.6.1 Multilevel Analysis .....	75
4. Discussion .....	78

4.1 Teacher Recognition of Child Problem Type.....	79
4.2 Teacher Ratings of Problem Severity.....	81
4.3 Teacher Ratings of Confidence in Service Referral.....	82
4.4 Teacher Referrals for Services.....	82
4.5 Implicit Association Test.....	84
4.6 Limitations.....	84
4.7 Future Directions.....	86
4.8 Conclusions.....	88
Appendix A: Informed Consent Documents.....	89
Appendix A1: Email solicitation for participation.....	89
Appendix A2: Online Consent Form.....	90
Appendix B: Study Materials.....	92
Appendix B1: Sample Vignettes.....	92
Appendix B2: Child Behavior Information Processing Questionnaire.....	93
Appendix B3: Implicit Association Test.....	97
References.....	99
Biography.....	106



## List of Tables

Table 1: Pilot study ratings of child names .....	25
Table 2: Bonferroni correction for family-wise error: p-value used for significance testing. ....	32
Table 3: Observed dependent variable means for each 2 (gender) x 2 (race) x 3 (behavioral condition) category. ....	35
Table 4: Multinomial logistic regression predicting externalizing vs. internalizing primary problem type and externalizing primary problem type vs. no problem .....	39
Table 5: Multiple regression results predicting number of internalizing and externalizing problems.....	41
Table 6: Multivariate multilevel logit models predicting externalizing problem type .....	44
Table 7: Multinomial logistic regression results predicting externalizing vs. internalizing primary problem type and externalizing primary problem type vs. no problem. ....	46
Table 8: Multiple regression results predicting number of internalizing and externalizing problems.....	48
Table 9: Multivariate multilevel regression models predicting number of externalizing problems.....	51
Table 10: Multiple regression results predicting teacher rating of problem severity. ....	53
Table 11: Multivariate multilevel regression models predicting teacher rating of problem severity. ....	56
Table 12: Multiple regression results predicting teacher rating of confidence in service referral. ....	58
Table 13: Multivariate multilevel regression models predicting teacher rating of confidence in service referral.....	60
Table 14: Multinomial logistic regression predicting parent contact or no service vs. mental health service, academic service vs. mental health service, and behavioral intervention vs. mental health service. ....	63

Table 15: Multivariate multilevel regression models predicting number of academic service referrals. ....	65
Table 16: Multivariate multilevel regression models predicting number of behavioral service referrals. ....	67
Table 17: Multivariate multilevel regression models predicting number of mental health service referrals. ....	68
Table 18: Multivariate multilevel regression models measuring mediation of gender effects on teacher referral decision-making.....	71
Table 19: Multivariate multilevel regression models measuring moderation effects of implicit bias as measured by the Implicit Association Test (IAT). ....	76

## List of Figures

Figure 1: Hypothesized teacher mental health decision-making model.....	9
-------------------------------------------------------------------------	---

## Acknowledgements

This project was funded by a Society for Research in Child Development (SRCD) Dissertation Research Funding Award. My graduate studies were also supported by a Duke Endowment Dean's Graduate Fellowship, NIDA Diversity Training Supplement R01 DA016903-07S1, and Duke University Summer Research Fellowship.

This project would not have been possible without the kind and brilliant guidance of my mentor, Dr. Ken Dodge. Ken has, ever so gently, pushed me out of my comfort zone and set me on the course to the career of my dreams. Many thanks to Ken and the rest of my committee, Drs. Laura Richman, Gary Bennett, and Kathryn Whetten for your wisdom and patience throughout this process.

Thank you also to my mentors and guides on this journey: Peggy Morrell and Clarice Montgomery at Duke and Drs. Artemio Brambila and Annette Crisanti at UNM. Special thanks to all the faculty at UNM's Division of Community Behavioral Health, especially Drs. Deb Altschul and Brian Isakson, for holding out hope.

My graduate school journey took some twists and turns and I would not have reached this point without the support of my parents, Sara Marsden and Filiberto Alonso, and my sister Cori Alonso-Yoder. I'm eternally grateful to my Duke crew (Jess, Alissa, Lori, and Sarah), my new New Mexico friends (Kati, Julia, Jess, and Clint), and in PIKE to my "sisters" Helen and Alysa. They have been with me through ups and downs

and I am better for having known them all. There are not enough words of thanks I could give to my future husband, Chase, for his boundless love, energy, encouragement, and most of all sense of humor. Chase, I cannot wait to spend the rest of my life with you and Anna and Nat (and Dany, Chewie, Arya, and all other fur babies we pick up along the way).

# 1. Introduction

Mental health problems are a major concern for children and adolescents. One recent epidemiological study estimated that 82% of children meet diagnostic criteria for at least one psychological disorder by the age of 21 (Copeland, Shanahan, Costello, & Angold, 2011). This estimate includes both well-specified disorders as well as those categorized as “Not Otherwise Specified,” disorders that cause significant life impairment but do not meet full diagnostic criteria for a Diagnostic and Statistical Manual (5th ed.; DSM-5; American Psychiatric Association, 2013) disorder. However, the majority of children suffering from mental illness do not receive treatment (Burns et al., 1995; Farmer, Burns, Phillips, Angold, & Costello, 2003; Leaf et al., 1996). Much of this unmet need is a byproduct of the complex processes by which children arrive at treatment facilities. Specifically, children rarely self-refer for psychiatric treatment, and are instead referred by parents, teachers, and other professionals (Angold et al., 1998; Finkelhor, Wolak, & Berliner, 2001; Srebnik, Cauce, & Baydar, 1996).

Teachers serve as one of the major gatekeepers for psychiatric treatment (Alegría et al., 2012; Hogan, 2003; Wu et al., 1999). They are in a uniquely influential position for identifying abnormal behavior as a result of their routine and lengthy observations of students’ classroom behaviors, academic functioning, and peer interactions (Hogan, 2003). Further, teachers have extensive knowledge of normative child development and are thus able to recognize behavior that is abnormal compared with that of same-age

peers. Among clinical samples, teachers are the initial referral source for up to 87 percent of cases (Costello & Janiszewski, 1990). Unfortunately, teacher education regarding identifying and referring child mental health issues continues to lag behind training in academic skills (Hogan, 2003; Walter, Gouze, & Lim, 2006), and most teachers report low confidence in their ability to manage mental health problems in the classroom (Walter et al., 2006). Without formalized training in diagnosing mental illness, teachers must make referral decisions based on limited colloquial knowledge, which is likely to be biased by pervasive cultural stereotypes. Thus, child demographic variables including race and gender may unduly influence teachers' decisions when identifying student mental illness and making referrals for treatment. The purpose of this study is to understand the cognitive processes through which teachers make decisions regarding child mental health referrals, and in particular how perceived race and gender influence these processes.

### ***1.1 Behavioral Presentations of Mental Illness in Childhood***

Childhood mental health problems are traditionally divided into two categories with distinctive behavioral profiles: internalizing disorders and externalizing disorders. Internalizing problems, including anxiety and depression, are characterized by sadness, tearfulness, nervousness, somatic complaints, and withdrawal from pleasant activities and social interactions (Achenbach & Rescorla, 2001). Internalizing disorders affect an estimated 40% of children under the age of 18 annually (Chen, Killea-Jones, & Vega,

2005). Externalizing disorders, on the other hand, present behaviorally as aggression, irritability, lying, stealing, and disobedience toward authority (Achenbach & Rescorla, 2001). Externalizing disorders, including oppositional defiant disorder (ODD), conduct disorder, and other disruptive behavior disorders, affect an estimated 36% of children per year (Chen et al., 2005).

Although they are frequently discussed as discrete problems, internalizing and externalizing disorders are also highly comorbid. For instance, in a national epidemiological sample, correlations between ODD diagnosis and anxiety and mood disorders ranged from 0.27 to 0.53 (Kessler et al., 2012). Beyond formal comorbidity, overlap in internalizing and externalizing symptomology is also common. For instance, children suffering from anxiety disorders may become oppositional when pressed to engage in anxiety-provoking activities (Mota-Castillo, 2004). Depressed mood frequently presents as irritability in children and adolescents (American Psychiatric Association, 2000). And finally, aggressive behavior in children is often a byproduct of decreased mood, low self-esteem, or interpersonal discomfort (Allen-Meares, 1991; Donnellan, Trzesniewski, Robins, Moffitt, & Caspi, 2005). Children with these behavioral profiles may not meet criteria for a second diagnosis, but they display a mix of internalizing and externalizing symptoms. Overall, children with mental illness frequently display complex clinical pictures with deficits in multiple domains.



## ***1.2 Disparities in Childhood Mental Health Treatment***

As noted above, the majority of children suffering from a mental disorder do not receive treatment. Furthermore, some children are more likely to receive treatment than others. Children with externalizing disorders receive services at higher rates than do children with internalizing disorders (Finkelhor et al., 2001; Gudiño, Lau, Yeh, McCabe, & Hough, 2009; Kazdin & Weisz, 2003; Thompson, 2005). This pattern is generally explained by the visibility of externalizing symptoms, which are easily observed by teachers, parents, and other adults (Cantwell, Lewinsohn, Rohde, & Seeley, 1997). Further, externalizing behaviors such as defiance and aggression can be highly disruptive both at home and at school. Thus, teachers and parents are motivated to seek services for the child in order to reduce conflict and their own distress related to behavior management (Pihlakoski et al., 2004). Children with internalizing disorders, on the other hand, may present with fewer overt symptoms and the distress they experience is largely internal (Cantwell et al., 1997).

In addition, the extant literature suggests that boys are more likely to receive mental health treatment than are girls, despite similar baseline levels of psychiatric problems (Achenbach & Edelbrock, 1981; Achenbach, Howell, Quay, & Conners, 1991). This bias may, in part, reflect the fact that boys are more likely to exhibit externalizing problems, whereas girls are more likely to display internalizing problems (e.g., Angold, Erkanli, Silberg, Eaves, & Costello, 2002; Kessler et al., 2012). Further, boys tend to

experience more severe academic difficulties related to mental illness than do girls (M. T. Green, Clopton, & Pope, 1996). Thus parents and teachers may be especially motivated to seek services for boys to prevent further educational problems. Finally, there is some indication that teachers view girls as more capable of maturing out of psychiatric problems, particularly internalizing problems, without intervention (M. T. Green et al., 1996). The teachers in this study did not express the same level of optimism regarding boys.

Finally, children from different racial backgrounds have unequal rates of unmet psychiatric need. Black children exhibit similar, or perhaps even higher, rates of disorder compared with White peers (Alonso-Marsden & Dodge, 2017; Garland et al., 2005; Roberts, 2000). However, Black children with psychiatric disorders receive treatment at lower rates than their White counterparts (Garland et al., 2005; Yeh, McCabe, Hough, Dupuis, & Hazen, 2003). This outcome is related, in part, to socioeconomic differences and reduced access to treatment for Black families (Alegría, Bijl, Lin, Walters, & Kessler, 2000). However, there is also increasing evidence of disparities in mental health referrals for children of color. Black children exhibiting internalizing disorders are particularly unlikely to be referred for, and to receive, psychiatric treatment (Alegría et al., 2012; Costello & Janiszewski, 1990; Rue & Xie, 2009). This is especially true for Black boys (Alegría et al., 2012; Costello & Janiszewski, 1990).

Moreover, the *type* of services that Black children receive differs from that of their White classmates. Black children are overrepresented in behavioral special education services (Bryan, Day-Vines, Griffin, & Moore-Thomas, 2012; Coutinho & Oswald, 2000; Skiba, Poloni-Staudinger, Gallini, Simmons, & Feggins-Azziz, 2006). Moreover, Black children who exhibit behavior problems are at increased risk for exclusionary discipline (Blake, Butler, Lewis, & Darensbourg, 2011; Butler, Lewis, Moore, & Scott, 2012; Skiba et al., 2011) and arrest and incarceration (Fite, Wynn, & Pardini, 2009; Martin et al., 2011; Martin & Grubb, 1990). However, data from psychiatric clinics are mixed. Some studies suggest an underrepresentation of Black children, including those with externalizing problems (Angold et al., 2002; Erath et al., 2009; Garland et al., 2005). Others suggest overrepresentation of Black children with externalizing disorders (Costello & Janiszewski, 1990; Gudiño et al., 2009). These disparate findings may reflect regional or historical differences in the samples. Further research is necessary to understand service referral differences between Black and White students with psychiatric problems.

### ***1.3 Critique of Previous Research Methods***

A handful of experimental studies has examined the effect of child race and gender on teacher judgments of problem severity and need for services. The goal of these studies has been to disentangle the contribution of a child's actual symptoms from teacher biases in explaining the different rates of identified disorders across race and gender groups. The basic methodology of these studies has utilized hypothetical

vignettes to present identical child behavioral profiles to teachers, manipulating only the gender (M. T. Green et al., 1996; S. P. Green, Shriberg, & Farber, 2008; Kelter & Pope, 2011; Percy, Clopton, & Pope, 1993) or race (Chang & Sue, 2003; Noltemeyer, Kunesh, Hostutler, Frato, & Sarr-Kerman, 2012) of the child. On the whole, these studies find little evidence for bias in teacher assessments of problem severity and referral decisions. However, each study has utilized vignettes that present an obvious clinical picture of solely internalizing or externalizing problems. This type of vignette lacks ecological validity given comorbidity of disorders and symptom overlap as described above. Further, confirmation bias is more likely to occur in response to ambiguous information (Klayman, 1995). Thus, one would not expect significant teacher bias in response to simplistic descriptions of behavior. Rather, bias is more likely to occur in response to complex vignettes that afford selective attention to information that confirms one's existing stereotypes while discounting inconsistent information. More information is needed to understand the cognitive steps by which teachers make referral decisions, particularly in the context of complex behavioral presentations.

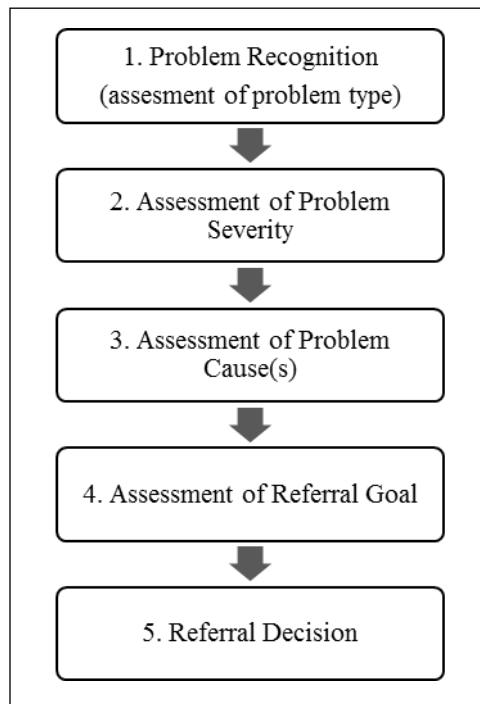
#### **1.4 Current Study**

Previous analyses of teacher mental health decision-making have posited that gender and racial disparities in referral are the result of implicit cognitive bias. However, none have proposed a theoretical model to explain the processes or steps at which bias may impact teacher conceptualization of mental health problems. Instead,

the majority of existing studies use disparities in ultimate referral choice as a proxy measure for bias, with little attention to the cognitive processes preceding this decision. A fuller conceptualization of the cognitive processes that teachers engage in prior to referral will contribute to our basic scientific understanding of mental health decision making by non-clinicians. Moreover, analyzing the effect of gender- and race-based bias on these processes has the potential to inform teacher education programs and other interventions targeted at training teachers to identify mental health symptoms in children. Incorporating training specific to race and gender bias will ultimately help reduce disparities in mental health referral.

#### **1.4.1 Social cognitive model**

As part of this study, I developed and tested a child behavior information processing model, adapted from a social information processing model (Dodge, 2011), that details the manner in which teachers make mental health decisions for their students (see Figure 1). This model posits that teachers engage in a series of mental steps prior to making a decision and ultimately engaging in the action of making a referral for mental health services.



**Figure 1: Hypothesized teacher mental health decision-making model**

The first step in referring a child with mental health concerns for services is recognizing that a problem in fact exists. In naturalistic settings, externalizing behaviors are more likely to be identified due to the overt and observable nature of these problems (Cantwell et al., 1997). However, as described above, children are likely to experience both internalizing and externalizing symptoms. Little is known about how teachers interpret mental health problems in the context of complex behavioral profiles. Existing social cognitive studies suggest that in the context of ambiguous information, individuals selectively attend to information that supports their existing stereotypes while discounting inconsistent information (Gurwitz & Dodge, 1977). Thus, I hypothesized that when a child presents with a variety of mental health symptoms,

teachers will disproportionately attend to those that are consistent with their preconceived notions about child behavior. As a result, boys, and Black boys in particular, will be more likely to be labeled with externalizing problems while girls, and White girls in particular, will be more likely to be labeled as internalizing.

The second step in making a referral is analyzing the seriousness or severity of the presenting problem. There are two competing theories regarding this process in the literature. Some have argued that when behavior is viewed as characteristic of the child's group, then it will be interpreted as normative and therefore non-problematic (Chang & Sue, 2003; Thompson, 2005). Others have suggested that stereotype-consistent behaviors, particularly those exhibited by ethnic minorities, will be over attended to and thus interpreted as more severe than similar behaviors for majority group children (Epstein et al., 2005; Hosterman, DuPaul, & Jitendra, 2008). Some researchers have attempted to test these theories empirically. One such study (Chavez, Shrout, Alegría, Lapatin, & Canino, 2010) manipulated the severity of symptoms in vignettes describing Anglo and Latino children. The authors found that both parent and providers rated behavior as more severe when ascribed to a child with a Latino name than an Anglo name. Another study conducted by Taylor, Gunter, and Slate (2001) found that disruptive behavior was rated as more problematic when depicted by a Black child or a male child than when depicted by a White girl. Thus, the balance of evidence indicates that behaviors will be labeled as more severe if they are consistent with a teacher's

stereotype of the problem that a child is likely to exhibit. This study was able to test these competing theories in the context of complex behavioral presentations and across racial and gender identities.

The third step in referral decision-making involves assessment of precipitating factors. In order to make an appropriate referral for services, the teacher must have some sense of the cause of the child's presenting problem(s). One previous experimental study of teacher referrals also included an assessment of teacher cognition regarding problem etiology (Chang et al., 2003). This study did not find any effects of Black or White race on teacher ratings of problem cause. However, the study did identify a main effect of problem type on this assessment, suggesting that teachers were more likely to attribute externalizing behaviors to transient factors (e.g., neglect, features of the school environment, psychological problems) and more likely to attribute internalizing behaviors to stable influences (e.g., personality, features of the home environment, cultural influences, interpersonal deficits; Chang et al., 2003). Thus, I hypothesized that teachers in this sample would endorse transient causal factors when a child is rated as externalizing and stable etiologic factors when internalizing problems are suspected. Again, this pattern is hypothesized to be most pronounced in situations where the problem domain is consistent with stereotyped expectations.

The fourth step toward a referral decision is assessment of referral goal. Before deciding on a course of action, the teacher must evaluate what the desired outcome of



the referral is, or in other words what he or she would hope to be different following services. As with Step 3, this process is hypothesized to relate to the identified problem domain. That is, the referral goal for externalizing problems will be decreased interpersonal conflict and improved academic performance, and the referral goal for internalizing problems will be improved mood and social interactions.

The fifth and final step of the referral decision process is referral decision. It is hypothesized that ultimate disparities in level and type of service referral is a product of the previous four steps of social cognition: assessment of the child's problem domain, problem severity, problem cause, and service goal. Additionally, the teacher must have confidence in this decision and the effectiveness of the intervention. This is one final point in which implicit bias may interfere with decision-making. Namely, teachers are more likely to have confidence in their referral in situations where the child's behavioral profile is consistent with their stereotyped expectations (Dodge & Newman, 1981). This assessment is crucial for generalizability of this study's findings, as teachers will be more likely to follow through on a referral if they are confident in their decision (Kassirer & Kopelman, 1989).

#### **1.4.2 Bias in teacher implicit cognition of mental health symptoms**

All existing studies of teacher bias in mental health referral have assumed an implicit social cognitive model (Bertrand, Chugh, & Mullainathan, 2005; Greenwald, McGhee, & Schwartz, 1998; Nosek & Riskind, 2012; Payne & Cameron, 2010). In other

words, bias in teacher assessments of behavior is thought to be the result of *implicit* cognitive processes including internalization of cultural stereotypes rather than the result of *explicit* preference for one group of children over another. However, there is a paucity of empirical research testing implicit bias in teachers' behavioral assessments. The current study included an objective measure of implicit cognition (the Implicit Association Test; Greenwald et al., 1998) to analyze individual differences in teachers' implicit associations between child characteristics and behavioral patterns. I hypothesized that teachers would, on average, demonstrate a bias toward associating externalizing symptoms with Black and male children and internalizing symptoms with White and female children. Moreover, I hypothesized that individual differences in the strength of this association would predict bias in teacher assessments of problem type, particularly in response to ambiguous stimuli with equal numbers of internalizing and externalizing symptoms. This bias would, in turn, predict disparities in subsequent levels of the referral decision making process.

Teacher characteristics including ethnicity/race, gender, age, education, and experience were also collected and analyzed as potential moderators of teacher decision-making. However, based on findings from previous studies of bias in mental health identification and referral (Chang & Sue, 2003; Taylor et al., 2001), I hypothesized that teacher demographic characteristics would have minimal impact on the hypothesized cognitive processes.

## **1.5. Study Aims and Hypotheses**

This study significantly contributes to our understanding of how teachers, one of the major gatekeepers of children's mental health services, make decisions regarding identifying problem behaviors and referrals for intervention. This study builds upon previous research on teacher bias in mental health referrals in three major ways: 1) by measuring racial- and gender-based bias at multiple steps of social information processing, 2) utilizing more ecologically valid behavioral stimuli that include a combination of internalizing and externalizing behavior, and 3) including an objective measure of bias (the IAT; Greenwald et al., 1998) to measure the role of implicit stereotypes in teachers' mental health decision-making. This experimental study had eight specific aims:

**Specific Aim #1:** Develop and pilot test a child behavior information processing model to explicate the social cognitive processes by which teachers make decisions regarding referrals (mental health and other) in response to problematic student behavior.

**Specific Aim #2:** Examine the way in which perceived student demographic variables of race and gender influence teacher referral decisions in the context of complex behavioral profiles.

**Specific Aim #2a:** Develop and pilot test behavioral vignettes that represent three types of complex clinical presentations: 1) children with predominantly

externalizing symptoms with some internalizing, 2) children with predominantly internalizing symptoms with some externalizing, and 3) children with an equal number of internalizing and externalizing symptoms.

**Specific Aim #2b:** Analyze the effect of student race (Black versus White) on teacher's assessment of problem type.

*Hypothesis 2b: Teachers will be more likely to apply an externalizing label to a vignette presented with a Black name than to an identical vignette with a White name.*

*Conversely, teachers will be more likely to apply an internalizing label to a vignette with a White name than one with a Black name.*

**Specific Aim #2c:** Analyze the effect of student gender on teacher's assessment of problem type.

*Hypothesis 2c: Teachers will be more likely to apply an externalizing label to a behavioral vignette if it is presented with a male name than the identical behavior presented with a female name. Teachers will be more likely to apply an internalizing label to a vignette with a female name than one with a male name.*

**Specific Aim #2d:** Examine the way in which bias in teachers' assessment of mental health problem type impacts their decision-making at several steps, ultimately resulting in differential referral patterns.

*Hypothesis 2d: Teachers will rate a problem as more severe, more needing of referral, and will have more confidence in the referral when the child's behavior is rated as consistent with a teacher's stereotype of the problem that a child is likely to exhibit.*

**Specific Aim #2e:** Examine main effects of child gender and race on ultimate referral decision.

*Hypothesis 2e: Teachers will be more likely to refer White children for mental health services and Black children for educational services. Girls are less likely than boys overall to be referred for services.*

**Specific Aim 3:** Examine the role of implicit social cognition in racial and gender bias in teacher referral decision-making.

*Hypothesis 3: Overall, teachers will demonstrate a bias toward associating Black children with externalizing symptoms and White children with internalizing symptoms.*

*Similarly they will demonstrate an average bias toward associating males with externalizing symptoms and females with internalizing symptoms. Moreover, individual differences in the strength of these associations will predict bias at each level of the referral decision-making process.*

This study helps explicate the social cognitive processes by which teachers make mental health and other referral decisions. Further, the results of this study aid in our understanding of the nuances of bias in these cognitive processes. Identifying the specifics of how bias impacts referral decisions has the potential to inform targeted

teacher training interventions and policies far more than simply recognizing that such biases exists.

## 2. Method

### 2.1 Study Design

This study utilized an online experimental design to present middle school teachers with vignettes describing student behavior and asking them to respond to a series of questions regarding child behavior information processing and referral decision-making. The vignettes that each teacher received were identical aside from block-randomized names indicative of child race and gender. Utilization of vignettes in an online experimental paradigm provides less ecological validity than direct measures of teacher decision-making in relation to real children in a classroom environment. However, this paradigm allows for in-depth assessment of the psychological processes at play in teacher decision-making, which has been minimally studied up to this point. Further, utilization of social psychological methodology allowed me to control for potentially confounding variables such as child appearance, interaction style, and socioeconomic status in order to isolate the unique effects of gender and racial bias in teacher assessments of mental health problems.

This study consists of a 2 (*gender*) \* 2 (*race*) \* 3(*problem type*: internalizing, externalizing, mixed) design, for a total of 12 conditions. Preliminary pilot testing with a convenience sample of eight college-educated adults indicated that the time to complete all 12 conditions (approximately 90 minutes) was prohibitively burdensome. As a result, the study design was modified to make problem type a between-subjects variable. Thus, teachers were randomly assigned to one of the three behavioral conditions (internalizing, externalizing, mixed) and each received four vignettes representative of the 2 (*gender*) \*

2 (*race*) conditions. This modified design allowed for measurement of the within-subject effects of implicit race and gender bias on decision-making while minimizing participant burden.

A block design was utilized to ensure that each of the 12 experimental vignette conditions was presented first an equal amount of the time to minimize practice effects. Study stimuli and questionnaires were administered through use of the Qualtrics web-based survey tool. This software was also utilized to collect all participant data including item responses and response latencies for the IAT measure.

## ***2.2 Power Considerations and Sample Size Analysis***

As noted above, experimental studies of bias in teachers' behavior ratings have generally yielded null results. Specifically, studies that have analyzed main effects of gender on teacher assessments of behavior severity and need for referral have frequently reported non-significant group differences (Noltemeyer et al., 2012; Percy et al., 1993). Among studies that reported sufficient information for calculating effect sizes (Cohen's *d*), gender effects on ratings of severity ranged from 0 (S. P. Green et al., 2008) to 0.25 (Kelter & Pope, 2011). In the latter case, oppositional behavior was rated as more disruptive when ostensibly exhibited by a girl. Similarly, gender effects on teacher ratings of need for service are generally less than 0.1, with no clear direction favoring males or females (M. T. Green et al., 1996; S. P. Green et al., 2008; Kelter & Pope, 2011; Percy et al., 1993). One exception to this trend was found by Kelter and Pope (2011), who noted a *d* of 0.35 indicating that teachers were significantly more likely to endorse a plan to call a student's parent in response to oppositional behavior of a female student



compared with a male. Thus a small effect size ( $b = 0.2$ ) is assumed for all gender effects in this analysis.

Similarly, effect sizes for racial disparities in teacher ratings of problem severity have generally been in the very small range. Chang and Sue (2003) and Noltemeyer and colleagues (2012) both reported null differences in teacher severity ratings of vignettes associated with Black or White students. Chang and Sue (2003) similarly noted null differences in ratings of need for referral between these two conditions. On the other hand, results of a meta-analysis conducted by Alonso-Marsden and Dodge (2017) indicated a small effect size difference between teacher- and self-report ratings of externalizing behaviors. Specifically, this analysis suggested that, while both types of informants rated Black children higher on externalizing, teachers noted a bigger racial disparity, on the order of  $d = 0.29$  units higher than by youth self-report. Based on the available literature, a small effect size ( $b = 0.2$ ) of race is hypothesized and used to estimate statistical power for the design.

Finally, existing studies have demonstrated a strong effect of externalizing problems (compared with internalizing) on ratings of symptom severity. These effect sizes are in the large range ( $b = 1.12$ , Chavez et al., 2010). Similarly, medium and large effect sizes have also been noted for the effect of externalizing problems on ratings of need for services ( $d = 0.36$ , M. T. Green et al., 1996;  $d = 1.134$ , Pearcy et al., 1993;  $b = -2.08$ , Chavez et al., 2010). Thus a large effect size ( $b = 0.9$ ) is hypothesized for the effect of externalizing vignettes compared with internalizing vignettes.

In order to determine the appropriate sample size for this study, a preliminary power analysis was conducted. This analysis utilized the MLPowSim software (Browne, Golalizadeh Lahi, & Parker, 2009) and the lme4 package for multilevel power analysis (Bates, 2013; Bates, Maechler, & Bolker, 2013) in R version 3.0.1 (R Core Team, 2013). A power analysis was conducted for a repeated measures model with four level-one observations per participant with the following parameters: 80% power,  $\alpha = 0.05$ , and three level-one predictors (gender, race, problem type) with the estimated effects of  $b = 0.2, 0.2, \text{ and } 0.9$ , respectively. Given the non-normal distribution of the outcome measures, a Bernoulli distribution was utilized. Results of this analysis indicated that 1075 participants (level-two) is necessary to achieve sufficient power.

### ***2.3 Recruitment***

During the course of the 2015-2016 academic year, I visited the websites of all public middle schools in the state of North Carolina (including public school districts and public charter schools) in order to create a database of eligible teachers for this study. There was significant variability in the publicly-available information for each school. However, the majority of schools listed teachers' first and last names, subject taught, and email address. Gaps in this information were filled, where possible, with internet searches. Finally, missing email addresses were extrapolated from school district email address conventions (e.g., first initial, last name @county.k12.nc.gov). The resulting database included 12,926 core subject middle school teachers. A randomly-selected sample of 190 teachers was solicited through direct email to participate in the pilot study (described below). Twenty-one of these teachers completed the pilot study, for a response

rate of 11%. Based on the low response rate for the pilot study, all of the remaining 12,736 identified teachers were invited to participate in the final study.

## **2.4 Procedure**

All procedures were reviewed and approved by the Duke University Institutional Review Board. Informed consent was obtained from all study participants via a digital consent form prior to solicitation of data. Since the goal of this study was to measure implicit biases, some deception regarding the intent of the study was necessary. Thus, it was presented to participants as a study of teacher decision-making about mental health with no specific information regarding the independent variables of race, gender, and child problem type.

Demographic self-report data were collected from all participants including age, gender, race/ethnicity, grade taught, years of experience, and educational attainment. School demographics were obtained from publicly available school district records.

Teachers were assigned randomly to one of three between-subject conditions, as *predominantly internalizing*, *predominantly externalizing*, or *mixed*. Then, they were presented with each of the four behavioral vignettes (based on a 2 (*race*) x 2 (*gender*) factorial design) within their assigned condition in randomized order. After reading a vignette, teachers were asked to respond to a series of questions to assess each step of the cognitive processes involved with referral decision-making: problem recognition, assessment of problem severity, assessment of problem cause(s), assessment of referral goal, and referral decision. The questions and response categories are presented in Appendix A.

Upon completion of all vignettes and associated questionnaires, teachers completed a series of Implicit Association Tests (IATs) to measure their automatic associations between mental health categories and child gender and race. All participants completed two IATs: one measuring implicit racialized associations with mental health symptoms and one measuring gendered associations. Following completion of the IAT, teachers were fully debriefed about the intent of the study and entered into a drawing for one of forty \$40 Amazon.com gift cards.

## **2.5 Vignette Development**

Twelve experimental vignettes were developed, each describing a child displaying a particular behavioral profile based on the Teacher's Report Form (TRF; Achenbach, 1991). The symptoms chosen for each vignette were those that are present in both male and female samples, and the total number of symptoms included in each vignette (13) corresponds to a T-score of roughly 67 for both boys and girls on the TRF Internalizing and Externalizing Problem scales. A T-score of 67 represents borderline clinical significance, behavior outside the expected range but not so high as to indicate clear disorder, thus introducing potential variability in response (Achenbach, 1991; M. T. Green et al., 1996). All vignettes included a mixture of internalizing and externalizing symptoms. The *predominantly internalizing* condition described a child with ten internalizing symptoms and three externalizing symptoms. The *predominantly externalizing* condition described a child with ten externalizing symptoms and three internalizing symptoms. The *mixed* condition described a child with seven externalizing symptoms, as these are more frequently endorsed on the TRF (Achenbach, 1991), and six

internalizing symptoms. Vignettes were based on normed behavioral presentations of 12-year-old children. Teachers were asked to imagine each vignette child as middle-school-aged to reduce variability in teacher responding based on perceived child age.

Per field convention, child race and gender were manipulated in the experimental vignettes through the use of normed given names. Specifically, the list of 20 common “Black” and “White” male and female first names identified by Levitt and Dubner (2005) was cross-referenced with a list of the most common baby names for the state of North Carolina between 1999 and 2004 (the birth years for current and recent middle schoolers; Social Security Administration, 2016). Through this procedure, I identified three names to represent each of the four race/gender conditions. Following the procedures of Noltemeyer and colleagues (2012), the 21 pilot study participants rated each of these 12 names. They were asked to indicate the racial group (Black or White) from which a person with that name is likely to be and rate their confidence on a 0-10 Likert scale. The results of this pilot study are described below and summarized in Table 1. Manipulating perceived child race and gender in this manner has been demonstrated to be an effective way of eliciting implicit bias (Bertrand & Mullainathan, 2004; Chavez et al., 2010).

**Table 1: Pilot study ratings of child names**

	Births '99-'04	Ratings of Race/Ethnicity			Child Known:Unknown
		Black	White	Other	
<b>Black Female</b>					
Jada	1,112	14	5	2	18:2
Jasmine	1,664	10	5	6	16:2
<b>Aaliyah</b>	532	18	1	2	12:6
<b>Black Male</b>					
Jalen	810	15	5	2	16:4
<b>Malik</b>	103	19	0	2	12:6
Xavier	820	12	2	7	16:1
<b>White Female</b>					
Emily	3,969	1	16	4	15:1
Katelyn	1,094	0	20	1	17:3
<b>Katie</b>	706	0	21	0	18:2
<b>White Male</b>					
Connor	1,314	0	18	3	17:1
Cody	1,342	2	18	1	16:3
<b>Jack</b>	958	0	20	1	17:3

*Note.* Names in bold were selected for use in the name manipulation. Child Known:Unknown = number of participants who reported having known vs. not having known a child with specified name.

## 2.6 Materials

**Decision-making questionnaires** See Appendix A for survey items and responses.

After reading a vignette, teachers were asked to choose a primary problem from a list of general diagnostic categories. Each teacher chose from a list of 10 descriptors indicating if the child's problem was primarily externalizing or internalizing. Teachers were then allowed to select any additional categories to describe the child's behavior.

Second, teachers rated the overall severity of the problem by response on a 10-point Likert scale to one question: "How severe would you describe this child's problem(s) in terms of his/her day to day functioning in the classroom and elsewhere?" Anchors for the scale included 1 (not at all severe) and 10 (extremely severe).

Third, teachers were asked to indicate the primary source of the child's observed behavior from various potential contributors, including both child-level variables and external factors. The participant was then be able to select any additional causes of this behavior.

Fourth, teachers selected a primary referral from a list of a number of services (decision-making step 5). Participants chose from among psychiatric interventions, punishment, and educational services. In addition, they rated their confidence that this intervention would be successful for the child in the vignette on a 1 (not at all confident) to 10 (extremely confident) scale.

Finally, teachers indicated what they hoped would be different following intervention (decision-making step 4). They chose a primary outcome as well as additional outcomes from a list that included the following items: improved classroom behavior, improved academic performance, improved relationships with teacher and peers, and improved emotional wellbeing.

**Implicit Association Test (IAT)** The goal of this study-specific IAT was to measure the relative strength of associations between children from one racial

background or gender with specific behaviors. Two separate IAT measures were developed for this study. The first of these measures compared strength of associations between Black children with externalizing behaviors and White children with internalizing behaviors. This measure consisted of seven blocks: three practice blocks and four test blocks. In the first practice block, teachers were shown a series of behavior words chosen from the CBCL to represent externalizing and internalizing behavior (e.g., hitting, lying, sad, anxious) and asked to classify internalizing words by pressing the *A* key on the keyboard or externalizing words using the *L* key. In the second block, pictures of children were shown and participants had to classify White children by pressing the *A* key and Black children by pressing the *L* key. In the third block, the first test block, these word and picture categories were paired such that teachers were asked to press the *A* key for internalizing words or White children and the *L* key for externalizing words or Black children. The fourth block, the second test block, was identical to the third block. In the fifth block, only behavior words were categorized but the keys were switched such that the *A* key was paired with externalizing words and *L* key paired with internalizing words. In the sixth block, the third test block, behavior words and pictures were again paired but in the opposite manner: externalizing words or White children were categorized with the *A* key and internalizing words or Black children were categorized with the *L* key. The seventh block, the fourth test block, was identical to the sixth block. Response latencies were recorded by the Qualtrics software



with the addition of Javascript code (Chakroff, 2013). IAT scores were calculated by comparing the response latencies from the two “stereotype consistent” test blocks (blocks 3 and 4) with the response latencies from the two “stereotype inconsistent” test blocks (blocks 6 and 7). Longer latencies indicate more difficulty on the second task and thus stereotype-consistent bias. Raw data were transformed into standardized  $d$  scores using the scoring algorithm proposed by Greenwald, Nosek, and Banaji (2003). Positive  $d$  scores indicate association of Black children with externalizing behaviors and negative scores indicate associating White children with externalizing. Half the sample was randomized to receive the “stereotype consistent” condition first, and half received the “stereotype inconsistent” condition first to minimize order and practice effects.

The IAT for gender bias followed the same pattern described above but with gendered words (Figure 2) in place of photo stimuli. In the first test blocks (blocks three and four) male words were paired with externalizing and females with internalizing. This pattern was reversed for the second test blocks (blocks six and seven). Positive  $d$  scores on this measure indicate association of males with externalizing and negative scores suggest association of females with externalizing. As with the race IAT, half the participants were randomized to receive the “stereotype consistent” condition first and the other half received the “stereotype inconsistent” condition first. The sequence of the race IAT and gender IAT was randomly varied across the sample to avoid order effects

such that half the participants completed the race IAT first and half completed the gender IAT first.

## **2.7 Pilot Study**

Twenty-one teachers randomly selected from the larger sample participated in the pilot study which had several aims. First, the pilot measured the amount of time required to complete the abbreviated protocol which included a demographic questionnaire, four behavioral vignettes (representing the 2 (*race*) × 2 (*gender*) conditions), and two IAT measures (one for race and one for gender). The shortened protocol took approximately 30 minutes to complete.

Second, the pilot questionnaire included a manipulation/deception check to determine whether the teachers attended to the name stimuli. Unfortunately, this question was asked after participants had completed the IAT tasks, and as such the deception was at least partially revealed prior to debriefing. This structure was necessary in order to obtain an accurate measure of the time to complete the protocol. Six teachers reported noticing that the child names were associated with specific racial and/or gender groups and three of these teachers also stated they noticed that the vignettes described similar behavioral profiles, and thus that they correctly identified the manipulation. Each of these three teachers reported identifying the naming pattern over the course of the survey as new child vignettes were presented. The remaining teachers reported that they did not surmise a “hidden” research question. As such, the

manipulation and deception were judged to be appropriate for the purposes of this study. However, a separate set of between-subject analyses was also conducted for each hypothesis using only the first vignette in order to control for the observed practice effect.

Third, the pilot study solicited feedback regarding the answer choices for each information processing question. Based on teacher feedback, several response choices were added or removed from the survey. These changes are summarized in Appendix B2.

Fourth, pilot study participants were asked to rate the three names selected to represent each of the four identity conditions (Black female, Black male, White female, White male). Responses to these items are summarized in Table 1. Based on these responses, the names selected for the final survey were Aaliyah (Black female), Malik (Black male), Katie (White female), and Jack (White male).

## ***2.8 Analysis Plan***

Analyses were conducted with SAS 9.4. Multilevel modeling was utilized to analyze each dependent variable and the relations among variables, with vignette at the first level and teacher at the second level. Vignette-level independent variables include child race and gender and dependent variables were the child behavior information-processing decisions made by a teacher. Teacher-level variables included implicit associations, teacher demographics, and school district characteristics.

As suggested by Hox (2010), all within subject models were systematically built, beginning with a baseline model of the teacher-level random intercept. This initial model was used to estimate teacher-level variation in each of the dependent variables. Teacher-level covariates (teacher age, gender, race/ethnicity, grade taught, years of experience, educational attainment, school demographics) were then added and included in all models. Some of these variables, including teacher race/ethnicity, years of teaching experience, and school demographics, were also tested as moderators of the proposed analyses. Once this covariate model was established, regression analyses were added to the model to test the individual hypotheses. Due to the large number of tests to be performed, family-wise error was controlled using a Bonferroni correction such that  $\alpha = .05$  was divided by the number of tests performed for each predictor variable. Table 2 summarizes the resulting significance level for each predictor utilized in this study.

In order to control for carryover effects, the first analysis was a between-subject model estimating the effects of the independent variable(s) on the participants' response to the first vignette they completed. All analyses controlled for teacher demographics (race, gender, age, education) and school district demographics (percentage of Black/African American students).

**Table 2: Bonferroni correction for family-wise error: p-value used for significance testing.**

Predictor	Number of comparisons	Significant <i>p</i> -value
Child Race	11	<.0045
Child Gender	11	<.0045
Child Race x Gender	9	<.0056
Child Race x Primary Problem	4	<.0125
Child Gender x Primary Problem	4	<.0125
Child Race x Condition	2	<0.025
Child Gender x Condition	2	<0.025

## 3. Results

### 3.1 Sample Characteristics

Of the 12,926 middle school teachers invited to participate, 1106 teachers provided complete responses for at least one vignette, for a response rate of 8.6%. Participants were 82.3% female and an average of 40.5 years old ( $SD = 10.5$ ). Participants were allowed to select multiple racial/ethnic categories; 85.9% indicated they were White, 12.7% endorsed Black/African American, 2.4% selected American Indian or Alaska Native, 1.6% were Hispanic/Latino/a, 0.7% were Asian, 0.2% were Native Hawaiian/Pacific Islander, and 1.5% reported other race/ethnicity including multiracial. Teachers reported an average of 12.6 years of teaching experience ( $SD = 7.8$ ), 48.9% had a 4-year bachelor's degree, 46.7% had a master's degree, and 1.5% had a doctoral or professional degree (e.g., Ed.D., Ph.D., J.D.).

The target population for this survey was current and recent public and charter middle school core subject teachers, and the recruitment methods were largely successful in meeting this goal. The majority of participants (97.6%) were current middle school teachers while 2.3% were currently teaching elementary and 4.2% teaching high school. Most (92.3%) taught in public schools, 7.1% taught in a charter school, and 0.5% taught in another school environments (e.g., alternative schools). In terms of classroom environment, 80.5% of participants were general education core subject teachers, 6.2% taught advanced/academically or intellectually gifted (AIG) classes, 3.3% taught

exceptional children (special education) classes, 0.6% taught enrichment/electives, and 9.4% reported “other” teaching environment including some combination of the above. Participants represented 99 county school districts and 47 charter schools.

Table 3 presents cell means for within-subject responses on several dependent variables in this analysis divided by vignette child race and gender and participant condition. Inferential statistical tests of significance are presented in the following sections.

**Table 3: Observed dependent variable means for each 2 (gender) x 2 (race) x 3 (behavioral condition) category.**

Trial	Number of Internalizing Problems			Trial	Number of Externalizing Problems		
	P Int <i>M (SD)</i>	Mixed <i>M (SD)</i>	P Ext <i>M (SD)</i>		P Int <i>M (SD)</i>	Mixed <i>M (SD)</i>	P Ext <i>M (SD)</i>
Black Female	1.40 (0.94)	1.29 (1.00)	1.21 (0.96)	Black Female	0.71 (0.85)	1.01 (0.90)	1.41 (1.02)
Black Male	1.44 (0.99)	1.28 (0.91)	1.20 (1.01)	Black Male	0.83 (0.94)	1.21 (1.08)	1.60 (1.05)
White Female	1.55 (0.96)	1.37 (0.96)	1.25 (0.96)	White Female	0.65 (0.85)	1.05 (0.99)	1.37 (1.05)
White Male	1.47 (1.03)	1.17 (0.92)	1.19 (1.04)	White Male	0.80 (0.89)	1.15 (1.00)	1.55 (1.07)

Trial	Severity Rating			Trial	Number of MH Referrals		
	P Int <i>M (SD)</i>	Mixed <i>M (SD)</i>	P Ext <i>M (SD)</i>		P Int <i>M (SD)</i>	Mixed <i>M (SD)</i>	P Ext <i>M (SD)</i>
Black Female	7.12 (1.45)	7.24 (1.44)	7.84 (1.26)	Black Female	0.71 (0.77)	0.69 (0.76)	0.90 (0.79)
Black Male	7.31 (1.44)	7.43 (1.45)	8.01 (1.32)	Black Male	0.78 (0.80)	0.72 (0.76)	0.90 (0.82)
White Female	7.03 (1.55)	7.23 (1.46)	7.78 (1.44)	White Female	0.81 (0.87)	0.74 (0.75)	0.93 (0.83)
White Male	7.39 (1.39)	7.44 (1.31)	7.86 (1.41)	White Male	0.84 (0.87)	0.81 (0.84)	0.91 (0.82)

Note. P Int = predominantly internalizing condition; P Ext = predominantly externalizing condition; Mixed = mixed condition; MH = mental health



### **3.2 Effect of Perceived Student Race on Problem Type**

*Hypothesis 2b: Teachers will be more likely to apply an externalizing label to a vignette presented with a Black name than to an identical vignette with a White name. Conversely, teachers will be more likely to apply an internalizing label to a vignette with a White name than one with a Black name.*

This first between-subject analysis looks at the effects of perceived child race on teacher assessment of primary problem type (internalizing, externalizing, or no problem). For the sake of these analyses, problem type was categorized as externalizing (*anger, attention difficulties, hyperactivity, and impulsivity*), internalizing (*anxiety/worry, sadness/depression, shyness, and social withdrawal*), and no problem (*none/age appropriate behavior*). The total sample of 1106 teachers responded to this first question in response to the first vignette. Unfortunately, 13 first vignette responses were eliminated due to typo on one story in the mixed condition in which a child was erroneously referred to as “Katie” in the vignette after being introduced as “Aaliyah”. This error was fixed as soon as it was discovered and affected a total of 53 responses. The missing data associated with this error occurs completely at random since participants were randomized to the affected condition(s). As such, all responses to the erroneous Aaliyah (Black female) vignettes were deleted from the data. A dummy variable was created to identify participants who received this nonconforming vignette at any point in the procedure. This dummy variable is included as a control variable in all within subject analyses.

SAS Proc CATMOD was utilized to measure the between subjects effect of vignette child race on the multinomial outcome variable of primary problem type ( $N = 1093$ ). A preliminary set of univariate analyses was conducted to measure the effect of the control variables (teacher demographics, school demographics, behavioral condition) on the categorical outcome of primary problem type. Covariates with significant main effects ( $p < .05$ ) on the outcome variable (teacher age, race, gender, and education, school district percentage of Black students) were kept in the model. Those that were not significant (years of experience, grade taught, special education classroom) were removed.

As anticipated, the randomly assigned behavioral condition had a significant main effect on teacher selection of child primary problem type. Specifically, participants in the *predominantly externalizing* condition were significantly less likely to select an internalizing versus externalizing problem type for the child in the first vignette ( $OR = 1.27, \chi^2(2) = 65.21, p < .0001$ ). The difference between selecting an externalizing problem and no problem was nonsignificant for this condition ( $OR = .72, \chi^2(2) = 1.01, p = .315$ ). Those assigned to the *predominantly internalizing* group were significantly more likely to select an internalizing problem ( $OR = 2.37, \chi^2(2) = 67.68, p < .0001$ ) or no problem ( $OR = 2.00, \chi^2(2) = 6.96, p = .008$ ), compared with an externalizing problem. Thus it appears that teachers were appropriately attending to the first vignette.

Table 4 summarizes the observed rates of vignettes with Black and White child names that were categorized as having an internalizing, externalizing, or no primary problem. Participant responses are similar across groups, with 32.62% of Black vignettes labeled externalizing (versus 32.15% of White vignettes), 66.67% labeled internalizing (versus 66.36%) and very few (0.72%) categorized as “no problem” (versus 1.50% of White vignettes). The CATMOD procedure was used to test the significance of the observed response pattern. Results of this multivariate analysis indicate that, controlling for teacher and school demographics, perceived child race was not a significant predictor of between-subject variability in selection of primary problem type. Specifically, odds of selecting an externalizing problem for a Black child were not significantly greater than odds of selecting an externalizing problem for a White child. This pattern is true for selection of internalizing primary problem and for selecting no problem as well.

**Table 4: Multinomial logistic regression predicting externalizing vs. internalizing primary problem type and externalizing primary problem type vs. no problem**

Model 1	Externalizing (n = 354) n (%)	Internalizing (n = 727) n (%)	Odds ratio, internalizing <sup>a</sup> (SE)	$\chi^2$	<i>p</i>	No problem (n = 12) n (%)	Odds ratio, no problem <sup>a</sup> (SE)	$\chi^2$	<i>p</i>
Child Race Black White (ref)	182 (32.62) 172 (32.15)	372 (66.67) 355 (66.36)	-0.06 (0.16)	0.17	.681	4 (0.72) 8 (1.50)	-1.10 (0.72)	2.33	.127
Condition									
P Externalizing	228 (61.46)	139 (37.47)	-1.24 (0.16)	57.37	<.0001	4 (1.08)	0.08 (0.90)	0.01	.928
P Internalizing	111 (31.18)	241 (67.70)	2.45 (0.31)	63.44	<.0001	4 (1.12)	2.99 (0.94)	10.19	.001
Mixed (ref)	15 (4.10)	347 (94.81)				4 (1.09)			
Teacher Race									
Black	53 (37.59)	84 (59.57)	-0.33 (0.23)	2.01	.156	4 (2.84)	0.99 (0.71)	1.97	.161
Other	14 (43.75)	18 (56.25)	-0.53 (0.45)	1.38	.240	0 (0.00)	*	*	*
White (ref)	287 (31.26)	623 (67.86)				8 (0.87)			
Teacher Male	67 (34.36)	128 (65.64)	-0.04 (0.20)	0.04	.841	0 (0.00)	*	*	*
Female (ref)	287 (32.00)	598 (66.67)				12 (1.34)			
Education Master's+	159 (29.01)	384 (70.07)	0.31 (0.16)	3.61	.057	5 (0.91)	-0.17 (0.68)	0.06	.801
Bachelor's (ref)	195 (25.78)	343 (62.94)				7 (1.28)			
	<i>M (SD)</i>	<i>M (SD)</i>				<i>M (SD)</i>			
Teacher Age	39.99 (10.07)	40.63 (10.32)	0.01 (0.01)	0.59	.443	47.7 (10.07)	0.09 (0.04)	5.98	.014
School Dist % Black	27.16 (17.69)	26.62 (18.28)	0.292 (0.45)	0.43	.513	41.06 (16.28)	3.56 (1.49)	5.74	.017
Model 2									
Child Race * P Externalizing			-0.28 (0.33)	0.74	.389		*	*	*
Child Race * P Internalizing			-0.44 (0.57)	0.62	.432		*	*	*

*Note.* Model 2 includes all main effects in Model 1 with added interaction effects. *SE* = standard error; P Externalizing = predominantly externalizing; P Internalizing = predominantly internalizing; ref = reference group for dummy coded multivariate analysis; Education Master's+ = teacher has master's degree or above; School Dist % Black = percentage of Black students in school district.

<sup>a</sup>Externalizing primary problem type is the reference category. \*No estimate due to small cell size.

In addition to a main effect of vignette child race on primary problem type, I hypothesized an interaction with the randomized behavioral condition. As displayed in Table 4, no significant interaction was observed. Likewise, the association between child race and primary problem type was not significantly moderated by teacher race, gender, or age.

In addition to selecting a primary problem type, teachers could also choose any other presenting problems from the same list of behavior categories described above. “Yes” responses to these items were summed along with the primary problem type to create a count variable of the number of internalizing and externalizing problems. Responses on these two count variables were sufficiently normally distributed for multiple regression analysis. Group means for the number of internalizing and externalizing problems endorsed are presented in Table 5. The mean number of internalizing problems selected for vignettes with a Black name ( $M = 1.97, SD = 1.02$ ) were similar to those for White name vignettes ( $M = 1.98, SD = 1.09$ ). Likewise the mean number of externalizing problems was similar for Black name vignettes ( $M = 1.08, SD = 1.02$ ) and White name vignettes ( $M = 1.00, SD = 1.01$ ). Controlling for teacher demographics, vignette child race was not a significant predictor of number of internalizing or externalizing symptoms.

**Table 5: Multiple regression results predicting number of internalizing and externalizing problems.**

Variable	Number of Internalizing Problems				Number of Externalizing Problems			
	M (SD)	Model 1 B	Model 2		M (SD)	Model 1 B	Model 2	
			B	95% CI			B	95% CI
Constant		1.95***	1.96***	[1.84, 2.08]		1.52***		
Condition								
P Externalizing	1.46 (0.92)	-0.4***	-0.47***	[-0.61, -0.33]	1.44 (1.04)	0.32***	0.32***	[0.18, 0.45]
P Internalizing	2.55 (0.96)	0.62***	0.62***	[0.48, 0.76]	0.55 (0.78)	-0.57***	-0.57***	[-0.71, -0.43]
Mixed (ref)	2.55 (0.96)				1.12 (0.99)			
Teacher Race								
Black	1.80 (1.03)	-0.18*	-0.18*	[-0.36, -0.01]				
Other	2.00 (0.88)	0.10	0.10	[-0.24, 0.44]				
White (ref)	2.00 (1.06)							
Teacher Age						-0.01***	-3.55	[-0.02, -0.00]
Child Race Black	1.97 (1.02)		-0.03	[-0.14, 0.09]	1.08 (1.02)		0.08	[-0.03, 0.19]
White (ref)	1.98 (1.09)				1.00 (1.01)	.14		
R <sup>2</sup>		.19	.19				0.15	
ΔR <sup>2</sup>			.00				-.01	

Note. CI = confidence interval; P Externalizing = predominantly externalizing; P Internalizing = predominantly internalizing; ref = reference group for dummy coded multivariate analysis.

\* family-wise  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

### 3.2.1 Multilevel Analysis

Prior to conducting the multilevel analysis, an analysis of missing data patterns was conducted. Overall, 450 observations were missing across the 1106 (participants)  $\times$  4 (observations per participant) expected, for a total missing data proportion of 450/4424 or 10.2%. These data are assumed to be missing at random, or in other words the missingness is correlated with participant-level variables measured in this analysis (Hox, 1999). The estimator used for all within subject analyses (Restricted Maximum Likelihood) is robust to this type of missing data (Hox, 1999). However, I also conducted a preliminary analysis of bivariate predictors of missingness. Results of this analysis indicated that the participant-level variables of special education status and behavioral condition significantly predicted missing data. Specifically, special education teachers were significantly more likely to have missing data than their general education/AIG counterparts ( $OR = 0.61$ ,  $\chi^2(1) = 6.88$ ,  $p = .009$ ). In addition, participants randomized to the mixed condition were significantly less likely to have missing data than those randomized to the predominantly externalizing ( $OR = 0.72$ ,  $\chi^2(1) = 31.89$ ,  $p < .0001$ ) or predominantly internalizing ( $OR = 0.27$ ,  $\chi^2(1) = 5.35$ ,  $p = .021$ ) conditions. As such, these two covariates are included in all within subject analyses to aid in accurate estimation of model parameters.

In the first step of the multilevel analysis, the intercept-only model was tested in order to obtain the intraclass correlation. Results of this analysis indicated that 9% of the

variability in externalizing classification was at the within-subject level ( $ICC = 0.093$ ), suggesting multilevel analysis is indeed appropriate to account for nesting. In this analysis, teachers are also nested within schools, implying a three-level structure. However, results of the intercept-only model indicate that very little additional variability in externalizing classification is explained by nesting within school districts ( $ICC = 0.001$ ). Thus, all subsequent models controlled for the fixed effect of level-three nesting, but no analyses attempt to predict variability in response at the school district level.

In the next step of the multilevel analysis, each control variable was tested independently to estimate its unique predictive value, controlling for within subject effects. PROC GLIMMIX does not allow for significance testing of random effects, nor for model comparisons based on fit indices. As such, the decision to keep random effects in the model was based on improved estimates for fixed effects. A saturated model with improved model fit was created (Table 6) stepwise by adding significant predictors one at a time. Once again, controlling for behavioral condition, teacher demographics, and within subject variability, vignette child race was not a significant predictor of externalizing primary problem type. Teacher race and behavioral condition did not significantly moderate this relationship.



**Table 6: Multivariate multilevel logit models predicting externalizing problem type**

Model 1	Estimates of Fixed Effects		
	Odds ratio (SE)	<i>t</i>	<i>p</i>
Child Race Black	-0.01 (0.02)	-0.66	.512
Condition			
P Externalizing	0.18 (0.02)	10.18	<.0001
P Internalizing	-0.23 (0.02)	-12.67	<.0001
Mixed (ref)			
Teacher Race			
Black	0.05 (0.03)	1.64	.102
Other	0.00 (0.06)	0.04	.969
White (ref)			
<hr/>			
Model 2			
Child Race x P Externalizing	0.02 (0.04)	0.47	.639
Child Race x P Internalizing	0.05 (0.08)	0.64	.522

*Note.* Model 2 includes all main effects in Model 1 with added interaction effects. *SE* = standard error; P Externalizing = predominantly externalizing; P Internalizing = predominantly internalizing; ref = reference group for dummy coded multivariate analysis.

### **3.3 Effect of Perceived Student Gender on Problem Type**

*Hypothesis 2c: Teachers will be more likely to apply an externalizing label to a behavioral vignette if it is presented with a male name than the identical behavior presented with a female name. Teachers will be more likely to apply an internalizing label to a vignette with a female name than one with a male name.*

The first step in testing Hypothesis 2c was a between-subject analysis of the effect of vignette child gender on the multinomial outcome of primary problem type. As

described above, this outcome is categorized into three groups: externalizing primary problem, internalizing primary problem, and no problem. Results of the PROC CATMOD multinomial analysis are presented in Table 7. Participant responses are similar across groups, with 30.63% of male vignettes labeled externalizing (versus 34.29% of Female vignettes), 68.66% of males labeled internalizing (versus 64.19% of females) and very few (0.70% of males) categorized as “no problem” (versus 1.52% of female vignettes). Controlling for teacher and school district demographics, vignette child gender was not a significant predictor of primary problem type.

In addition to main effect of gender, I hypothesized an interaction between child gender and race. The observed categorization of primary problem for each race/gender condition is presented in Table 7. The observed proportions vary somewhat across trial, with Black male children less likely to be assigned an externalizing primary problem (29.15%), and more likely to receive an internalizing primary problem (70.85%) compared with other trials. No Black males were assigned “no problem.” However, accounting for teacher and school district demographics, this race x gender interaction was not significant ( $OR = 0.26$ ,  $\chi^2(1) = 0.69$ ,  $p = .405$ ). Similarly, the interaction between vignette child gender and behavioral condition was non-significant.

**Table 7: Multinomial logistic regression results predicting externalizing vs. internalizing primary problem type and externalizing primary problem type vs. no problem.**

Model 1	Externalizing (n = 354) n (%)	Internalizing (n = 727) n (%)	Odds ratio, internalizing <sup>a</sup> (SE)	$\chi^2$	<i>p</i>	No problem (n = 12) n (%)	Odds ratio, no problem <sup>a</sup> (SE)	$\chi^2$	<i>p</i>
Child Gender Male	174 (30.63)	390 (68.66)	-0.04 (0.20)	0.04	.844	4 (0.70)	*	*	*
Female (ref)	180 (34.29)	337 (64.19)				8 (1.52)			
<hr/>									
Model 2									
Child Gender * Child Race	N/A	N/A	0.26 (0.31)	0.69	.405	N/A	*	*	*
Black Female	96 (36.50)	163 (61.98)				4 (1.52)			
Black Male	86 (29.15)	209 (70.85)				0 (0.00)			
White Female	84 (32.06)	174 (66.41)				4 (1.53)			
White Male	88 (32.23)	181 (66.30)				4 (1.47)			
<hr/>									
Model 3									
Child Gender x P Externalizing			-0.11 (0.33)	0.11	.745		*	*	*
Child Gender x P Internalizing			-0.58 (0.63)	0.86	.353		*	*	*

*Note.* Model 1 accounts for the following covariates (presented in Table 4): behavioral condition; vignette child race; teacher race, gender, education, and age; school district percentage of Black students. Model 2 includes all main effects in Model 1 with added interaction effects. *SE* = standard error; ref = reference group for dummy coded multivariate analysis; P Externalizing = predominantly externalizing condition; P Internalizing = predominantly internalizing condition; N/A = Not applicable.

<sup>a</sup>Externalizing primary problem type is the reference category. \*No estimate due to small cell size.

Next, Hypothesis 2c was tested with a multiple regression analysis measuring the effect of vignette child gender on number of internalizing and externalizing symptoms. Results of this analysis are presented in Table 8. The observed mean number of internalizing problems was similar for male vignettes ( $M = 1.97, SD = 1.02$ ) and female vignettes ( $M = 1.98, SD = 1.09$ ). Controlling for behavioral condition and teacher demographics, there was no significant main effect of vignette child gender on internalizing symptoms. However, the mean number of externalizing problems endorsed was higher for male vignettes ( $M = 1.10, SD = 1.06$ ) than for female vignettes ( $M = 0.98, SD = 0.96$ ). There was a significant main effect of vignette child gender on externalizing symptoms ( $r^2 = 0.06, t = 2.70, p = .007$ ). However, the interaction between vignette race and gender was not a significant predictor of number of internalizing or externalizing problems.

**Table 8: Multiple regression results predicting number of internalizing and externalizing problems.**

Variable	Number of Internalizing Problems					
	<i>M (SD)</i>	Model 2			Model 3	
		Model 1 <i>B</i>	<i>B</i>	95% CI	<i>B</i>	95% CI
Constant		1.95***	1.95***	[1.83, 2.07]	2.01***	[1.86, 2.15]
Condition						
P Externalizing	1.46 (0.92)	-0.4***	-0.47***	[-0.61, -0.33]	-0.47***	[-0.61, -0.33]
P Internalizing	2.55 (0.96)	0.62***	0.62***	[0.48, 0.76]	0.62***	[0.48, 0.76]
Mixed (ref)	2.55 (0.96)					
Teacher Race						
Black	1.80 (1.03)	-0.18*	-0.18*	[-0.36, -0.01]	-0.18*	[-0.36, 0.01]
Other	2.00 (0.88)	0.10	0.10	[-0.24, 0.44]	0.10	[-0.23, 0.44]
White (ref)	2.00 (1.06)					
Child Gender Male	1.97 (1.02)		-0.004	[-0.14, 0.09]	-0.09	[-0.26, 0.07]
Female (ref)	1.98 (1.09)					
Child Race Black					-0.12	[-0.28, 0.05]
Child Gender x Child Race					0.18	[-0.05, 0.41]
Race						
Black Female	1.90 (1.93)					
Black Male	2.03 (1.01)					
White Female	2.02 (1.07)					
White Male	1.95 (1.11)					
<i>R</i> <sup>2</sup>		.19	.19		.19	
$\Delta R^2$			.00		.00	

Number of Externalizing Problems						
Variable	<i>M (SD)</i>	Model 2			Model 3	
		Model 1 <i>B</i>	<i>B</i>	95% CI	<i>B</i>	95% CI
Constant		1.52***	1.43***	[1.19, 1.68]	1.39***	[1.13, 1.64]
Condition						
P Externalizing	1.44 (1.04)	0.32***	0.33***	[0.19, 0.46]	0.33***	[0.19, 0.46]
P Internalizing	0.55 (0.78)	-0.57***	-0.57***	[-0.70, -0.43]	-0.57***	[-0.71, -0.43]
Mixed (ref)	1.12 (0.99)					
Teacher Age		-0.01***	-0.01***	[-0.02, -0.00]	-0.01***	[-0.02, -0.00]
Child Gender Male	1.10 (1.06)		0.15	[0.04, 0.26]	0.15	[-0.00, 0.31]
Female (ref)	0.98 (0.96)					
Child Race Black					0.08	[-0.08, 0.24]
Child Gender x Child Race					-0.005	[-0.23, 0.22]
Race						
Black Female	1.02 (0.99)					
Black Male	1.13 (1.05)					
White Female	0.94 (0.93)					
White Male	1.06 (1.08)					
<i>R</i> <sup>2</sup>		.14	.15		.15	
$\Delta R^2$			.01		.00	

CI = confidence interval; P Externalizing = predominantly externalizing; P Internalizing = predominantly internalizing; ref = reference group for dummy coded multivariate analysis.

\* family-wise  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

### 3.3.1 Multilevel Analysis

Based on the above between-subjects results, I chose to conduct a multilevel analysis looking at the within subject effects of vignette on number of externalizing problems. The first step in this analysis is an intercept-only model to estimate the ICC. This analysis indicated that 25% of variability in number of externalizing problems assigned was at the within-subject level (ICC = 0.25). Again, a negligible amount of variability was explained by the school district level above participant-level effects (ICC = 0.002). Thus, fixed and random effects are modeled for participants and only fixed effects of school district.

As described above, a saturated model was then built by adding covariates on at a time. Variables were retained in the model if they were significant at the  $p < .05$  level or added to model fit. The final model is presented in Table 9. Results of this multilevel regression suggest that, accounting for within-subject effects and school district demographics, vignette child gender was a significant predictor of number of externalizing problems ( $r^2 = 0.15$ ,  $t = 5.89$ ,  $p < .0001$ ). This main effect was not significantly moderated by behavioral condition or vignette child race.

**Table 9: Multivariate multilevel regression models predicting number of externalizing problems.**

Model 1	Estimates of Fixed Effects		
	$r^2$ (SE)	$t$	$p$
Child Gender Male Condition	0.15 (0.03)	5.89	<.0001
P Externalizing	0.35 (0.05)	6.49	<.0001
P Internalizing	-0.38 (0.06)	-6.97	<.0001
Mixed (ref)			
School Dist % Black	-0.05 (0.12)	-0.42	.676
<hr/>			
Model 2			
Child Gender x P Externalizing	0.01 (0.07)	0.17	.864
Child Gender x P Internalizing	-0.03 (0.07)	-0.42	.673
<hr/>			
Model 3			
Child Gender x Child Race	0.03 (0.05)	0.54	.587

*Note.* Model 2 and Model 3 include all main effects in Model 1 with added interaction effects. *SE* = standard error; P Externalizing = predominantly externalizing condition; P Internalizing = predominantly internalizing condition; ref = reference group for dummy coded multivariate analysis.

### **3.4 Predictors of Ratings of Problem Severity, Referral, and Confidence**

*Hypothesis 2d: Teachers will rate a problem as more severe, more needing of referral, and will have more confidence in the referral when the child's behavior is rated as consistent with a teacher's stereotype of the problem that a child is likely to exhibit.*



### 3.4.1 Predictors of Teacher Ratings of Problem Severity

The first set of analyses to test Hypothesis 2d models variation in the continuous outcome variable of problem severity. Teachers responded to this question with a slider button that allowed them to assess severity on a 0 to 10 scale measured to one decimal place (e.g., 6.4 and 6.5 are distinct possible responses). Due to this measurement, responses are sufficiently normally distributed for regression analysis.

As with previous analyses, the first step in modeling teacher assessment of problem severity is a between-subject analysis of responses to the first vignette presented. Results of this multivariate regression model are presented in Table 10. Teachers rated males slightly higher ( $M = 7.61$ ,  $SD = 1.25$ ) than females ( $M = 7.46$ ,  $SD = 1.36$ ) on severity. Controlling for significant teacher demographic covariates, this association was marginally significant ( $r^2 = 0.18$ ,  $t = 2.30$ ,  $p = .022$ ) at the Bonferroni family-wise adjusted significance level of  $p < .0045$ . Ratings of severity were similar for vignettes with a Black name ( $M = 7.56$ ,  $SD = 1.27$ ) and those with a White name ( $M = 7.52$ ,  $SD = 1.34$ ). Controlling for teacher demographics, vignette child race was not a significant predictor of teacher severity rating. Likewise, the interaction between vignette child race and gender was not significant.

**Table 10: Multiple regression results predicting teacher rating of problem severity.**

Problem Severity Rating				
Model 1	<i>M (SD)</i>	<i>r</i> <sup>2</sup> ( <i>SE</i> )	<i>t</i>	<i>p</i>
Child Gender Male	7.61 (1.25)	0.18 (0.08)	2.30	.022
Female (ref)	7.46 (1.36)			
Condition				
P Externalizing	7.93 (1.19)	0.55 (0.09)	5.81	<.0001
P Internalizing	7.27 (1.41)	-0.13 (0.09)	-1.35	.178
Mixed (ref)	7.39 (1.21)			
Special Education Class	7.20 (1.61)	-0.41 (0.14)	-2.89	.004
General/AIG Class (ref)	7.57 (1.27)			
	<i>R</i>			
Years Teaching Experience	0.92	0.02 (0.00)	3.26	.001
Model 2	<i>M (SD)</i>	<i>r</i> <sup>2</sup> ( <i>SE</i> )	<i>t</i>	<i>p</i>
Child Race Black	7.56 (1.27)	0.06 (0.08)	0.79	.430
White (ref)	7.52 (1.34)			
Model 3				
Child Gender x Child Race		0.13 (0.15)	0.84	.401
Black Female	7.46 (1.32)			
Black Male	7.64 (1.23)			
White Female	7.45 (1.41)			
White Male	7.58 (1.27)			
Model 4				
Externalizing Problem	7.83 (1.35)	0.29 (0.10)	2.10	.036
Internalizing Problem (ref)	7.39 (1.35)			
Model 5				
Child Gender x Ext Problem		-0.03 (0.16)	-0.19	.847
Model 6				
Child Race x Ext Problem		-0.20 (0.16)	-1.22	.222

*Note.* Models 2 through 6 include all covariates in Model 1 with added main effects and interactions. *SE* = standard error; P Externalizing = predominantly externalizing; P Internalizing = predominantly internalizing; ref = reference group for dummy coded multivariate analysis; Ext Problem = externalizing primary problem.

I had also hypothesized that main effects of race and gender on teacher assessment of problem severity would be moderated by their assessment of primary problem type, such that stereotype consistent behavior would be rated as more severe. To test this hypothesis, a bivariate marker for externalizing primary problem type was added to the regression model. As shown in Table 10, this variable had a significant main effect on ratings of problem severity ( $r^2 = 0.29$ ,  $t = 2.10$ ,  $p = .036$ ). Specifically, children labeled with an externalizing primary problem were rated more severe ( $M = 7.83$ ,  $SD = 1.35$ ) than those labeled internalizing or no problem ( $M = 7.39$ ,  $SD = 1.35$ ). However, primary problem type did not significantly moderate the effect of gender or race on ratings of problem severity.

### **3.4.2 Multilevel Analysis**

As detailed above, the first step in building a multivariate multilevel model is establishing the baseline intercept-only model. Results of this first analysis indicated a high intraclass correlation of 0.69, suggesting that 69% of variation in ratings of severity are at the within-subject level. Stated another way, this result indicates that individual teachers provided very similar ratings of problem severity across conditions. As such, a multilevel model is necessary to account for this within-subject effect.

Next a saturated multivariate model was built by including all covariates that were significant univariate predictors of severity ratings and those that contributed to overall model fit. The full model is presented in Table 11. Results of this analysis again

point to vignette child gender as a significant predictor of problem severity ( $r^2 = 0.19$ ,  $t = 4.96$ ,  $p < .0001$ ), with males rated as more severe than females. Controlling for teacher demographics and within-subject effects, vignette child race was not a significant predictor of severity rating. As with the between-subject analysis, interactions between vignette child race and gender, child gender and primary problem type, and child race and primary problem type were all non-significant.

### **3.4.3 Predictors of Referral for Services**

I had next planned to test predictors of non-referral for services. However, the base rate of this phenomenon was too low for reliable statistical testing. At the between-subject level, no first vignette was rated as not needing services. In other words, teachers selected at least one intervention for every first vignette. Across participants, “no services” was selected only nine times, or 0.25% of all 3,896 observations. Descriptively, four of these “no referral” answers were in response to a Black female vignette, two for a Black male vignette, one for a White female vignette, and two for a White male vignette. Further analysis of referral patterns is presented in section 3.5 below.

**Table 11: Multivariate multilevel regression models predicting teacher rating of problem severity.**

Model 1	Estimates of Fixed Effects		
	$r^2$ (SE)	$t$	$p$
Child Gender Male	0.19 (0.04)	4.96	<.0001
Condition			
P Externalizing	0.52 (0.08)	6.21	<.0001
P Internalizing	-0.15 (0.08)	-1.75	.081
Mixed (ref)			
Teacher Age	0.01 (0.00)	1.45	.147
Teacher Gender	-0.25 (0.09)	-2.90	.004
Education Master's Plus	0.09 (0.07)	1.38	.168
Years Teaching Experience	0.01 (0.01)	1.32	.188
Model 2			
Child Race Black	0.04 (0.04)	1.03	.302
Model 3			
Child Gender x Child Race	-0.01 (0.08)	-0.19	.847
Model 4			
Child Gender x Ext Problem	0.01 (0.08)	0.11	.915
Model 5			
Child Race x Ext Problem	0.06 (0.08)	0.76	.450

*Note.* Models 2 through 5 include all main effects in Model 1 with added main effects and interactions. *SE* = standard error; P Externalizing = predominantly externalizing; P Internalizing = predominantly internalizing; ref = reference group for dummy coded multivariate analysis; Ext Problem = externalizing primary problem.

### 3.4.4 Predictors of Referral Confidence

As with severity rating, teacher confidence in their selected referral was rated with a sliding scale from 0 to 10 with one decimal place increments. The resulting outcome measure is sufficiently normally distributed for regression analysis.

The first between-subject analysis looks at main effects of child gender and race on ratings of confidence in the selected service referral for the first vignette presented. Results of these analyses are presented in Table 12. Confidence ratings were similar for vignettes with male ( $M = 6.32$ ,  $SD = 1.64$ ) and female ( $M = 6.35$ ,  $SD = 1.67$ ) names and gender did not have a significant main effect on confidence in services ( $r^2 = -0.02$ ,  $t = -0.24$ ,  $p = .813$ ). Teachers reported slightly lower levels of confidence in services for vignettes with a Black name ( $M = 6.26$ ,  $SD = 1.64$ ) than those with a White name ( $M = 6.41$ ,  $SD = 1.66$ ). However, controlling for teacher demographics, this main effect was not significant ( $r^2 = -0.16$ ,  $t = -1.55$ ,  $p = .121$ ). Contrary to the hypothesized pattern, primary problem type did not significantly moderate the effect of gender or race on teacher ratings of confidence in services.

**Table 12: Multiple regression results predicting teacher rating of confidence in service referral.**

Confidence in Service Referral				
Model 1	<i>M (SD)</i>	<i>r</i> <sup>2</sup> ( <i>SE</i> )	<i>t</i>	<i>p</i>
Child Gender Male	6.32 (1.64)	-0.02 (0.10)	-0.24	.813
Female (ref)	6.35 (1.67)			
Teacher Race				
Black	6.98 (1.59)	0.69 (0.15)	4.62	<.0001
Other	6.43 (1.76)	0.12 (0.29)	0.40	.688
White (ref)	6.23 (1.64)			
Teacher Male	5.96 (1.70)	-0.46 (0.13)	-3.51	<.001
Female (ref)	6.42 (1.63)			
	<i>R</i>			
Years Teaching Experience	-0.08	-0.02 (0.01)	-2.70	.007
Model 2	<i>M (SD)</i>			
Child Race Black	6.26 (1.64)	-0.16 (0.10)	-1.55	.121
White (ref)	6.41 (1.66)			
Model 4				
Externalizing Problem	6.38 (1.64)	-0.16 (0.11)	-1.51	.132
Internalizing Problem (ref)	6.25 (1.67)			
Model 5				
Child Gender x Ext Problem		-0.04 (0.21)	-0.18	.856
Model 6				
Child Race x Ext Problem		0.24 (0.21)	1.12	.262

*Note.* Models 2 through 6 include all covariates in Model 1 with added main effects and interactions. *SE* = standard error; ref = reference group for dummy coded multivariate analysis; Ext Problem = externalizing primary problem.

### 3.4.5 Multilevel Analysis

Results of the intercept-only model indicated 67% of variability in confidence ratings is at the within-subject level ( $ICC = 0.67$ ). Thus all following models account for the fixed and random effects of teacher ID on ratings of confidence in services. Teacher and school district predictors were each tested for univariate effects on confidence ratings. Covariates that were significant or improved model fit were maintained in the final models, presented in Table 13. Results of the first model indicated a significant main effect of gender on teacher confidence, whereby participants reported more confidence in services provided for children with a female name than those with a male name ( $r^2 = -0.10$ ,  $t = -3.19$ ,  $p = .001$ ). Race was not a significant predictor of confidence in the multivariate model. Primary problem type had a significant main effect on teacher confidence in services such that teachers were less confident in the success of services for children categorized as externalizing than those categorized as internalizing or no problem ( $r^2 = -0.12$ ,  $t = -2.96$ ,  $p = .003$ ).

Again, a significant interaction effect was hypothesized whereby teachers would be more confident in services for children with stereotype consistent primary problems. However, results of interaction testing (Table 13) indicated that the effect of primary externalizing problem did not significantly moderate the effects of gender on ratings of confidence in services.



**Table 13: Multivariate multilevel regression models predicting teacher rating of confidence in service referral.**

Model 1	Estimates of Fixed Effects		
	$r^2$ (SE)	$t$	$p$
Child Gender Male	-0.10 (0.04)	-3.19	.001
Teacher Race			
Black	0.68 (0.14)	5.01	<.0001
Other	0.32 (0.27)	1.18	.238
White (ref)			
Teacher Age	-0.01 (0.00)	-1.90	.058
Teacher Gender	-0.39 (0.12)	-3.24	.001
<hr/>			
Model 2			
Child Race Black	-0.02 (0.03)	-0.55	.584
<hr/>			
Model 3			
Externalizing Primary Problem	-0.12 (0.04)	-2.96	.003
<hr/>			
Model 4			
Child Gender x Ext Problem	0.09 (0.07)	1.18	.240
<hr/>			
Model 5			
Child Race x Ext Problem	-0.002 (0.06)	-0.05	.963

*Note.* Models 2 through 5 include all covariates in Model 1 with added main effects and interactions. *SE* = standard error; ref = reference group for dummy coded multivariate analysis; Ext Problem = externalizing primary problem.

### **3.5 Predictors of Ratings of Service Referral Type**

*Hypothesis 2e: Teachers will be more likely to refer White children for mental health services and Black children for educational services. Girls are less likely than boys overall to be referred for services.*

The next set of between- and within-subject analyses look at the effect of vignette child demographics on service referral type. For these analyses, the referral choices are categorized into four groups: parent contact (*parent contact or none*), academic (*academic remediation or academic/special education evaluation*), behavioral (*behavioral remediation or office referral*), or mental health (*assessment by school psychologist, psychiatric medication, psychotherapy/family therapy, or guidance counselor*).

The first analysis to test Hypothesis 2e was a between-subjects multinomial regression analysis examining predictors of primary service referral in response to the first vignette presented. Results of this analysis are presented in Table 14. As with previous models, covariates that had significant ( $p < .05$ ) effect on the outcome of primary referral type were maintained in the multivariate model. As expected, teachers in the predominantly externalizing condition were more likely to select a behavioral intervention for the vignette child than those in the mixed condition ( $OR = 1.12, \chi^2 (3) = 22.28, p < .0001$ ). Teachers in the predominantly internalizing condition were significantly less likely to select parent contact ( $OR = -0.48, \chi^2 (3) = 5.41, p = .020$ ) or a behavioral service ( $OR = -1.22, \chi^2 (3) = 11.28, p < .001$ ) compared to those in the mixed

condition. Special education teachers were less likely than other teachers to refer a child for academic services ( $OR = 1.48, \chi^2 (3) = 8.03, p = .005$ ). Finally, teachers with more experience were less likely to choose a behavioral service referral ( $OR = -0.05, \chi^2 (3) = 11.18, p < .001$ ). Thus it appears that teachers were responding to this question in a thoughtful and attentive manner and the results suggest that this categorization maps on to meaningful group differences in referral patterns. Interestingly, teachers in school districts with a high percentage of Black students were less likely to refer the vignette child for academic services ( $OR = -3.50, \chi^2 (3) = 4.36, p = .037$ ).

However, results of this multivariate analysis also indicate that vignette child gender and race did not significantly predict primary referrals of parent contact, academic, or behavioral services, compared with the reference group of mental health referral. Likewise, the interaction of child gender and race did not significantly predict this categorization.

**Table 14: Multinomial logistic regression predicting parent contact or no service vs. mental health service, academic service vs. mental health service, and behavioral intervention vs. mental health service.**

Model 1	Mental Health (n = 709) n (%)	Parent Contact (n = 204) n (%)	Odds ratio, parent contact <sup>a</sup> (SE)	$\chi^2$	p	Academic (n = 21) n (%)	Odds ratio, academic <sup>a</sup> (SE)	$\chi^2$	p	Behavioral (n = 133) n (%)	Odds ratio behavioral <sup>a</sup> (SE)	$\chi^2$	p
Child Gender Male	357 (64.32)	113 (20.36)	-0.09 (0.08)	1.07	.300	15 (2.70)	-0.45 (0.25)	3.17	.075	70 (12.61)	-0.11 (0.11)	1.17	.279
Female (ref)	352 (68.75)	91 (17.77)				6 (1.17)				63 (12.30)			
Condition													
P Externalizing	198 (54.40)	73 (20.05)	0.19 (0.20)	0.91	.339	9 (2.47)	0.33 (0.53)	0.38	.539	84 (23.08)	1.12 (0.24)	22.28	<.0001
P Internalizing	280 (79.10)	58 (16.38)	-0.48 (0.21)	5.41	.020	5 (1.41)	-0.54 (0.60)	0.79	.373	11 (3.11)	-1.22 (0.36)	11.28	<.001
Mixed (ref)	231 (66.19)	73 (20.92)				7 (2.01)				38 (10.89)			
Teacher Race													
Black	82 (59.85)	27 (19.71)	0.08 (0.25)	0.09	.761	4 (2.92)	1.10 (0.63)	2.98	.084	24 (17.52)	0.31(0.28)	1.18	.278
Other	19 (59.38)	4 (12.50)	-1.12 (0.75)	2.22	.136	3 (9.38)	2.04 (0.73)	7.85	.005	6 (18.75)	0.03 (0.55)	0.00	.952
White (ref)	606 (67.63)	173 (19.31)				14 (1.56)				103 (11.50)			
Special Education Class	60 (69.77)	12 (13.95)	-0.44 (0.33)	1.76	.185	6 (6.98)	1.48 (0.52)	8.03	.005	8 (9.30)	-0.52 (0.43)	1.44	.230
General/AIG Class (ref)	649 (66.16)	192 (19.57)				15 (1.53)				125 (12.74)			
	<u>M (SD)</u>	<u>M (SD)</u>				<u>M (SD)</u>				<u>M (SD)</u>			
Years Teaching Exp	12.95 (7.84)	13.46 (7.80)	0.01 (0.01)	0.51	.473	12.71 (8.53)	-0.01 (0.03)	0.24	.623	10.08 (7.34)	-0.05 (0.01)	11.18	<.001
School Dist % Black	26.10 (18.44)	28.86 (17.65)	0.88 (0.45)	3.73	.053	19.51 (13.28)	-3.50 (1.68)	4.36	.037	30.20 (18.04)	1.05 (0.58)	3.24	.072
Model 2	<u>n (%)</u>	<u>n (%)</u>				<u>n (%)</u>				<u>n (%)</u>			
Child Race Black	363 (66.73)	112 (20.59)	-0.09 (0.08)	1.04	.307	10 (1.84)	0.12 (0.23)	0.26	.610	59 (10.85)	0.16 (0.10)	2.31	.129
White (ref)	346 (66.16)	92 (17.59)				11 (2.10)				74 (14.15)			
Model 3													
Child Gender x Child Race			-0.03 (0.08)	0.15	.695		0.52 (0.32)	2.70	.101		0.03 (0.11)	0.10	.758

Note. Models 2 and 3 include all covariates in Model 1 with added main effects and interaction. SE = standard error; P Externalizing = predominantly externalizing; P Internalizing = predominantly internalizing; ref = reference group for dummy coded multivariate analysis; School Dist % Black = percentage of Black students in school district.

<sup>a</sup>Mental health service referral is the reference category.

### 3.5.1 Multilevel Analysis Predicting Number of Academic Service Referrals

To test within-subject effects of vignette child demographics, a set of multilevel models was built with outcome variables of number of services selected within each of the above referral types. Each of these count variables was sufficiently normally distributed for regression analysis with PROC MIXED. The first of these models analyzed predictors of number of academic service referrals. Results of the intercept-only model indicated that 7% of variability in the outcome measure is at the within-subject level ( $ICC = 0.07$ ). Teacher and school district covariates that were significant at the  $p < .05$  level or which added to model fit were retained in the model. Table 15 displays the final model. As with the between-subject analysis, teachers in the predominantly externalizing condition made more referrals for academic services ( $r^2 = 0.11, t = 3.49, p < .001$ ).

Results of this multilevel analysis indicated that vignettes with male names were referred for significantly more academic services than vignettes with female names ( $r^2 = 0.07, t = 4.96, p < .0001$ ). Similarly, vignettes with Black names were referred for more academic services than vignettes with White names. This effect was marginally significant ( $r^2 = 0.03, t = 2.34, p = .019$ ) at the family-wise error rate of  $p < .0045$  (Table 2). Main effects of race and gender are qualified by a marginally significant interaction between these two predictors ( $r^2 = 0.06, t = 2.04, p = .042$ ). Specifically, the largest number of academic referrals was in response to Black male vignettes ( $M = 0.33, SD = 0.57$ ),

followed by White males ( $M = 0.27$ ,  $SD = 0.51$ ), White females ( $M = 0.23$ ,  $SD = 0.50$ ), and Black females ( $M = 0.22$ ,  $SD = 0.49$ ).

**Table 15: Multivariate multilevel regression models predicting number of academic service referrals.**

Model 1	Estimates of Fixed Effects		
	$r^2$ (SE)	$t$	$p$
Child Gender Male Condition	0.07 (0.01)	4.96	<.0001
P Externalizing	0.11 (0.03)	3.49	<.001
P Internalizing	0.01 (0.03)	0.42	.676
Mixed (ref)			
Teacher Age	0.001 (0.001)	1.42	.155
Education Master's Plus	0.06 (0.02)	2.60	.009
<hr/>			
Model 2			
Child Race Black	0.03 (0.01)	2.34	.019
<hr/>			
Model 3			
Child Gender x Child Race	0.06 (0.03)	2.04	.042

*Note.* Models 2 and 3 include all covariates in Model 1 with added main effects and interaction. *SE* = standard error; P Externalizing = predominantly externalizing; P Internalizing = predominantly internalizing; ref = reference group for dummy coded multivariate analysis.

### 3.5.2 Multilevel Analysis Predicting Number of Behavioral Service Referrals

The next model looks at vignette child demographics as predictors of number of behavioral service referrals. The intercept-only model found that 16% of variability in teacher ratings of behavioral services was at the within-subject level ( $ICC = 0.16$ ). Results of the multivariate analysis are presented in Table 16. Teachers in the predominantly externalizing condition selected more behavioral services than those in the mixed condition ( $r^2 = 0.27, t = 7.30, p < .0001$ ). Likewise, those in the predominantly internalizing condition made fewer behavioral service referrals ( $r^2 = -0.18, t = -4.74, p < .0001$ ).

Controlling for within-subject effects and teacher and school district demographics, vignette child gender was a significant predictor of number of behavioral services ( $r^2 = 0.05, t = 2.95, p = .003$ ). Specifically, vignettes with male names received more behavioral service referrals than those with female names. Vignette child race did not have a significant main effect on number of behavioral service referrals and did not significantly moderate the main effect of gender.

**Table 16: Multivariate multilevel regression models predicting number of behavioral service referrals.**

Model 1	Estimates of Fixed Effects		
	$r^2$ (SE)	$t$	$p$
Child Gender Male Condition	0.05 (0.02)	2.95	.003
P Externalizing	0.27 (0.04)	7.30	<.0001
P Internalizing	-0.18 (0.04)	-4.74	<.0001
Mixed (ref)			
Model 2			
Child Race Black	-0.01 (0.02)	-0.55	.584
Model 3			
Child Gender x Child Race	-0.01 (0.03)	-0.24	.810

*Note.* Models 2 and 3 include all covariates in Model 1 with added main effects and interaction. *SE* = standard error; P Externalizing = predominantly externalizing; P Internalizing = predominantly internalizing; ref = reference group for dummy coded multivariate analysis.

### 3.5.3 Multilevel Analysis Predicting Number of Mental Health Service Referrals

The intercept-only model indicated that 35% of the variability in number of mental health service referrals was at the within-subject level ( $ICC = 0.35$ ). In other words, teachers tended to respond similarly to mental health referral options across vignettes. Results of this analysis are presented in Table 17. Between subject condition was again a significant predictor of number of mental health services, with those in the predominantly externalizing ( $r^2 = 0.19$ ,  $t = 2.88$ ,  $p = .004$ ) and predominantly internalizing ( $r^2 = 0.14$ ,  $t = 2.15$ ,  $p = .031$ ) conditions selecting more mental health referrals than those in the mixed condition.



Vignette child race had a significant main effect on number of mental health referrals such that vignettes with Black names received fewer mental health referrals than those with White names ( $r^2 = -0.07$ ,  $t = -3.12$ ,  $p = .002$ ). There was no significant main effect of vignette child gender on number of mental health service referrals and gender did not significantly moderate the effect main of race.

**Table 17: Multivariate multilevel regression models predicting number of mental health service referrals.**

Model 1	Estimates of Fixed Effects		
	$r^2$ (SE)	$t$	$p$
Child Gender Male Condition	0.01 (0.02)	0.35	.728
P Externalizing	0.19 (0.06)	2.88	.004
P Internalizing	0.14 (0.07)	2.15	.031
Mixed (ref)			
<hr/>			
Model 2			
Child Race Black	-0.07 (0.02)	-3.12	.002
<hr/>			
Model 3			
Child Gender x Child Race	-0.01 (0.05)	-0.14	.889

*Note.* Models 2 and 3 include all covariates in Model 1 with added main effects and interaction. *SE* = standard error; P Externalizing = predominantly externalizing; P Internalizing = predominantly internalizing; ref = reference group for dummy coded multivariate analysis.

### **3.5.4 Mediators of Child Vignette Demographics on Service Referral Decision-Making, Exploratory Analyses**

It was hypothesized that any observed group differences in service referral would be the result of group differences in earlier steps of the decision making process. As such, I tested a series of multilevel models to examine potential mediators for observed significant differences in service referrals. Estimates of direct and indirect effects were calculated with PROC MIXED code developed by Bauer, Preacher, and Gil (2006).

Vignette male gender was a significant predictor of number of academic service referrals (Table 15) and number of behavioral service referrals (Table 16). In addition, male gender significantly predicted the number of externalizing problems selected (Table 8). Results of these analyses suggest that observed gender differences in number of academic and behavioral referrals may be mediated through number of externalizing problems. In particular, teachers may have selected more academic and behavioral services for boys who were rated as having more externalizing problems. Results of these analyses are presented in Table 18. Indirect effects of gender on number of academic and behavioral referrals, mediated through number of externalizing problems were not significant.

A significant effect of vignette child race on number of mental health referrals was also observed, with vignettes with Black names referred to fewer mental health services than those with White names (Table 17). I hypothesized that this effect may be

mediated through differences in teacher ratings of problem cause (teacher decision making step 3). To test this hypothesis, causes were dichotomized into stable (*culture, family, interpersonal deficits, personality/temperament*) and shifting sources (*age, classroom dynamics, academic challenges, hormones, peer influence, psychological problems, and school environment*) based on the work of Chang and colleagues (2003). However, child vignette race was not a significant predictor this binomial stable/shifting etiology variable ( $OR = -0.01, \chi^2 (2) = -0.39, p = .697$ ). One alternative categorization was comparing external causes of behavior (*classroom dynamics, culture, family/home environment, peer influence, and school environment*) versus internal factors (*age, academic challenges, hormones, interpersonal deficits, personality/temperament, and psychological problems*). Again, race did not significantly predict differences in this binomial outcome ( $OR = 0.03, \chi^2 (2) = 1.05, p = .295$ ). Thus I was unable to identify any effects of race on teacher decision making steps one through four to explain the observed differences in mental health referrals at step five; no mediational model was tested.

**Table 18: Multivariate multilevel regression models measuring mediation of gender effects on teacher referral decision-making.**

Model	Direct Effects			Mediator	Indirect Effects		
	$r^2$ (SE)	$z$	$p$		$r^2$ (SE)	$z$	$p$
1. Child Gender on Number of Academic Referrals	0.05 (0.00)	2.73	.003	Number of Externalizing Problems	0.01 (0.00)	0.82	.206
2. Child Gender on Number of Behavioral Referrals	0.07 (0.00)	5.00	<.0001	Number of Externalizing Problems	0.00 (0.00)	0.62	.267

*Note.* SE = standard error.

### **3.6. Measures of Implicit Bias (Implicit Association Test)**

*Hypothesis 3: Overall, teachers will demonstrate a bias toward associating Black children with externalizing symptoms and White children with internalizing symptoms. Similarly they will demonstrate an average bias toward associating males with externalizing symptoms and females with internalizing symptoms. Moreover, individual differences in the strength of these associations will predict bias at each level of the referral decision-making process.*

The goal of the final set of analyses is to test within- and between-subjects effects of implicit bias (as measured by the IAT) on each step of the teacher mental health decision-making process. Unfortunately, the majority of participants did not provide complete data on one or both IAT measure. Several factors contributed to this missing data. First, the IAT was collected at the end of the protocol to avoid revealing the implicit race and gender manipulations. Thus participants who ended the survey early did not complete the IAT. Second, coding within the Qualtrics system made it possible for participants to bypass all or some IAT sections by clicking the “Next” button. I worked with Qualtrics support staff to identify survey options or Javascript code to prohibit skipping IAT sections, but was unable to identify a repair for this issue. Finally, the picture stimuli for the race IAT did not appear properly on some computers. In particular, school firewall security may have blocked the picture stimuli from loading. This disproportionately affected participants who completed the survey on work computers in one large county (Wake). Unfortunately the IAT scoring algorithm

(Greenwald et al., 2003) does not provide guidance on incomplete data and requires responses to all 4 IAT test blocks. As a result, *D* scores were only calculated for a total of 404 gender IATs (36.5% of participants) and 324 race IATs (29.3% of participants).

Importantly, there were no significant demographic differences between participants with missing data on the gender or race IAT and those with complete data (all  $p > .10$ ). Likewise, randomized behavioral condition was not a significant predictor of missing IAT data. Multilevel modeling is used for all IAT analyses to account for within-subject effects. Further, the estimator used for PROC MIXED, Restricted Maximum Likelihood, is robust to missing at random (MAR) data as in this case. However, given the large amount of missing data, results of the following analyses should be interpreted with caution.

*D* score measures of implicit bias were calculated using the algorithm presented by Greenwald and colleagues (2003). First, trials with response latencies greater than 10,000 milliseconds (ms) were eliminated. Next, subjects with more than 10% of trials faster than 300 ms were eliminated. Then a mean and standard deviation was calculated for each block. These values were used to assign a “penalty” for each incorrect response: the mean value plus two times the standard deviation. After assigning penalties for incorrect responses, an adjusted mean was calculated for each block. These adjusted means were used to calculate two difference scores: Block 6 minus Block 3 and Block 7 minus Block 4. Each of these difference scores was divided by the adjusted pooled

standard deviation for the associated blocks (Block 3 and 6 or Block 4 and 7). The final  $D$  score is an average of these two values.

As anticipated, average observed  $D$  scores were positive for both the race ( $M = 0.19$ ,  $SD = 0.33$ ) and gender ( $M = 0.26$ ,  $SD = 0.30$ ) IATs. Thus participants, on average, showed a bias toward associating Black pictures with externalizing symptoms and White pictures with internalizing symptoms. Likewise they showed a bias toward associating male words with externalizing and female words with internalizing. Average  $D$  scores for race and gender were in the “slight” bias category ( $D > 0.15$ ) identified by Nosek, Greenwald, & Banaji (2006). In summary, the race IAT showed an average slight bias toward associating Black children with externalizing symptoms; the gender IAT showed an average slight bias toward associating male children with externalizing symptoms.

Black teachers had significantly lower scores on the race IAT ( $M = 0.05$ ,  $SD = 0.26$ ) than non-Black teachers ( $M = 0.21$ ,  $SD = 0.34$ ,  $t = 2.57$ ,  $p = 0.003$ ). Male teachers also had lower race IAT scores ( $M = 0.11$ ,  $SD = 0.32$ ) than female teachers ( $M = 0.21$ ,  $SD = 0.34$ ,  $t = 1.91$ ,  $p = 0.046$ ). Finally, teacher age was significantly correlated with gender  $D$  score, such that older teachers had higher scores ( $r^2 = 0.003$ ,  $t = 2.08$ ,  $p = 0.038$ ). No other demographic variables significantly predicted variability in race or gender  $D$  scores at the  $p < .05$  level.

### **3.6.1 Multilevel Analysis**

It was hypothesized that implicit bias, as measured by the IAT would moderate each step of the mental health decision-making process. Specifically, I hypothesized that teachers with higher degrees of bias would be more likely to make stereotype-consistent decisions. To this end, an interaction term was added to each of the previously presented multilevel models estimating the moderation effect of gender bias on gender effects and racial bias on race effects. Parameter estimates for each of these interaction effects are presented in Table 19. Results of these analyses indicate that bias, as measured by the IAT, did not significantly moderate any of the previously presented main effects.



**Table 19: Multivariate multilevel regression models measuring moderation effects of implicit bias as measured by the Implicit Association Test (IAT).**

Model	Number of Externalizing Problems			Severity Rating			Confidence Rating		
	<i>r</i> <sup>2</sup> (SE)	<i>t</i>	<i>p</i>	<i>r</i> <sup>2</sup> (SE)	<i>t</i>	<i>p</i>	<i>r</i> <sup>2</sup> (SE)	<i>t</i>	<i>p</i>
Model 1									
Child Gender Male	0.15 (0.03)	5.89	<.0001	0.19 (0.04)	4.96	<.0001	-0.10 (0.04)	-3.19	.001
Model 2									
Child Gender x Gender IAT	0.20 (0.14)	1.41	.157	-0.02 (0.19)	-0.09	.927	0.08 (0.16)	0.51	.609
Model 3									
Child Race Black	0.08 (0.06)	1.43	.152	0.04 (0.04)	1.03	.302	-0.02 (0.03)	-0.55	.584
Model 4									
Child Race x Race IAT	0.25 (0.15)	1.66	.094	0.01 (0.19)	0.06	.950	-0.23 (0.16)	-1.44	0.150

Model	Number of Academic Referrals			Number of Behavioral Referrals			Number of Mental Health Referrals		
	$r^2$ (SE)	$t$	$p$	$r^2$ (SE)	$t$	$p$	$r^2$ (SE)	$t$	$p$
Model 5									
Child Gender Male	0.07 (0.01)	4.96	<.0001	0.05 (0.02)	2.95	.003	0.01 (0.02)	0.35	.728
Model 6									
Child Gender x Gender IAT	0.09 (0.07)	1.26	.208	0.09 (0.08)	1.06	.288	0.00 (0.11)	0.00	.996
Model 7									
Child Race Black	0.03 (0.01)	2.34	.019	-0.01 (0.02)	-0.55	.584	-0.07 (0.02)	-3.12	.002
Model 8									
Child Race x Race IAT	0.01 (0.07)	0.08	.936	0.00 (0.07)	0.06	.955	-0.07 (0.09)	-0.75	.455

Note. SE = standard error.

## 4. Discussion

The purpose of the present study was to better understand the social cognitive processes that precede teacher decision-making about service referral in cases of problematic child behavior, and how these processes may be influenced by race and gender biases. This line of inquiry is important because teachers serve a crucial role as primary gatekeepers for mental health treatment (Alegría et al., 2012; Hogan, 2003; Wu et al., 1999). However, teachers are provided limited formal training in identifying and referring children with mental health conditions (Hogan, 2003; Walter et al., 2006). Thus, they are overly reliant on colloquial knowledge, which may be influenced by pervasive cultural stereotypes about the types of problems children are likely to exhibit. Such biases may contribute to observed disparities in mental health treatment for girls and children of color (Garland et al., 2005; Hogan, 2003; Walter et al., 2006; Yeh et al., 2003).

In order to experimentally test implicit bias in student mental health referrals, previous studies have presented behavioral vignettes to teachers in which perceived child race and gender were manipulated through given names. A similar protocol was utilized for the current study. Building upon previous research, the vignettes in this study presented descriptions of complex behavior patterns including both internalizing and externalizing problems. These vignettes provided increased ecological validity given large amounts of comorbidity and symptom overlap in child clinical populations (Allen-Meares, 1991; Donnellan et al., 2005; Kessler et al., 2012). Moreover, these

complex descriptions afforded ambiguity, thus increasing the opportunity for implicit biases to influence decision-making (Klayman, 1995).

This study also built upon previous research on teacher's implicit bias by measuring several decision-making steps leading up to ultimate referral choice. These preliminary steps included: problem recognition, assessment of problem severity, assessment of problem cause(s), and assessment of referral goal. It was theorized that bias in decision-making would be mediated primary through attribution of problem type. In particular, I hypothesized that vignettes with male and/or Black names would be rated as externalizing more often than White female vignettes. Given that externalizing problems are generally rated as more severe (Cantwell et al., 1997) and children with externalizing problems are more likely to be referred for mental health treatment (Finkelhor et al., 2001; Gudiño et al., 2009; Kazdin & Weisz, 2003; Thompson, 2005), I expected attribution of problem type to account for any observed differences in service referral outcomes.

#### ***4.1 Teacher Recognition of Child Problem Type***

The behavioral profiles utilized in this study were based on the extensively normed Teacher Report Form (TRF; Achenbach, 1991). Specifically, the number and type of symptoms were selected to depict a marginally clinical profile of a middle-school aged child. Symptoms were selected to represent common behaviors of both boys and girls. The rationale behind utilizing marginally clinical vignettes was that teachers

would be inclined to select “no problem” for a large portion of the vignettes presented. However, this did not occur. Teachers selected the “no problem” option in response to only 1% of vignettes (Table 4). As such, I was unable to reliably measure predictive models of how race and gender bias impact detecting a mental health problem versus categorizing behavior as normative.

Overall, the name manipulation utilized in this experiment had minimal within- or between-subjects effects on teacher categorizations of internalizing versus externalizing problems. Specifically, vignettes with Black names were no more likely to be labeled externalizing than those with White names. Likewise, there were no significant effects of vignette child race on number of internalizing or externalizing symptoms. Contrary to my hypothesis, vignette condition (predominantly internalizing, predominantly externalizing, or mixed) did not significantly moderate the effect of race on problem classification. In other words, teachers who received the most ambiguous vignettes (those in the mixed condition) did not exhibit increased effects of implicit bias compared with the other two conditions.

The results of the gender manipulation were more mixed. Child vignette gender did not have a significant effect on categorization of internalizing versus externalizing primary problem. Likewise, gender did not significantly predict number of internalizing problems selected. However, the number of externalizing problems assigned was higher for vignettes with a male name than those with a female name. This result suggests that,

as hypothesized, teachers may over-attend to symptoms that are consistent with their stereotypes of problematic male behavior. This likewise means that they may overlook these problems when exhibited by a female.

As described above, there are two conflicting theories for how implicit bias may influence assessments of mental health problems. The first proposes that teachers and other important adults will be more likely to overlook stereotype-consistent problematic behavior because it is deemed normative for the child's group (Chang & Sue, 2003; Thompson, 2005). The alternative model suggests that adults are primed to notice stereotype-consistent problems and overlook those that are inconsistent with expectations (Epstein et al., 2005; Hosterman et al., 2008). The gender effect identified in this study adds to a growing body of support for the latter theory.

## ***4.2 Teacher Ratings of Problem Severity***

Overall, teacher ratings of severity for the experimental vignettes was quite high ( $M = 7.48$  out of 10,  $SD = 1.44$ ), suggesting that the behavioral profiles presented were out of the ordinary for average middle school students. I hypothesized that ratings of severity would be higher for children exhibiting stereotype consistent behaviors. Results of interaction testing (primary problem type  $\times$  gender and primary problem type  $\times$  race) did not support this hypothesis. However, there was a marginally significant main effect of vignette child gender on ratings of severity, with males rated higher than females. In other words, mental health symptoms were rated as more severe when attributed to a

male than when attributed to a female, regardless of the problem type. This result is consistent with previous research suggesting that teachers view females as more mature and capable of dealing with mental health issues than their male peers (M. T. Green et al., 1996).

### ***4.3 Teacher Ratings of Confidence in Service Referral***

As with severity ratings, I hypothesized that ratings of confidence in the success of services would be highest for stereotype-consistent behavior. In other words, teachers would be most confident about the success of services viewed as appropriate for a child's group. Results of interaction analyses (primary problem type x race and primary problem type x gender) did not support this hypothesis. However, a main effect of primary problem type was observed. Specifically, teachers reported significantly less confidence in the success of services for children categorized as externalizing. This notion has a basis in reality, as children with externalizing problems may have low motivation for changing their behavior (Kazdin, Mazurick, & Siegel, 1994). At the same time, this assumption may prevent teachers from referring children with externalizing problems for services, which could result in such children not having the opportunity for effective intervention.

### ***4.4 Teacher Referrals for Services***

I hypothesized that perceived child race and gender would have significant effects on the type of service to which a vignette child was referred. Based on previous

research, I expected that Black children would be referred for more behavioral and special education services than White children (Bryan et al., 2012; Coutinho & Oswald, 2000; Skiba et al., 2006) and that boys would be referred for more services overall than girls (Achenbach & Edelbrock, 1981; Achenbach et al., 1991). I further hypothesized that this effect would be mediated through attribution of primary problem type.

Results of this analysis suggested that vignettes with male names were referred for more academic and behavioral services than those with female names. However, this effect was not significantly mediated through primary problem type. Likewise, vignettes with a Black name were marginally more likely to be referred for academic services than those with a White name. The observed results may help explain the over-representation of Black and male children in special education services (Skiba et al., 2006).

Interestingly, there was also a main effect of race on number of mental health services selected. In particular, vignettes with a Black name were referred to fewer mental health services than those with a White name. This result suggests an implicit assumption that mental health services are more appropriate for White children than for Black children. This implicit bias may help explain disparities in receipt of mental health treatment for Black children (Garland et al., 2005; Yeh et al., 2003). Unfortunately, race was not a significant predictor of variability in any other decision-making step. As such, no mediational analyses were possible. In other words, the hypothesized social cognitive



model does not appear to account for a significant amount of variability in the observed disparities in service referral.

#### ***4.5 Implicit Association Test***

An adaptation of the Implicit Association Test (IAT; Greenwald et al., 1998) was developed for this study and pilot tested with a sample of 20 middle school teachers. This measure was designed to identify individual differences in two types of implicit associations: 1) pictures of White children with internalizing symptoms and pictures of Black children with externalizing symptoms and 2) female words with internalizing symptoms and male words with externalizing symptoms. On average, teachers demonstrated slight bias on both the race and gender IAT measures. However, Black teachers exhibited significantly lower (almost zero) racial bias. Male teachers also exhibited lower levels of bias on the race IAT than their female counterparts. Teacher age was significantly correlated with gender IAT score, with older teachers demonstrating more bias. However, contrary to the hypothesized pattern, IAT scores did not significantly predict within- or between-subject variability in race or gender effects in any step of the decision-making process.

#### ***4.6 Limitations***

The results of this analysis have the potential to significantly contribute to the existing literature on implicit bias and referrals for child mental health treatment. However, several limitations of the present study should be noted.

First, there was limited variability in teacher ratings of problem type and severity. This pattern is likely related to the vignettes that were developed for this study. Specifically, the behavior profiles included too many overtly problematic behaviors. The intention was to describe marginally clinical behavior patterns. However, without including any positive or prosocial behavior descriptors, vignette children were rated as highly pathological compared with the average middle school child.

Second, the solicitation of participants through email resulted in a very low participation rate (8.5%). It is impossible to empirically compare participants and non-participants, but it can be assumed that participants had a higher interest in, and likely greater knowledge of, psychological research than their non-participating peers. This savvy participant group may have been less susceptible to the deception involved in this protocol. As such, estimates of implicit bias may be artificially low. Alternatively, these participants may have been susceptible to demand characteristics and thus altered their responses to conform to experimenter expectations. Alternative recruitment styles are encouraged for future research in this domain in order to produce a more representative sample of teachers.

Further, the time to complete the full study (approximately 30 minutes) may have discouraged many participants, and likely contributed to missing data, particularly on the IAT. The protocol was developed to allow for repeated measures, with the intention of modeling the within-subject effects of race and gender bias. However,

observed within-subject effects were small, and were consistent with between-subject observations. Further, behavioral descriptions were very similar within each condition (predominantly internalizing, predominantly externalizing, mixed). The repeated measure design resulted in at least some carryover effects, whereby teachers were alerted to the manipulation over the course of multiple trials and modified their response pattern accordingly. Thus, the study design may have suppressed within-subject effects of bias. Between-subject-only designs are recommended for future research utilizing the vignette procedure.

Finally, the use of an online data collection system presented a number of challenges. In particular, the researcher did not have direct contact with the majority of participants. This prohibited the opportunity for addressing any confusion about the protocol and likely contributed to the low response rate. Further, online data collection allowed less control over data quality and completion of the full protocol. Issues with the online interface particularly affected collection of IAT data, which was largely missing or incomplete. These factors should be taken into account when selecting data collection procedures for future studies.

#### ***4.7 Future Directions***

Overall the findings of this study map on to previous research regarding implicit bias and referral for mental health treatment. However, the hypothesized decision-making model tested in this study predicted only small amounts of variation in ultimate

referral decision. As such, more research is necessary to explicate the social cognitive processes that impact teacher decision-making about mental health referrals.

Perhaps unsurprisingly, several teachers contacted the researcher to express discomfort about the measurement of implicit racial bias. They were concerned about teachers being portrayed as prejudiced towards the children in their own classrooms. There is no reason to believe that teachers possess more bias than the average adult; however, training on the topic of implicit bias is useful for anyone in a position of power who may make important decisions without fully understanding the social cognitive processes at play. Interestingly, teachers did not express the same level of discomfort regarding assessments of gender bias. On the contrary, one pilot study participant reported the gender IAT was difficult because, "I know it was hard for me to remember that boys and internal behaviors were paired together because I don't hear them together often, and with the girls and external behavior." As observed within this study, both race and gender biases are common and increased open dialogue about these processes can help minimize their impact on decision-making. Such conversations should be incorporated into teacher training on mental health intervention.

Finally, the IAT measure developed for this study appeared to be measuring the intended construct in that participants, on average, demonstrated a bias in the hypothesized directions. However, the *D* score measure was not significantly predictive of observed patterns of implicit bias. As such, I was unable to establish convergent

validity for this adaptation of the IAT. Establishing convergent validity of this child behavior IAT will be an important next step in this line of research.

#### **4.8 Conclusions**

The goal of the present study was to gain further insight into the social cognitive processes that influence teacher assessments of child mental health problems and need for treatment, with a particular emphasis on the effects of perceived child race and gender. Results of this analysis suggest that implicit racial and gender bias may have an impact on teacher referral of students for mental health treatment. In particular, behavioral vignettes were assigned more externalizing problems and rated as more severe when presented with a male name than with a female name. Vignettes with a male name were more likely to be referred for academic or behavioral services than those with a female name. Likewise, a vignette with a Black child name was slightly more likely to be referred for academic services and less likely to be referred for mental health treatment than a vignette with a White child name. The observed patterns may help explain group differences in referrals for mental health treatment. Future research will be necessary to further explicate the social cognitive processes underlying these implicit biases. In addition, more research is necessary to identify appropriate training and policy interventions to support teachers in making mental health referrals and reducing the impact of race and gender bias.

## Appendix A: Informed Consent Documents

### ***Appendix A1: Email solicitation for participation***

Subject: Research Participation Invitation: Teacher Decision-Making about Student Mental Health

This email message is an approved request for participation in research that has been approved by the Duke University Institutional Review Board (IRB).

Hello First Name,

You were selected for participation in this study as a current North Carolina middle school teacher/administrator. This goal of this study is to collect information about how teachers make decisions regarding potentially problematic child behavior. This survey will take approximately 30 minutes to complete. Survey must be completed on a computer (laptop or desktop), NOT a mobile device (phone or tablet). **Upon completion, you will be entered into a drawing for a \$40 Amazon gift card.** Odds of winning are approximately 1 in 27.

Participation in this research study is completely voluntary. You have the right to withdraw at any time or refuse to participate entirely. All data obtained from participants will be kept confidential and will only be reported in an aggregate format (by reporting only combined results and never reporting individual ones). Demographic information and email address (for payment) will be collected and stored separately from survey responses.

This project Teacher Decision-Making about Student Mental Health was approved by the Duke University IRB on 11/08/2016. Pertinent questions or concerns about the research, research participants' rights, and/or research-related injuries to participants should be directed to supervising investigator, Kenneth Dodge (919-613-9303, [dodge@duke.edu](mailto:dodge@duke.edu)) and Director of Human Subjects Protection, Lorna Hicks (919-684-3030, [ors-info@duke.edu](mailto:ors-info@duke.edu)).

Questions about this research should be addressed to Shelley Alonso-Marsden, (919-813-0472 or [sma30@duke.edu](mailto:sma30@duke.edu)).

## ***Appendix A2: Online Consent Form***

### **Teacher Decision-Making about Student Mental Health**

#### Introduction

This study attempts to collect information about how teachers make decisions regarding potentially problematic child behavior.

#### Procedures

First, you will be asked to complete some basic demographic questions regarding yourself and your teaching experience. Then, you will be shown a series of four paragraphs describing a child and his/her behavior and asked to complete a questionnaire about each child. Each questionnaire consists of 10 questions which are designed to determine how you make decisions about child behavior and possible referrals for services. Following these four stories, you will be asked to complete a task that involves quickly sorting children into groups based on their behavior and other characteristics. This questionnaire will be conducted with an online Qualtrics-created survey and should take you 20-30 minutes to complete.

Survey must be completed on a computer (laptop or desktop), NOT a mobile device (phone or tablet). Please minimize distractions for most accurate responses.

#### Risks/Discomforts

Risks are minimal for involvement in this study. However, you may feel emotionally uneasy when asked to make judgments based on the behavioral descriptions provided. Although we do not expect any harm to come upon any participants due to electronic malfunction of the computer, it is possible though extremely rare and uncommon.

#### Benefits

There are no direct benefits for participants. However, it is hoped that through your participation, researchers will learn more about how teachers make decisions regarding problematic child behavior.

#### Confidentiality

All data obtained from participants will be kept confidential and will only be reported in an aggregate format (by reporting only combined results and never reporting individual ones). All questionnaires will be concealed, and no one other than the primary investigator listed below will have access to them. The data collected will be stored in the HIPPA-compliant, Qualtrics-secure database until it has been deleted by the primary

investigator. Demographic information and email address (for payment) will be collected and stored separately from questionnaire responses.

#### Compensation

Participants will be entered into a drawing for a \$40 Amazon gift card. Odds of winning are approximately 1 in 22.5. Upon completion of the questionnaire, you will be asked to provide an email address to be entered into this drawing. If you are selected, you will be provided contact number for our reimbursement personnel.

#### Participation

Participation in this research study is completely voluntary. You have the right to withdraw at any time or refuse to participate entirely. If you desire to withdraw, please close your internet browser and notify the principal investigator at this email: sma30@duke.edu.

#### Questions about the Research

If you have questions regarding this study, you may contact graduate researcher Shelley Alonso-Marsden, at 919-813-0472 or sma30@duke.edu, or research advisor Kenneth Dodge, PhD at 919-613-9303 or dodge@duke.edu.

#### Questions about your Rights as Research Participants

If you have questions you do not feel comfortable asking the researcher, you may contact Dr. Kenneth Dodge 919-613-9303, dodge@duke.edu. Or contact the director of Duke University's Institutional Review Board, Lorna Hicks, 919-684-3030, ors-info@duke.edu.

I have read, understood, and printed a copy of, the above consent form and desire of my own free will to participate in this study.

Yes

No



## **Appendix B: Study Materials**

### ***Appendix B1: Sample Vignettes***

INSTRUCTIONS: Please read the following child description carefully and then answer the questions that follow. Remember that you are rating behavior for a MIDDLE SCHOOL AGED child.

#### **Predominantly Internalizing Vignette:**

Malik tends to worry more than his peers. Overall, he appears high strung, at times refuses to talk, and is overtired. He seems to fear he might think or do something bad, is overly anxious to please, and is very secretive. Malik hangs around with kids who get in trouble, has sudden changes in mood, and demands a lot of attention. He is overall very fearful/anxious, is afraid to make mistakes, and feels that others are out to get him.

#### **Predominantly Externalizing Vignette:**

Katie has a pattern of teasing other students. She lies and cheats, uses obscene language, and physically attacks others. She talks out of turn and disrupts other students. She doesn't seem to feel guilty after misbehaving. Katie is also high strung and tense, overtired, and feels that no one likes her. She is easily frustrated and her demands must be met immediately. She talks too much and shows explosive, unpredictable behavior.

#### **Mixed Symptom Vignette:**

Jack is more self-conscious than others his age. He complains of loneliness and feeling worthless. He has sudden changes in mood, disrupts class discipline, and can be defiant and disobedient with school staff. Jack appears depressed much of the time, feels hurt when criticized, and does not engage much with peers. He has a quick temper, is easily jealous, and often appears sullen or irritable.

## **Appendix B2: Child Behavior Information Processing Questionnaire**

Response choices with a strikethrough were eliminated following pilot testing. Response choices in bold were added following pilot testing.

Assume that you are this child's teacher. With the information you have, please answer questions 1-10.

- 1) How would you describe this child's PRIMARY problem? (Select one.)
  - a. None/age appropriate behavior
  - b. Anger
  - c. Anxiety/worry
  - d. Attention difficulties
  - e. Hyperactivity
  - f. Impulsivity
  - ~~g. Physical problems~~
  - ~~h. Rule breaking~~
  - i. Sadness/depression
  - j. Shyness
  - k. Social withdrawal
  
- 2) What other problems, if any, do you think this child has? (Select all that apply.)
  - a. None/age appropriate behavior
  - b. Anger
  - c. Anxiety/worry
  - d. Attention difficulties
  - e. Hyperactivity
  - f. Impulsivity
  - ~~g. Physical problems~~
  - ~~h. Rule breaking~~
  - i. Sadness/depression
  - j. Shyness
  - k. Social withdrawal
  
- 3) How severe would you describe this child's problem(s) in terms of his/her day to day functioning in the classroom and elsewhere?



- 4) What do you think is the PRIMARY cause of this child's behavior? (Select one.)
- a. **Academic challenges**
  - b. Age
  - c. Classroom dynamics
  - d. Culture
  - e. Family/home environment
  - f. ~~Genetics~~
  - g. Hormones
  - h. Interpersonal deficits
  - i. Peer influence
  - j. Personality/temperament
  - k. Psychological problems
  - l. **School environment**
  - m. ~~Teacher~~
- 5) What other causes, if any, do you think have influenced this behavior? (Select all that apply.)
- a. **Academic challenges**
  - b. Age
  - c. Classroom dynamics
  - d. Culture
  - e. Family/home environment
  - f. ~~Genetics~~
  - g. Hormones
  - h. Interpersonal deficits
  - i. Peer influence
  - j. Personality/temperament
  - k. Psychological problems
  - l. **School environment**
  - m. ~~Teacher~~
- 6) What is the PRIMARY referral you would make for this child? (Select one.)
- a. None
  - b. Academic remediation

- c. Academic/special education evaluation
- d. Assessment by school psychologist
- e. Behavioral remediation
- f. ~~Corporal punishment~~
- g. Guidance counselor**
- h. Office referral
- i. Parent contact**
- j. Psychiatric medication evaluation
- k. Psychotherapy/Family therapy

7) What additional services (if any) would be appropriate for this child? (Select all that apply.)

- a. None
- b. Academic remediation
- c. Academic/special education evaluation
- d. Assessment by school psychologist
- e. Behavioral remediation
- f. ~~Corporal punishment~~
- g. Guidance counselor**
- h. Office referral
- i. Parent contact**
- j. Psychiatric medication evaluation
- k. Psychotherapy/Family therapy

8) How confident are you that this/these service(s) will be successful for this child?



9) What would you MOST want to see improved following service(s)? (Select one.)

- a. Nothing/no service needed
- b. Academic performance
- c. Classroom behavior
- d. Emotional wellbeing
- e. Relationships with teachers and peers

- 10) What else, if anything, would you like to see improved by services? (Select all that apply.)
- a. Nothing/no service needed
  - b. Academic performance
  - c. Classroom behavior
  - d. Emotional wellbeing
  - e. Relationships with teachers and peers

## **Appendix B3: Implicit Association Test**

INSTRUCTIONS: In the next task, you will be presented with a set of words or images to classify into groups. This task requires that you classify items as quickly as you can while making as few mistakes as possible. Going too slow or making too many mistakes will result in an uninterpretable score. This part of the study will take about 5 minutes. The following is a list of category labels and the words that belong to each of those categories.

Category	Items
Externalizing Behaviors ("Acting Out")	Angry, Argues, Cheats, Fights, Lies, Loud, Mean, Screams, Steals, Swears
Internalizing Behaviors ("Shutting Down")	Anxious, Cries, Depressed, Fearful, Nervous, Sad, Shy, Timid, Unhappy, Worries
Female	Daughter, Female, Girl, Her, Hers, She, Sister
Male	Boy, Brother, He, Him, His, Male, Son

*Keep in mind*

- Keep your index fingers on the 'A' and 'L' keys to enable rapid response.
- Two labels at the top will tell you which words or images go with each key.
- Each word or image has a correct classification. Most of these are easy.
- The test gives no results if you go slowly – Please try to go as fast as possible.
- Expect to make a few mistakes because of going fast. That's OK.
- For best results, avoid distractions and stay focused.

**Sample Picture Stimuli**



## References

- Achenbach, T. M. (1991). *Manual for the Teacher's Report Form and 1991 profile integrative guide to the 1991 CBCL/4-18, YSR, and TRF Profiles*. Burlington, VT: University of Vermont, Department of Psychology.
- Achenbach, T. M., & Edelbrock, C. S. (1981). Behavioral problems and competencies reported by parents of normal and disturbed children aged four through sixteen. *Monographs of the Society for Research in Child Development*, 46(1). doi: 10.2307/1165983
- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles: Child Behavior Checklist for Ages 6-18, Teacher's Report Form, Youth Self-Report, an integrated system of multi-informant assessment*. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.
- Achenbach, T. M., Howell, C. T., Quay, H. C., & Conners, C. K. (1991). National survey of problems and competencies among four- to sixteen-year-olds: Parents' reports for normative and clinical samples. *Monographs of the Society for Research in Child Development*, 56(3). doi:10.2307/1166156
- Alegría, M., Bijl, R. V., Lin, E., Walters, E. E., & Kessler, R. C. (2000). Income differences in persons seeking outpatient treatment for mental disorders: A comparison of the United States with Ontario and The Netherlands. *Archives of General Psychiatry*, 57, 383-391.
- Alegría, M., Lin, J. Y., Green, J. G., Sampson, N. A., Gruber, M. J., & Kessler, R. C. (2012). Role of referrals in mental health service disparities for racial and ethnic minority youth. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(7), 703-711. doi: 10.1016/j.jaac.2012.05.005
- Allen-Meares, P. (1991). A study of depressive characteristics in behaviorally disordered children and adolescents. *Children and Youth Services Review*, 13(4), 271-286. doi: 10.1016/0190-7409(91)90064-O
- Alonso-Marsden, S., & Dodge, K. A. (2017). *Disparities in conduct problem ratings among African American, Latino, and White children and adolescents: A research synthesis*. Manuscript in preparation.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Angold, A., Erkanli, A., Farmer, E. M. Z., Fairbank, J. A., Burns, B. J., Keeler, G., & Costello, E. J. (2002). Psychiatric disorder, impairment, and service use in rural African American and White youth. *Archives of General Psychiatry*, 59(10), 893-904. doi: 10.1001/archpsyc.59.10.893



- Angold, A., Messer, S. C., Stangl, D., Farmer, E. M. Z., Costello, E. J., & Burns, B. J. (1998). Perceived parental burden and service use for child and adolescent psychiatric disorders. *American Journal of Public Health, 88*(1), 75-80. doi: 10.2105/AJPH.88.1.75
- Bates, D. (2013). *Linear mixed model implementation in lme4*. University of Wisconsin - Madison.
- Bates, D., Maechler, M., & Bolker, B. (2013). *lme4: Linear mixed-effects models using Eigen and Eigenfaces*. Retrieved from <http://CRAN.R-project.org/package=lme4>
- Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations. *Psychological methods, 11*(2), 142.
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review, 94*(4), 991-1013.
- Bertrand, M., Chugh, D., & Mullainathan, S. (2005). Implicit discrimination. *American Economic Review, 95*(2), 94-98.
- Blake, J. J., Butler, B. R., Lewis, C. W., & Darenbourg, A. (2011). Unmasking the inequitable discipline experiences of urban Black girls: Implications for urban educational stakeholders. *The Urban Review, 43*(1), 90-106. doi: 10.1007/s11256-009-0148-8
- Browne, W. J., Golalizadeh Lahi, M., & Parker, R. M. A. (2009). *A guide to sample size calculations for random effect models via simulation and the MLPowSim software package*. Bristol, United Kingdom: University of Bristol.
- Bryan, J., Day-Vines, N. L., Griffin, D., & Moore-Thomas, C. (2012). The disproportionality dilemma: Patterns of teacher referrals to school counselors for disruptive behavior. *Journal of Counseling & Development, 90*(2), 177-190. doi: 10.1111/j.1556-6676.2012.00023.x
- Burns, B. J., Costello, E. J., Angold, A., Tweed, D., Stangl, D., Farmer, E. M., & Erkanli, A. (1995). Children's mental health service use across service sectors. *Health Affairs, 14*(3), 147-159.
- Butler, B. R., Lewis, C. W., Moore, J. L., III, & Scott, M. E. (2012). Assessing the odds: Disproportional discipline practices and implications for educational stakeholders. *Journal of Negro Education, 81*(1), 11-24.
- Cantwell, D. P., Lewinsohn, P. M., Rohde, P., & Seeley, J. R. (1997). Correspondence between adolescent report and parent report of psychiatric diagnostic data. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*(5), 610-619. doi: 10.1097/00004583-199705000-00011
- Chakroff, A. (2013). IAT (2).qsf [Qualtrics Javascript code]. Retrieved from <https://sites.google.com/site/amazonturkshop/documents>

- Chang, D. F., & Sue, S. (2003). The effects of race and problem type on teachers' assessments of student behavior. *Journal of Consulting and Clinical Psychology, 71*(2), 235-242. doi: 10.1037/0022-006X.71.2.235
- Chavez, L. M., Shrout, P. E., Alegría, M., Lapatin, S., & Canino, G. (2010). Ethnic differences in perceived impairment and need for care. *Journal of Abnormal Child Psychology, 38*(8), 1165-1177. doi: 10.1007/s10802-010-9428-8
- Chen, K. W., Killeya-Jones, L. A., & Vega, W. A. (2005). Prevalence and co-occurrence of psychiatric symptom clusters in the U.S. adolescent population using DISC predictive scales. *Clinical Practice and Epidemiology in Mental Health, 1*(1), 22. doi: 10.1186/1745-0179-1-22
- Copeland, W., Shanahan, L., Costello, E. J., & Angold, A. (2011). Cumulative prevalence of psychiatric disorders by young adulthood: A prospective cohort analysis from the Great Smoky Mountains study. *Journal of the American Academy of Child & Adolescent Psychiatry, 50*(3), 252-261. doi: 10.1016/j.jaac.2010.12.014
- Costello, E. J., & Janiszewski, S. (1990). Who gets treated? Factors associated with referral in children with psychiatric disorders. *Acta Psychiatrica Scandinavica, 81*(6), 523-529. doi: 10.1111/j.1600-0447.1990.tb05492.x
- Coutinho, M. J., & Oswald, D. P. (2000). Disproportionate representation in special education: A synthesis and recommendations. *Journal of Child and Family Studies, 9*(2), 135-156. doi: 10.1023/A:1009462820157
- Dodge, K. A. (2011). Social information processing models of aggressive behavior. In M. Mikulincer & P. R. Shaver (Eds.), *Understanding and reducing aggression, violence, and their consequences*. Washington, DC: American Psychological Association.
- Dodge, Kenneth A., & Newman, J. P. (1981). Biased decision-making processes in aggressive boys. *Journal of Abnormal Psychology, 90*(4), 375-379. doi: 10.1037/0021-843X.90.4.375
- Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., Moffitt, T. E., & Caspi, A. (2005). Low self-esteem is related to aggression, antisocial behavior, and delinquency. *Psychological Science, 16*(4), 328-335. doi: 10.1111/j.0956-7976.2005.01535.x
- Epstein, J. N., Willoughby, M., Valencia, E. Y., Tonev, S. T., Abikoff, H. B., Arnold, L. Eugene, & Hinshaw, S. P. (2005). The role of children's ethnicity in the relationship between teacher ratings of attention-deficit/hyperactivity disorder and observed classroom behavior. *Journal of Consulting and Clinical Psychology, 73*(3), 424-434. doi: 10.1037/0022-006X.73.3.424
- Erath, S. A., Keiley, M. K., Pettit, G. S., Lansford, J. E., Dodge, K. A., & Bates, J. E. (2009). Behavioral predictors of mental health service utilization in childhood through adolescence. *Journal of Developmental and Behavioral Pediatrics, 30*(6), 481-488. doi: 10.1097/DBP.0b013e3181c35938

- Farmer, E. M. Z., Burns, B. J., Phillips, S. D., Angold, A., & Costello, E. J. (2003). Pathways into and through mental health services for children and adolescents. *Psychiatric Services, 54*(1), 60-66. doi: 10.1176/appi.ps.54.1.60
- Finkelhor, D., Wolak, J., & Berliner, L. (2001). Police reporting and professional help seeking for child crime victims: A review. *Child Maltreatment, 6*(1), 17-30. doi: 10.1177/1077559501006001002
- Fite, P. J., Wynn, P., & Pardini, D. A. (2009). Explaining discrepancies in arrest rates between Black and White male juveniles. *Journal of Consulting and Clinical Psychology, 77*(5), 916-927. doi: 10.1037/a0016626
- Garland, A. F., Lau, A. S., Yeh, M., McCabe, K. M., Hough, R. L., & Landsverk, J. A. (2005). Racial and ethnic differences in utilization of mental health services among high-risk youths. *The American Journal of Psychiatry, 162*(7), 1336-1343. doi: 10.1176/appi.ajp.162.7.1336
- Green, M. T., Clopton, J. R., & Pope, A. W. (1996). Understanding gender differences in referral of children to mental health services. *Journal of Emotional and Behavioral Disorders, 4*(3), 182-190. doi: 10.1177/106342669600400305
- Green, S. P., Shriberg, D., & Farber, S. L. (2008). What's gender got to do with it? Teachers' perceptions of situation severity and requests for assistance. *Journal of Educational & Psychological Consultation, 18*(4), 346-373. doi: 10.1080/10474410802463288
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464-1480. doi: 10.1037/0022-3514.74.6.1464
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*(2), 197.
- Gudiño, O. G., Lau, A. S., Yeh, M., McCabe, K. M., & Hough, R. L. (2009). Understanding racial/ethnic disparities in youth mental health services: Do disparities vary by problem type? *Journal of Emotional & Behavioral Disorders, 17*(1), 3-16.
- Gurwitz, S. B., & Dodge, K. A. (1977). Effects of confirmations and disconfirmations on stereotype-based attributions. *Journal of Personality and Social Psychology, 35*(7), 495-500. doi: 10.1037/0022-3514.35.7.495
- Hogan, M. F. (2003). The President's New Freedom Commission: Recommendations to transform mental health care in America. *Psychiatric Services, 54*(11), 1467-1474. doi: 10.1176/appi.ps.54.11.1467
- Hosterman, S. J., DuPaul, G. J., & Jitendra, A. K. (2008). Teacher ratings of ADHD symptoms in ethnic minority students: Bias or behavioral difference? *School Psychology Quarterly, 23*(3), 418-435. doi: 10.1037/a0012668
- Hox, J. J. (1999). A review of current software for handling missing data. *Kwantitatieve Methoden, 62*, 123-138.

- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York, NY US: Routledge/Taylor & Francis Group.
- Kassirer, J. P., & Kopelman, R. I. (1989). Cognitive errors in diagnosis: Instantiation, classification, and consequences. *The American Journal of Medicine*, 86(4), 433-441.
- Kazdin, A. E., Mazurick, J. L., & Siegel, T. C. (1994). Treatment outcome among children with externalizing disorder who terminate prematurely versus those who complete psychotherapy. *Journal of the American Academy of Child & Adolescent Psychiatry*, 33(4), 549-557.
- Kazdin, A. E., & Weisz, J. R. (2003). Introduction: Context and background of evidence-based psychotherapies for children and adolescents. In A. E. Kazdin & J. R. Weisz (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. 3-20). New York: Guilford Press.
- Kelter, J. D., & Pope, A. W. (2011). The effect of child gender on teachers' responses to oppositional defiant disorder. *Child & Family Behavior Therapy*, 33(1), 49-57. doi: 10.1080/07317107.2011.545013
- Kessler, R. C., Avenevoli, S., Costello, E. J., Green, J. G., Gruber, M. J., McLaughlin, K. A., . . . Merikangas, K. R. (2012). Severity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication Adolescent Supplement. *Archives of General Psychiatry*, 69(4), 381-389. doi: 10.1001/archgenpsychiatry.2011.1603
- Klayman, J. (1995). Varieties of confirmation bias. *Psychology of Learning and Motivation*, 32, 385-418.
- Leaf, P. J., Alegria, M., Cohen, P., Goodman, S. H., Horwitz, S. M., Hoven, C. W., . . . Regier, D. A. (1996). Mental health service use in the community and schools: Results from the four-community MECA study. *Journal of the American Academy of Child & Adolescent Psychiatry*, 35(7), 889-897. doi: 10.1097/00004583-199607000-00014
- Levitt, S. D., & Dubner, S. J. (2005). *Freakonomics: A rogue economist explores the hidden side of everything*. New York, NY: HarperCollins.
- Martin, M. J., McCarthy, B., Conger, R. D., Gibbons, F. X., Simons, R. L., Cutrona, C. E., & Brody, G. H. (2011). The enduring significance of racism: Discrimination and delinquency among Black American youth. *Journal of Research on Adolescence*, 21(3), 662-676. doi: 10.1111/j.1532-7795.2010.00699.x
- Martin, T. W., & Grubb, H. J. (1990). Race bias in diagnosis and treatment of juvenile offenders: Findings and suggestions. *Journal of Contemporary Psychotherapy*, 20(4), 259-272. doi: 10.1007/BF00946040
- Mota-Castillo, M. (2004). Eliminate conduct disorders and ODD...this is the right time! *Psychiatric Times*, 4, 3-5.
- Noltemeyer, A., Kunesh, C., Hostutler, C., Frato, P., & Sarr-Kerman, B. J. . . (2012). The effects of student and teacher characteristics on teacher impressions of—and responses to—student behaviors. *International Education Studies*, 5(4), 96-111.

- Nosek, B. A., & Riskind, R. G. (2012). Policy implications of implicit social cognition. *Social Issues & Policy Review*, 6(1), 113-147. doi: 10.1111/j.1751-2409.2011.01037.x
- Payne, B. K., & Cameron, C. D. (2010). Divided minds, divided morals: How implicit social cognition underpins and undermines our sense of social justice. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. (pp. 445-460). New York, NY US: Guilford Press.
- Pearcy, M. T., Clopton, James R., & Pope, A. W. (1993). Influences on teacher referral of children to mental health services: Gender, severity, and internalizing versus externalizing problems. *Journal of Emotional and Behavioral Disorders*, 1(3), 165-169. doi: 10.1177/106342669300100304
- Pihlakoski, L., Aromaa, M., Sourander, A., Rautava, P., Helenius, H., & Sillanpää, M. (2004). Use of and need for professional help for emotional and behavioral problems among preadolescents: A prospective cohort study of 3- to 12-year-old children. *Journal of the American Academy of Child & Adolescent Psychiatry*, 43(8), 974-983. doi: 10.1097/01.chi.0000127574.41662.fa
- R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Roberts, R. E. (2000). Depression and suicidal behaviors among adolescents: The role of ethnicity. In F. A. Paniagua & A. Yamada (Eds.), *Handbook of multicultural mental health: Assessment and treatment of diverse populations* (pp. 359-388). San Diego, CA: Academic Press.
- Rue, D. S., & Xie, Y. (2009). Disparities in treating culturally diverse children and adolescents. *Psychiatric Clinics of North America*, 32(1), 153-163. doi: 10.1016/j.psc.2008.11.001
- Skiba, R. J., Horner, R. H., Chung, C., Rausch, M. K., May, S. L., & Tobin, T. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review*, 40(1), 85-107.
- Skiba, R. J., Poloni-Staudinger, L., Gallini, S., Simmons, A. B., & Feggins-Azziz, R. (2006). Disparate access: The disproportionality of African American students with disabilities across educational environments. *Exceptional Children*, 72(4), 411-424.
- Social Security Administration. (2016). Popular names by state.
- Srebnik, D., Cauce, A. M., & Baydar, N. (1996). Help-seeking pathways for children and adolescents. *Journal of Emotional and Behavioral Disorders*, 4(4), 210-220. doi: 10.1177/106342669600400402
- Taylor, P. B., Gunter, P. L., & Slate, J. R. (2001). Teachers' perceptions of inappropriate student behavior as a function of teachers' and students' gender and ethnic background. *Behavioral Disorders*, 26(2), 146-151.

- Thompson, R. (2005). The course and correlates of mental health care received by young children: Descriptive data from a longitudinal urban high-risk sample. *Children & Youth Services Review, 27*(1), 39-50. doi: 10.1016/j.childyouth.2004.07.003
- Walter, H. J., Gouze, K., & Lim, K. G. (2006). Teachers' beliefs about mental health needs in inner city elementary schools. *Journal of the American Academy of Child & Adolescent Psychiatry, 45*(1), 61-68. doi: 10.1097/01.chi.0000187243.17824.6c
- Wu, P., Hoven, C. W., Bird, H. R., Moore, R.E., Cohen, P., Alegria, M., . . . Roper, M. T. (1999). Depressive and disruptive disorders and mental health service utilization in children and adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 38*(9), 1081-1090. doi: 10.1097/00004583-199909000-00010
- Yeh, M., McCabe, K., Hough, R. L., Dupuis, D., & Hazen, A. (2003). Racial/ethnic differences in parental endorsement of barriers to mental health services for youth. *Mental Health Services Research, 5*(2), 65-77. doi: 10.1023/A:1023286210205

## Biography

Shelley Alonso-Marsden was born on October 25, 1984, in Colorado Springs, Colorado. She was raised in Denver, Colorado and graduated from Denver's George Washington High School in 2002. Ms. Alonso-Marsden graduated from Northwestern University's Medill School of Journalism in 2006 with a Bachelor of Science in Journalism and psychology major. Ms. Alonso-Marsden's began her graduate education at Duke in 2009 and earned a Master of Arts in Psychology in 2013 with her Major Area Paper, "Disparities in Conduct Problem Symptoms among African American, Latino, and White Children and Adolescents: A Research Synthesis." She completed her Clinical Internship at the University of New Mexico Health Sciences Center in 2015. Her graduate fellowships and honors include: the Duke Endowment Dean's Graduate Fellowship, NIDA Diversity Training Supplement R01 DA016903-07S1, Duke University Summer Research Fellowship, and Society for Research in Child Development (SRCD) Dissertation Research Funding Award. In 2013, Ms. Alonso-Marsden was elected to the Communications and Policy and Student and Early Career Committees of SRCD. She has one first author publication entitled, "Family risk as a predictor of initial engagement and follow-through in a universal nurse home visiting program to prevent child maltreatment," published in *Child Abuse & Neglect* in 2013. Ms. Alonso-Marsden has been awarded the APA/SAMHSA Mental Health and Substance Abuse Services Postdoctoral Fellowship for 2017.