2020

Effects of Race and Gender on Strengths and Difficulties Questionnaire Scores

Amanda Haas
Walden University

Walden University

College of Social and Behavioral Sciences

This is to certify that the doctoral dissertation by

Amanda Haas

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Neal McBride, Committee Chairperson, Psychology Faculty
Dr. Lana Nassen, Committee Member, Psychology Faculty
Dr. Michael Plasay, University Reviewer, Psychology Faculty

Chief Academic Officer and Provost
Sue Subocz, Ph.D.

Walden University
2020

Abstract

Effects of Race and Gender on Strengths and Difficulties Questionnaire Scores

by

Amanda Haas

MS, University of Phoenix, 2011

BA, Arizona State University, 2004

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Clinical Psychology

Walden University

May 2020

Abstract

Researchers can use the Strengths and Difficulties Questionnaire-Teacher Form (SDQ-TF) for ages 4 to 10 as a behavioral screening tool to determine the severity of a child's problematic behaviors; however, SDQ-TF is subject to potential rater biases. Implicit bias refers to people remaining unaware of their biases, which may influence the ways that student behaviors are rated. Therefore, this study was conducted to measure the effect of race and gender congruency on teacher ratings of student problematic behaviors by the total difficulties score using the SDQ-TF 4-10. A nonprobability convenience sample of 98 teachers in the greater Phoenix, Arizona area completed a SDQ-TF for an anonymous student. A two-way factorial analysis of variance with fixed effects was used to determine whether a significant interaction existed between race and gender affecting student SDQ-TF rating scores. The results showed gender incongruency between the teacher and student influenced the total difficulties scores, whereas the race congruency or incongruency between teachers and students had no effect on the total difficulties score. No interaction effect was observed. From a social change perspective, teachers can use these results to recognize potential rater biases, thereby increasing their awareness and becoming more objective while modeling that objectiveness for children and adolescents.

Effects of Race and Gender on Strengths and Difficulties Questionnaire Scores

by

Amanda Haas

MS, University of Phoenix, 2011

BS, Arizona State University, 2004

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Clinical Psychology

Walden University

May 2020

Dedication

This dissertation is dedicated to my daughter, Michigan, and to Daniel. They have motivated me, supported me, and encouraged me throughout these challenging years. Without their support and love, I would not have accomplished everything I have. I love you both so very much. Thank you for sharing this journey with me to achieve my life goal.

Acknowledgments

I would like to thank Dr. McBride and Dr. Nassen, as they had the desire to teach aspiring doctors and lead me on the path to success. They spent endless hours working with me. I am grateful for their support, guidance, and patience throughout this process.

I must recognize my Aunt Linda because, without her help, I would not have been able to finish the program. Additionally, I would like to thank my colleagues, Dr. Stapert, my family, and my friends who have helped me. They have not only provided me with tremendous support, but also they have also helped me while I traveled, provided me with feedback, and cheered me on over the years. Again, I am very thankful and grateful for everyone who encouraged me and helped me.

Table of Contents

List of Tables

Chapter 1: Introduction to the Study

Leaders often ask teachers to complete behavioral assessments or screeners to support the process of diagnosis, treatment, and services a student may need. No person can completely avoid bias (Snowden, 2005), meaning teachers are not exempt from attitudes, beliefs, expectations, and cultural identities that influence their perceptions of the world around them. Researchers have attributed race to skewing teachers' perceptions of problematic behaviors (Goble, Myers, & Pianta, 2016; Downey & Pribesh, 2004; McGrady & Reynolds, 2013) as well as gender (Caldarella et al., 2009; Glock, 2016; Glock & Kleen, 2017; Sargisson, Stanley, & Hayward, 2016; Zwirs et al., 2011). The purpose of this study was to examine whether teacher race congruency and gender congruency impact the total difficulties score using the Strengths and Difficulties Questionnaire–Teacher Form for students ages 4 to 10 (SDQ-TF 4-10). The goal was to understand how race and gender similarities, or differences may influence the behavioral ratings of students when teachers rate the students. When considering any raters' responses on behavioral assessments, psychologists must remain aware of potential rater biases that can compromise the validity of the assessment scores (Mason, Gunersel, & Ney, 2014). In this research, I focused on the SDQ, as understanding potential biases that may influence the use of the SDQ is imperative for gaining accurate behavioral ratings, which can assist in providing accurate treatment as needed.

Background

Research dating back to 1968 has indicated that teacher expectations and student awareness of teacher biases influence a student's achievement, behavior, and success as

well as create self-doubt and anxiety for students (Chang & Demyan, 2007). Research has shown evidence of ethnic and/or cultural bias of teachers toward their students (Mason et al., 2014). For example, teachers have rated African American students as having more problematic behaviors than any other race of students and have used more disciplinary actions (Horner, Fireman, & Wang, 2010; McGrady & Reynolds, 2013; Wright, 2015). Teachers suspend African American students from school three times more than Caucasian students (Scott, Gage, Hirn, & Han, 2019). The racial differences among students and teachers have a significant impact on how teachers perceive behaviors and require further attention in specific locations where race may differ from previous research.

Literature has also shown differences in student treatment based on gender. Research has shown that teachers perceive male students as having higher rates of behavioral problems than female students (Kulinna, 2008; Riley, 2014); boys are under more scrutiny about their expected behaviors and definitions of masculinity (Gilliam, Maupin, Reyes, Accavitti, & Shic, 2016; O'Neil et al., 2007). Additionally, Katz (2017) found that the gender of the student influenced the perceptions of teachers, and Cushman (2010) found that male teachers would hold stronger, more stereotypical gender views compared to female teachers. Therefore, I found gender and race worth researching to determine whether any impact would occur on teachers' behavior ratings of problematic students when using the SDQ.

Problem Statement

The SDQ-TF is under-researched, especially in the United States and compared to the parent and child (self) rating forms (Downs, Strand, Heinrichs, & Cerna, 2012). Thus, research on the SDQ-TF has been limited (Mason et al., 2014), which is a concern because leaders and researchers use teachers as a source of information regarding children's behaviors. There was a gap in the literature from 2015 to 2020, and most current researchers have focused on comparing teacher and parent ratings using the SDQ or focused on preschool-aged children per the current literature search strategy. Because rater bias might be influenced by race and gender (Cushman, 2010), and the SDQ-TF was under-researched, I conducted this study to determine whether teacher biases existed when using the SDQ-TF.

Purpose

The purpose of this quantitative study was to investigate the effect of race and gender congruency on teacher ratings of problematic student behaviors, as measured by the total difficulties score using the Strengths and Difficulties Questionnaire – Teacher Form for students ages 4 to 10 for children ages 4 to 10. The dependent variable was the SDQ total difficulties score. The independent variable, teacher race, had two levels: different or same with the rated student. The independent variable, gender, had two levels: different or same with the rated student.

Teachers who participated in this study completed one SDQ-TF for one student whom they perceived as having behavioral problems in the classroom. Teachers chose a student whom they had worked with for at least 6 months and had first-hand knowledge

about the student's behaviors. Additionally, teachers completed the demographic form (Appendices B-E) to identify the racial and gender congruency or incongruency between the teacher and student, and demographic information for descriptive purposes.

Research Questions and Hypotheses

The general research question guiding this study was the following: Is the SDQ's total difficulties score significantly affected by teachers' race and gender similarity with a problematic student they rated? Consequently, I investigated the following research questions and associated hypotheses:

RQ1: Does teachers' race congruency have a significant effect on rated students' SDQ total difficulties scores?

$H_0$1: Teachers' race congruency has no significant effect on rated students' SDQ total difficulties scores.

$H_1$1: Teachers' race congruency has a significant effect on rated students' SDQ total difficulties scores.

RQ2: Does teachers' gender congruency have a significant effect on rated students' SDQ total difficulties scores?

$H_0$2: Teachers' gender congruency has no significant effect on rated students' SDQ total difficulties scores.

$H_1$2: Teachers' gender congruency has a significant effect on rated students' SDQ total difficulties scores.

RQ3: Is there a significant interaction between teachers' race congruency and teachers' gender congruency, affecting rated students' SDQ total difficulties scores?

$H_0$3: There is no significant interaction between teachers' race congruency and teachers' gender congruency affecting rated students' SDQ total difficulties scores.

$H_1$3: There is a significant interaction between teachers' race congruency and teachers' gender congruency affecting rated students' SDQ total difficulties scores.

Conceptual Framework

A researcher may use several theories to describe what influences people to perceive the world in the manners that they do and whether they are aware of their perceptions. Social psychologists explore the root of perceptions, attitudes, and stereotypes (Kempf, 2020). The biases stemming from these phenomena can appear either positive or negative. For instance, implicit bias refers to people maintaining negative biases in their unconscious minds (Ungvarsky, 2017). The concept of implicit bias refers to a person's unconscious prejudices or negative attitudes about people and things; thus, the person remains unaware and will most likely deny bias if challenged (Ungvarsky, 2017). This finding indicates that people are not in control or intentional in the processes regarding social perceptions, impression formations, and judgments that influence behavior (Greenwald & Krieger, 2006). In the 1970s, the idea that people were unaware of their negative attitudes and prejudices interested social psychologists, and in 1998, researchers developed tests to measure implicit bias (Greenwald, McGhee, & Schwartz, 1998). Researchers have continued to use these tests to continue in a variety of areas to test implicit biases of race, gender, sexuality, age, and weight (Ungvarsky, 2017).

Nature of the Study

For this quantitative study, I used a two-way analysis of variance (ANOVA) to analyze the data. I found this analysis technique appropriate for comparing mean differences between groups and for determining the presence of interaction effects between two independent variables on a dependent variable (see Laerd Statistics, 2013). The independent variables included teacher race and teacher gender as congruent or incongruent in each category. The SDQ total difficulties score was the dependent variable. I used the IBM Statistical Package for the Social Sciences (SPSS) 25 Software to analyze and determine whether there was a main effect for race congruency and gender congruency and whether a significant interaction between race and gender affected student SDQ total difficulties scores.

Parents, teachers, and adolescents can use the SDQ as a behavioral screener to determine the severity of problematic behaviors (Goodman, 1997). The initial normative data that represented the United States only included parents and adolescent self-reports (Brown, Ciara, Bartlett, & Horn, 2006), but the teacher form was later included because they are a significant source of information regarding student behavior. The SDQ consists of 25 questions rating a child's behavior using a 3-point Likert Scale (0 = *Not True*, 1 = *Somewhat True*, 2 = *Certainly True*). There are five subscales of the SDQ: (a) emotional symptoms, (b) peer problems, (c) conduct problems, (d) hyperactivity, and (e) prosocial behaviors. A researcher can use the SDQ to provide a form for the parent and teacher to self-report answers. The parent and teacher use the same form but identify as the parent or teacher. Adolescents can also complete a self-report form. The researcher can separate

the forms into two age groups: 4 to 10 years and 11 to 17 years. The total difficulties score equates to the sum of four subscales, which does not include the prosocial behavior subscale. The total of all five subscales is the impact score (Hawes & Dadds, 2004). In this study, I did not evaluate the impact score; I focused on the total difficulties score, and I only used the form for children ages 4 to 10.

Definitions

Gender: Gender refers to the social and cultural traits related to a person's sex ("Gender," n.d.). I did not confirm the gender of the participant and anonymous student rated; the participant self-reported a gender.

Gender identity: Gender identity refers to a person's internal sense of being male, female, neither, or some combination of those identities ("Gender identity," n.d.).

Race: A group of individuals who share distinctive physical traits and common culture ("Race," n.d.). In this study, the participant self-reported race congruency or incongruency, and I did not verify these results.

Strengths and difficulties total difficulties score: This score equates to the sum of the four subscales of emotional symptoms, peer problems, conduct problems, and hyperactivity (Hawes & Dadds, 2004).

Assumptions

The main assumption of this study was that the SDQ was an exact measure of student behavior. This screener was selected based on previous research showing it as a valid and reliable tool due to consistent and accurate outcomes produced (Goodman, 1997, 2001; Stone, Otten, Engels, Vermulst, & Janssens, 2010). Furthermore, researchers

have identified the SDQ as a valid and reliable tool in several countries, as further supported throughout Chapter 2.

Another assumption was that the participating teachers followed the instructions accordingly, responded openly and honestly when completing the SDQ in its entirety, and did not give responses based on what they defined as socially acceptable. I also expected that the participants had the reading levels and comprehension needed to complete the SDQ correctly, as they had college educations and teacher certifications. I also assumed that participants had implicit biases, as most were unaware of their biases or how such biases could influence the teacher-student relationship.

Scope and Delimitations

I conducted this study due to the scarcity of research on the SDQ-TF and racial and gender biases in the specific population of the greater Phoenix, Arizona area. Thus, it was of interest to expand the research on the SDQ due to its cost and time efficiency as well as its level of reliability and validity in European populations. Teachers can use the findings to understand if biases may influence the SDQ-TF, which can assist in raising awareness of their potential biases and helping them to rate behaviors objectively in the future. However, the data cannot be generalized beyond the greater Phoenix, Arizona area.

This study included 98 teachers who worked with children aged 4 to 10 in the greater Phoenix, Arizona area. The purpose of including these boundaries for participants was to use the SDQ-TF for ages 4 to 10 elementary age students as a subgroup of the school system. Including a wider range of students might have skewed the data because a

junior high and high school student's behaviors are developmentally different than elementary-aged students. Given the research questions, the parent and self-forms of the SDQ were not used.

I selected the SDQ to study because it was in the public domain and free to use. Furthermore, the SDQ did not require training or licensure to administer or score. A researcher can administer the SDQ and have it completed in 5 minutes versus other tools that could take significantly longer, require training, and require purchase. Given the nature of this study's focus on teachers, the SDQ-TF was ideal for respecting participants' time while collecting similar data in comparison to lengthier tools.

Limitations

Limitations of the study included the time at which data were collected, as student behavior might fluctuate depending on circumstances. The results of this study also do not have the capacity to inform treatment recommendations for the rated students, as the SDQ was only to be used in conjunction with a full assessment battery administered by a licensed professional. After, clinically significant behaviors that warrant treatment can be identified. The results can also not be generalized beyond the greater Phoenix, Arizona area because Phoenix, Arizona consisted of cocultures and subcultures within schools and school districts within the county, state, and country that had unique characteristics. In Arizona, the major cocultures are Hispanic and Native American, although many other cocultures were present in the study. In other regions of the United States, the coculture presence might be different, which might affect the strength or weakness of racially induced biases.

The study findings might also miss nonnormally distributed phenomena, although a two-way ANOVA is robust when considering skewed distributions (Laerd Statistics, 2013; Schmider, Ziegler, Danay, Beyer, & Buehner, 2010). Future researchers might find a larger sample useful. Lastly, the race and gender congruency or incongruency of the participant and the student they rated was solely documented by the self-report of the participant and was not confirmed; thus, some data could have been missed or documented incorrectly because of the participant's perception.

Significance

Researchers have shown that SDQ outcomes predict clinically significant conduct problems in populations outside the United States. Given the tool's cost and time efficiency, researchers have suggested its potential value in U.S. populations as worth exploring (Goodman, Renfrew, & Mullick, 2000). Significant findings of this study confirmed the use of the SDQ, which could lead to the improvement of identifying and treating problem behaviors in children as well as raise awareness of teacher bias. Increased use of the SDQ may supply a time and cost-effective screening tool for clinicians to treat children for conduct problems earlier in life and prevent long-term issues. I contributed to the positive social change of recognizing biases to become more objective when rating student behavior. From a larger social change perspective, if teachers become more aware of their biases when interpreting student behavior and make changes to become more objective, they will model a more objective stance toward others for their students.

Summary

I used the SDQ-TF to show whether teacher race and gender congruency had any effect on the SDQ-TF total difficulties scores. Exploring the use of the SDQ-TF in a diverse population adds to the existing research of the SDQ-TF. Future researchers may broaden their scope of schools within the United States and may build on this study to educate teachers about the influence of biases in their perceptions of students.

The following chapter will include an extensive review of existing research about the SDQ, including its application in other populations and the teacher forms. Chapter 3 entails a discussion of the research methods for this study. Chapter 4 contains the results of the data analysis, and Chapter 5 includes an interpretation of the data analysis findings, the limitations of the study, changes or direction for future study given the current findings, and the implications of the current findings on society.

Chapter 2: Literature Review

Introduction

The purpose of this quantitative study was to investigate the effect of race and gender congruency on teacher ratings of problematic student behaviors, as measured by the total difficulties score using the Strengths and Difficulties Questionnaire – Teacher Form for students ages 4 to 10 for children ages 4 to 10. The research on the SDQ-TF has been limited (Mason et al., 2014), and the existing research on the SDQ has not shown any position on the potential for rater biases to influence the SDQ outcome scores when teachers rated children with possible behavioral issues within the United States. However, researchers in New Zealand have found racial biases among teachers who used the SDQ-TF to rate students of different races than themselves (Zwirs et al., 2011). Because the SDQ has been shown to be predictive of significant behavioral and emotional disorders (Becker, Hagenberg, Roessner, Woerner, & Rothenberger, 2004; Goodman et al., 2000; Goodman, Ford, Corbin, & Meltzer, 2004; He, Burstein, Schmitz, & Merikangas, 2012), researching potential rater bias was imperative for expanding the knowledge of the SDQ-TF.

This chapter includes the strategy used to collect established research supporting the purpose of this current research. The terms, search engines, and combinations of terms used to discover existing research are presented. Following the research strategy, a discussion of implicit bias and its relationship with this research is offered. Lastly, this literature review includes the existing data related to the SDQ, the impact race and gender

have on existing assessment tools, and the influence teachers have on evaluating student behavior.

Literature Search Strategy

In a preliminary search of the literature for this research, the most commonly used research databases included PsychINFO and PsychARTICLES. Other searches included Google Scholar, which often led to Walden University's library of peer-reviewed articles. I also used national websites to access current rates of special education services and the cultural diversity of Arizona. The search terms and combinations of search terms used to retrieve current literature included the following: *Strengths and Difficulties Questionnaire* (e.g., *SDQ*), *SDQ teacher form; the Rutter Children's Behaviour Questionnaire; the Conners-Third Edition assessment (Conners 3); and the Child Behavior Checklist (CBCL); teachers (teacher form) with United States, validity, reliability, race,* and *gender; rater bias* and *behaviors; teacher(s) with race, gender, bias, implicit bias,* and *problematic behaviors; implicit bias with teachers, student behaviors, rating behaviors, behavioral assessment,* and *bias*; and *interracial bias*.

Regarding the research for this study, there was no limit to the age of the data. The SDQ was introduced in 1997; since then, data were reviewed to establish consistency to support claims. Most of the literature reviewed consisted of peer-reviewed articles. Some of the literature included excerpts from books and national websites.

Conceptual Framework

This research was driven by social psychology's implicit bias theory, which indicates that people maintain biases unconsciously (Ungvarsky, 2017). Researchers

became interested in this phenomenon of bias and defined explicit biases as biases conscious to the beholder (Greenwald & Krieger, 2006), which indicated that because people remained aware of their personal biases, they intentionally behaved in a manner supportive of their beliefs. Researchers then argued that biases might not be as explicit as expected; rather, people would likely remain unaware of the biases they maintained (Ungvarsky, 2017). Most often, people will not admit to biases even when challenged because they either do not want to admit to such behavior, or they are unaware of their biases; thus, unconscious or implicit biases are maintained (Greenwald & Krieger, 2006). Additionally, although many people are found to be unaware of implicit biases, some do not believe their implicit biases will have a negative impact on how they treat others (Avant, Weed, Connelly, Hincapie, & Penn, 2018).

One common myth is that explicit biases prevent people from displaying their biases because they are aware of these biases and would not act in a manner indicating they are biased. However, implicit biases may not align with explicit biases, and even those who maintain equality can unknowingly act in a manner reflecting their implicit biases (Staats, 2016). Therefore, no one is immune from implicit biases, which further supports the theory that humans are not completely aware of their biases, regardless of their ages, races, genders, or ethnicities. Additionally, implicit biases occur in most professions and throughout several domains in life, including education, health care, law and emergency buildings, and the U.S. judicial system (Staats, 2016).

The most difficult aspect of implicit bias is how to measure something that is intangible. Therefore, researchers developed the Implicit Association Test (IAT) to

measure the reaction time of the tester who would categorize a series of words of pictures into groups (Jolls & Sunstein, 2006). Researchers have shown the IAT as a reliable and good predictor of behavior as well as less likely to be influenced by the self-report responses subjected to self-presentation (Cunningham, Preacher, & Banaji., 2001; Steffens, 2004).

Researchers have defined the implicit bias of prejudice attitudes as a stronger predictor of teacher's expectations for students than explicit bias (Van den Bergh, Denessen, Hornstra, Voeten, & Holland, 2010). In children as young as preschool age, teachers perceive behaviors based on the children's races (Yates & Maccelo, 2014). Researchers have found that implicit bias can influence the ways that teachers view their students and can cause harsher discipline for African American students (Okonofua & Eberhardt, 2015; Westerberg, 2016). For example, in a 2018 study, researchers asked preservice teachers to identify the emotions and judge hostility of videos and pictures of African American and Caucasian males (Halberstadt, Castro, Chu, Lozada, & Sims, 2018). The results showed that the African American videos and pictures were less likely to be accurately perceived. African American faces were more likely to be perceived as angry when they were not angry, and their misbehaviors were perceived as more hostile than the Caucasian faces. Moreover, preservice teachers recognized more hostility in African American males than Caucasian males across varying levels of behavior severity. Thus, the severity of the behavior did not appear to be the cause for rated hostility, but students' race impacted the perceptions of teachers (Halberstadt et al., 2018).

Further, researchers have found the implicit biases of teachers to impact student achievement, causing a gap in achievement among African American and Caucasian students (Mortenson, 2018). This issue could become harmful because the teacher's lack of awareness of his or her perceptions influences his or her reactions or interpretations of the child's behaviors, influencing the child's development in all areas of life (Westerberg, 2016). Several studies have shown ways that teachers' unconscious attitudes have negatively impacted their interpretations of students' behaviors, especially when the student was of a different race. For instance, Kozlowski (2015) found African American and Hispanic students rated their efforts higher than what their teacher perceived their efforts as such, and these students were less likely to receive positive ratings even when admitting their efforts were lacking. Thys and Van Houtte (2016) also found teacher expectations for minority students were low.

Research has also shown that the race and gender of the teacher has an impact on implicit bias toward students. Gilliam et al. (2016) found that teachers viewed a video containing African American boys longer than any other races when they were primed to expect challenging behaviors, and the teachers' implicit biases differed by the teachers' races. Additionally, the level of implicit bias was lower when teachers and students were of the same races and higher when the students and teachers were of different races. Further, implicit bias, measured as a function of gender, indicated that teachers believed boys required more attention than girls, and these boys were at greater risk of being removed from the classroom.

Leaders have implemented training and education for teachers; however, such efforts appear hindered by avoidance, misunderstanding, or lack of training to curb the impact of implicit biases on children's behaviors (Whitford & Emerson, 2019). Additionally, leaders who implement education programs have considered that teachers received these pieces of training and education before becoming a teacher. However, researchers have shown interventions to impact teacher bias, as Whitford and Emerson (2019) used the IAT and Race IAT with a brief intervention designed to solicit empathy for African American students. Whitford and Emerson gave the IAT before and after the intervention and found that Caucasian female preservice teachers' IAT scores were significantly decreased after the intervention was presented. Therefore, the intervention appeared to decrease implicit biases and should be further explored.

Another concern with teacher biases is that whether implicit or explicit, biases lead to other forms of bias, such as confirmation bias. Confirmation bias refers to a person's tendency to discount information that does not match predetermined beliefs or judgments (Kappes, Harvey, Lohrenz, Montague, & Sharot, 2020). Although a review of confirmation bias was beyond the scope of this research, the halo/horns effect is a similar concept that applied to the study, which refers to a person positively or negatively perceiving behaviors of another based on the positive or negative attitudes that the rater upholds (Hoyt, 2000). In 1976, Foster and Ysseldyke found that teachers have negative expectations for children labeled as emotionally disturbed, learning disabled, or mentally retarded (i.e., intellectually disabled). Foster and Ysseldyke found that teachers maintained negative expectations for these children despite the child demonstrating

positive behaviors that were inconsistent with the label. Shifrer (2013) also found this phenomenon, as teachers and parents of students labeled with a learning disability had lower educational expectations for their students compared to adolescents who behaved well and were not labeled with a disability.

Additionally, rater bias may cause errors in behavioral ratings, which may cause errors when trying to identify psychiatric disorders in children and adolescents. In the classroom, the halo effect can be problematic because teachers may make inaccurate conclusions about their students and do not adequately address their needs (Lasky, 2015). Current researchers have continued to support the horns effect as teachers maintain negative expectations toward children labeled with behavioral, emotional, or cognitive limitations (Allday, Duhon, Blackburn-Ellis, & Van Dycke, 2011). Additionally, researchers have supported the halo effect on teachers' evaluations of student behaviors (Álvarez-García, García, & González-Castro, 2014). Teachers have faced errors when rating a student's behaviors because a well-behaved student was likely perceived as smart, intelligent, and engaged before the teacher objectively measured the student in these areas (Rasmussen, 2008). Likewise, a child with attention-deficit/hyperactivity disorder may disrupt the classroom; thus, a teacher may judge this student as not putting in effort or acting inappropriately (Rasmussen, 2008).

Literature Review Related to Key Concepts

The Strengths and Difficulties Questionnaire

The SDQ refers to a screening assessment developed by Goodman (1997) that was designed for use in conjunction with other assessments to address behavioral and

emotional problems. Researchers can use the SDQ to evaluate a child or adolescent's behaviors (ages 4 to 17) and identify any problem areas using five subscales: (a) emotional symptoms, (b) peer problems, (c) conduct problems, (d) hyperactivity, and (e) prosocial behaviors. The SDQ consists of three individual report forms: (a) the parent report form, (b) the teacher report form, and (c) a self-report form. In this study, the form included a 3-point Likert scale related to 25 questions. The total of all five subscale scores yields the impact score (Hawes & Dadds, 2004), and the total difficulties score is the sum of Scales 1 through 4. I scored the total difficulty scores ranging from 0 to 11 as normal, 12 to 15 as borderline, and 16 to 40 as abnormal.

The SDQ remains in the public domain at http://www.sdqinfo.com. The forms remain free to use and do not require any training to administer, complete, or score (http://www.sdqinfo.com). Researchers have translated this screener into over 60 languages and used it widely in many western countries (Ruchkin, Koposov, Vermeiren, & Schwab-Stone, 2012). Many researchers have focused on Western countries with little research on American samples and the teacher form (Downs et al., 2012). However, the SDQ was normed on children in the United States. The sample consisted of 9,878 children who completed the self-report form, and one parent completed the parent form for each child. These norms were based only on data collected from parents.

Validity. Researchers have found the SDQ, overall, as valid and reliable (Goodman, 1997, 2001; Stone et al., 2010). In a meta-analysis of 48 studies, Stone et al. (2010) determined that emotional symptoms were correctly identified in 36% of the cases, and 93% of children with conduct problems were correctly identified.

Additionally, according to Goodman et al. (2000, 2004), teachers' ratings were sensitive to psychiatric disorders 82.2% of the time. Given these statistics, researchers assumed that the SDQ could predict diagnosable behavioral disorders in children (Becker, Hagenberg, et al., 2004; He et al., 2012). Although the SDQ total difficulties score has been found predictive of a behavioral disorder when using multiple informants (Goodman et al., 2000), the SDQ alone cannot be used for diagnostic purposes; rather, it has been used in conjunction with a full evaluation. Furthermore, the SDQ multi-information approach is recommended as children's behaviors may vary among environments and require multiple reports for screening behaviors.

In addition to the SDQ, researchers have developed the Development and Well-Being Assessment (DAWAB) for the *International Statistical Classification of Diseases and Related Health Problems* (*ICD-10*) and *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-IV*) diagnoses of mental health disorders in children ages 5 to 17 (Goodman et al., 2000). Children identified as having a psychiatric disorder by the DAWAB have been correctly classified as having a disorder in 77.3% of the cases using the SDQ algorithm (Goodman et al., 2000). Furthermore, when the DAWAB indicated a child had hyperactivity, 91% were rated as probable for a hyperactive disorder using the SDQ; of those identified as having a conduct-oppositional disorder by the DAWAB, 44% were rated as probable using the SDQ (Hysing, Elgen, Gillberg, Lie, & Lundervold, 2007).

In comparison to other behavioral screening instruments such as the CBCL, Becker, Hagenberg, et al. (2004) indicated that the German CBCL, CBCL-Teacher

Report Form (CBCL-TRF), and the SDQ-TF could differentiate between children with and without clinically significant symptoms for clinical diagnosis; thereby supporting its validity. The SDQ total difficulties scores among teachers, parents, and the self-report form positively correlated with diagnostic criteria for a mental health disorder (Goodman, 1997; Goodman, Ford, Simmons, Gatward, & Meltzer, 2003; Goodman & Goodman, 2011) with similar results found in Australia (Mathai, Anderson, & Bourne, 2004). Furthermore, in a meta-analysis, the weighted correlation for teacher ratings between the SDQ and CBCL totals was 0.76 (Stone et al., 2010).

Additionally, when compared to the Rutter questionnaires (Rutter et al., 1974), Goodman (1997) found the total scores of the SDQ to correlate, indicating concurrent validity highly. Goodman found a high correlation ($r = 0.92$) between the SDQ-TF total difficulties score and the Rutter total deviance score. Gowers et al. (1999) also found a moderate correlation ($r = 0.46$) between the SDQ-TF total difficulties score and the Health of the Nation Outcome Scales for Children and Adolescents.

Reliability. Regarding the SDQ in the United States, Brown et al. (2006) collected normative data using the SDQ parent form for 9,878 children. Results indicated the total difficulties score was internally consistent (Cronbach's $\alpha = 0.83$), and the impairment scales were consistent (Cronbach's $\alpha = 0.80$; Brown et al., 2006). The subscales of conduct problems, emotional symptoms, hyperactivity, and prosocial behaviors also showed good internal consistency (Cronbach's $\alpha = 0.63$ to 0.77); however, the peer problems subscale presented with poor consistency (Cronbach's $\alpha = 0.46$). But Goodman, Meltzer, and Bailey (1998) indicated the internal consistency, Cronbach's α,

for subscales and total difficulties scores as follows: 0.82 for total difficulties, 0.75 for emotional symptoms, 0.72 for conduct problems, 0.69 for hyperactivity, 0.65 for prosocial behaviors, and 0.61 for peer problems. However, Palmieri and Smith (2007) described a couple of the subscales as having uncertain levels of internal consistency.

In a French sample, d'Acremont and Van der Linden (2008) found some concerns about the peer problems subscale because its reliability was .64, which was not a large issue; however, when compared to all other subscales, was the lowest result found. Capron, Thérond, and Duyme (2007) researched the confirmatory factor analysis of the SDQ in a French sample and concluded the SDQ-TF accurately measured mental disorders in youths. The researchers suggested that others continued to use the SDQ as a screening tool.

Downs et al. (2012) cited multiple researchers of the psychometrics of the SDQ in Europe, Asia, United States, and Australia, and Downs et al. concluded that the overall performance of the SDQ measured well when used by parents and teachers for children ages 4 to 16. Additionally, although inter-rater agreement between parents and teachers using the SDQ was low (Fält, Wallby, Sarkadi, Salari, & Fabian, 2018), the SDQ had shown satisfactory test-retest reliability and internal consistency for 4 to 12-year-olds (Stone et al., 2010), indicating the discrepancies between parents and teachers might be due to environmental factors on behaviors. Moreover, Fergusson, Boden, and Horwood (2009) posited that most children had situation-specific behaviors, and only a minority had generalized conduct problems in multiple environments. Thus, this finding could be the cause of the differences between mother and teacher reports of behaviors (Fergusson

et al., 2009). Cheng et al. (2018) found that parents and teachers agreed more about externalizing behaviors than internalizing behaviors, and agreement on externalizing and internalizing behaviors varied among seven European countries. With its general level of reliability and validity, researchers can expand the SDQ for understanding factors that may contribute to individual questions, subscales, and total difficulties scores.

In a Chinese sample, Du, Kou, and Coghill (2008) found that the peer problems subscale to have the lowest Cronbach's α among the parent, teacher, and self-report versions of the Chinese SDQ. Regarding the Chinese teacher form, the prosocial behaviors and hyperactivity/inattention subscales had the highest reliability at α 0.83 and 0.82, respectively. Despite some concerns about the reliability of the peer problem subscale, internal consistency and test-retest stability was stronger for teachers than for the self-report form for adolescents aged 5 to 15. Du et al. posited these concerns could be addressed by examining the wording and meaning of the questions as translated from English to Chinese. As recently as 2013, Liu et al. defined the Chinese SDQ parent, teacher, and self-report forms as reliable and valid for use among Taiwanese children and adolescents.

Additionally, Du et al. (2008) researched the SDQ teacher and parent in China. Du et al. defined the SDQ as reliable and valid for use in Chinese culture. The SDQ was also found reliable in comparison to other behavioral screening tools, such as the Rutter Questionnaire (Rutter et al., 1974), CBCL, and Conners in a Chinese population (Du et al., 2008). Du et al. (2008) noted the helpfulness of the SDQ being related to only

including 25 questions; thus, the tool was shorter than any other behavioral screening tool while remaining reliable in comparison to longer behavioral assessments.

Stone et al. (2010) found the German translation of the SDQ-TF to correlate with the CBCL-TRF highly, just as Du et al. (2008) found. Stone et al. (2010) found that the SDQ-TF had strong psychometric properties and posited the teacher form aids in the multi-informant approach for screening child and adolescent behaviors. However, Niclasen Skovgaard, Andersen, Sømhovd, and Obel (2013) found that higher internal consistency for older children rated by teachers than younger children rated by their parents in a Danish sample. In Russia (Ruchkin et al., 2012) and Sweden (Björnsdotter, Enebrink, & Ghaderi, 2013), researchers defined the SDQ teacher form as a useful screener among children and adolescents.

Other factors of reliability and validity. Thus far, the research on reliability and validity has consisted of relatively homogeneous populations. Race and ethnicity were not mentioned in these studies as a contributing factor to SDQ scores. However, Mieloo et al. (2014) found significant differences between Dutch and non-Dutch participants using the SDQ teacher and parent form. Mieloo et al. explained that the total difficulties scores were valid and reliable within different ethnic groups; however, some differences were noted with reliability and validity, which made interpretation of the total difficulties scores harder for some ethnic groups. This finding indicated that this current study would show racial differences as a factor in SDQ scores.

Some researchers have suggested the teacher form was not culturally sensitive because certain questions of the SDQ were not pertinent to the Italian culture (Tobia,

Buckwalter, & Stich, 2013); conversely, Mieloo et al. (2014) defined the teacher form as valid for different ethnic groups within a Dutch population. Research conducted in the Netherlands (van den Heuvel et al., 2017), Japan (Shibata, Cattaneo, Leach, & Galloway, 2014), and Russia (Ruchkin et al., 2012) showed the teacher form as reliable and valid in identifying behavioral and emotional problems in elementary school children.

Overall, researchers have defined the SDQ as useful. Research has indicated the SDQ and CBCL are highly correlated when used by parents (Goodman & Scott, 1999; Goodman et al., 2000). Goodman (1997) found the SDQ highly correlated with the Rutter Questionnaire (Rutter et al., 1974), and Du et al. (2008) indicated the SDQ highly correlated with the Connors.

Race and Gender Biases When Using Behavioral Assessments

Using the CBCL-TRF, Berg-Nielsen, Solheim, Belsky, and Wichstrom (2012) found that preschool teachers might have same-gender biases with female children; female teachers rated female students lower in externalizing behaviors than male students. As discussed later in this chapter, this finding was consistent with the research that female students were typically rated lower in externalizing behavioral problems than their counterparts. Additionally, researchers of the CBCL-TRF found teacher biases of student behavior likely influenced the significant differences between parental and teacher evaluations, but these researchers did not study what factors might lead to teacher biases (Glaser, Kronsnoble, & Forkner, 1997).

Regarding the Conners 3, researchers have focused on the agreement of a child's behaviors between the teacher and parent and validating the Conners 3 in different

cultures. However, in 1998, Epstein, March, and Conners found that African American students were more likely to have been rated as antisocial, while Caucasian female students were rated as inattentive. Epstein et al. (1998) found that African American children were rated higher than Caucasian children on externalizing behaviors across all genders. Regarding the Rutter's Questionnaire, one study showed teachers rated Western Indian immigrant children as having more behavioral problems than Caucasian students in London, England (Rutter et al., 1974). Research on the original Behavioral Assessment System for Children also showed similar evidence that Caucasian teachers perceived Caucasian students more positively than African American students, as well as female teachers, rated female students more positively than male students (Rong, 1996).

More recently, using the Eyberg Child Behavior Inventory (ECBI), African American children in preschool had a greater likelihood of being rated by their teachers as having a behavioral problem, while preschool females were rated lower for behavioral problems by teachers (Munzer et al., 2018). Because parent and teacher ratings of these preschool students differed significantly, Munzer et al. (2018) suggested that this issue might continue to support the therapy of implicit bias. Furthermore, this finding was consistent with existing research that teachers would interpret the behaviors of African American children as more troublesome, which would lead to harsher discipline.

Regardless of which behavioral assessment is used, teachers have more negative perceptions of students based on their genders and races. Overall, researchers have examined the factors of race and gender in relation to behavioral questionnaires and demonstrated that race and gender could positively or negatively impact the behavioral

ratings. This finding poses a threat to the children's development academically, behaviorally, and personally (Westerberg, 2016).

Teachers as a Source of Information About Student Behaviors

The American Psychiatric Association (2013) indicated that for several childhood behavioral diagnoses, the criteria for a behavioral or emotional disorder must remain present in several environments because a child's behaviors must stay consistent in more than one environment. Other researchers have concurred that multiple informants must be used when identifying behavioral or other psychiatric disorders (Goodman et al., 2004; Hunsley & Mash, 2007; Stone et al., 2010). Therefore, parents, teachers, counselors, and social workers who have consistently worked with and observed the child are potential informants to report on observable behaviors.

Professionals can use the multi-informant approach to determine the consistency of behaviors in different contexts to show ways that leaders should apply interventions to the person, especially for children with conduct problems (De Los Reyes et al., 2015; Dirks, De Los Reyes, Briggs-Gowan, Cella, & Wakschlag, 2012; Stone et al., 2010). Professionals qualified to diagnose and make treatment recommendations rely on multiple informants; however, when he or she notes discrepancies among informants, the professional must determine whether those discrepancies occurred due to rater bias or the person's behaviors are exclusive to only certain environments (De Los Reyes, 2011).

Researchers can use teachers' perceptions of behaviors to measure the severity of the behaviors through behavioral rating scales considered standard practice for assessment (Mason et al., 2014). Several common behavioral assessments completed by

teachers include the CBCL-TRF, Conners 3, The Rutter Questionnaire (Rutter et al., 1974), Behavioral Assessment System for Children, Second Edition, Behavioral and Emotional Screening System, Social Skills Improvement System-Performance Screening Guide, the Student Internalizing and Externalizing Behavior Screeners, and the Student Risk Screening Scale-Internalizing and Externalizing. Researchers of these assessments have used teachers as a source of gathering information because the multi-informant approach is empirically supported and criteria for diagnostic purposes.

Hodgins, Larm, Ellenbogen, Vitaro, and Tremblay (2013) found elementary school teachers were good predictors of criminal behaviors into adulthood when rating student conduct problems and hurtful and uncaring behaviors. As early as age 6, student behavioral ratings by teachers predicted criminalistic behaviors in adolescence and adulthood with the most significant prediction found among 10-year-olds (Hodgins et al., 2013). In relation to the SDQ, teachers' ratings of students' behaviors are relatively like the ratings by the students' parents (Becker, Woerner, Hasselhorn, Banaschewski, & Rothenberger, 2004; Sargisson et al., 2016). Therefore, teachers appear as a source of information when gathering evidence about student behavior.

Most importantly to this research, the SDQ included a teacher form and was recommended to use with the parent, child (if age appropriate), and teacher to obtain enough information to determine the severity of the child's behaviors. Although the SDQ forms were labeled as "P or T," an adult could complete this form with knowledge about the child's behaviors, which might contribute to identifying areas of concern.

Teacher Racial and Gender Biases on Student Behavior

Regarding gender and race factors when using the SDQ-TF, researchers conducted one study in 2011 among elementary teachers in the Netherlands. Zwirs et al. (2011) researched the ethnically diverse population in the Netherlands to determine if race and gender affected the SDQ outcomes. This study consisted of 2,185 children aged 6 to 10 of the ethnic groups, including Dutch, Moroccan, Turkish, and Surinamese. Results showed the mean scores on the emotional problems, hyperactivity, conduct problems, and prosocial behavior subscales varied significantly among ethnicity and gender (Zwirs et al., 2011). Furthermore, as discussed later in this chapter, males and females differed between peer problem scores but not ethnicity.

Racial differences. Researchers have noted racial differences can influence multiple aspects of a person's life. Race or ethnicity of the teacher and child are important to the teacher's judgments about student behaviors (Downer et al., 2016). In relation to the SDQ, Zwirs et al. (2011) found that racial differences between students and teachers impacted the SDQ ratings. Zwirs et al. found that Turkish immigrant children in the Netherlands were rated differently between Dutch and Turkish teachers, whereas Dutch and Surinamese immigrant teachers interpreted problematic behaviors among Dutch, Moroccan, Turkish, and Surinamese children similarly.

Researchers have found that students' races influence students behaviorally, academically, and emotionally. Researchers have found that African-American students are more likely to be poorly rated by their teachers and to experience harsher discipline when their teacher is of a different race (Horner et al., 2010; Wright, 2015); students are

often rated as less disruptive when the teacher is of the same race (Downer et al., 2016; Downey & Pribesh, 2004; McGrady & Reynolds, 2013; Munzer et al., 2018). These researchers have reported the general conclusion that Caucasian teachers rate African American students' behaviors poorly versus Caucasian students' behaviors. When an African-American student is aware of a teacher's biases, the student is more likely to act in a counterproductive manner when facing social rejection, may display inadequate anger control, and may tend to act aggressively (Thomas, Coard, Stevenson, Bentley, & Zamel, 2009). Races were the most significant factor found among African American students who were more highly disciplined than Caucasians, Hispanics, or other races (Horner et al., 2010), thus indicating the African American student acts out more so than other students.

Research indicated Caucasian teachers' perceptions of Hispanic students generally did not differ from their perceptions of Caucasian students; however, Caucasian, African American, and Hispanic students have better behavioral ratings when assessed by a teacher of the same race than by a teacher of a different race. African American teachers' ratings of students' behaviors did not differ between Caucasian and African American students, but Caucasian teachers rated African American students lower than Caucasian students in social skills, leadership skills, and social desirability. Teachers viewed overcontrolled behaviors of the Asian students as more typical and less typical of the Caucasian student (Chang & Sue, 2003); however, Asian students were perceived more positively than Caucasian students (McGrady & Reynolds, 2013).

Researchers have assumed cultural awareness is a contributing factor influencing the perceptions of child behaviors by teachers. Thomas et al. (2009) indicated that students who lacked knowledge about their cultural heritage were found to have higher rates of problematic behaviors when rated by their teachers, while students with personal cultural awareness were less likely to be reported as having problematic behaviors. However, Mason et al. (2014) argued, in their review of 13 studies, that generally mixed evidence about teacher biases existed, but teachers might demonstrate racial biases when the student violated positive stereotypes.

More favorable assessments of student's behaviors occurred when the teacher and students were of the same race (Downey & Pribesh, 2004; Gilliam et al., 2016; Ouazad, 2014). Although fewer minority teachers exist in the field, teachers are positive role models for students of the same race (Egalite, Kisida, & Winters, 2015). Yet, Bradshaw, Mitchell, O'Brennan, and Leaf (2010) found racial similarity among African American teachers, and students did not reduce the number of discipline referrals African American students received. In a Texas study, the racial similarity of the teacher and student dropped discipline referrals by 13.5%, and the researchers found that when a student was enrolled in a school where most teachers shared the same race or ethnicity, discipline referrals decreased by 11.2% (Blake et al., 2016).

Another influential factor in the perception of behaviors is the time of school year when a student is rated. Researchers determined that the behavior at the beginning of the school year showed no significant differences depending on the congruency or incongruency between the student and teacher; however, by the end of the school year,

Caucasian teachers identified African American students as having more challenging behaviors (Gilliam et al., 2016). Some opined that African American teachers in comparison to Caucasian teachers had an advantage of understanding African American children culturally; their perceptions of the behaviors were not viewed as having increased throughout the school year for African American students (Downer et al., 2016; Gilliam et al., 2016); this finding indicated that teachers' races would influence their perceptions of child behaviors (Saft & Pianta, 2001).

African American students make up the majority of behavioral referrals and receive harsher discipline in comparison to Caucasian students (Gilliam et al., 2016; Fenning & Rose, 2007), indicating African American students have more behavioral problems. However, African American students are overrepresented in the statistics because the number of referrals is not proportionate to the number of students (McFadden, Marsh, Price, & Hwang, 1992). In Arizona, in the 2011 to 2012 school year, of 61,700 teachers, 80.1% identified as Caucasian, 13.1% identified as Hispanic, 2.8% identified as African American, and 1.7% identified as Asian (National Center for Education Statistics, 2012). Regarding students, in the 2015 to 2016 school year, Arizona had approximately 45% Hispanic students, 39% Caucasian students, 5% African American students, and 4.5% Native American Students (National Center for Education Statistics, 2016).

When children show severe behaviors that disrupt their learning, special education services or accommodations may be required to assist the student. Therefore, a psychiatrist, licensed psychologist, licensed professional counselor, licensed clinical

social worker, or certified school psychologist should evaluate students (Arizona Department of Education, 2018a) to determine the extent to which the student's behaviors are adversely affecting their education. Emotional disturbance refers to one of the disability categories that may qualify a student for special education services or accommodations.

The Arizona Department of Education (2018a) defined emotional disturbance as a condition with one or more of the following characteristics over a long period that could negatively influence a student's education: (a) an inability to learn that cannot be explained by intellectual, sensory, or health factors; (b) an inability to build or maintain satisfactory interpersonal relationships with peers and teachers; (c) inappropriate types of behavior or feelings under normal circumstances; (d) a general pervasive mood of unhappiness or depression; and (e) a tendency to develop physical symptoms or fears associated with personal or school problems. Some of the behaviors in students with emotional disturbance include aggression, hyperactivity, withdrawal, immaturity, and learning difficulties (Center for Parent Information & Resources, 2017). Specifically related to this research, conduct disorder was an emotional disturbance that could qualify a student for special education services.

In the 2018 to 2019 school year, Arizona had slightly over 1 million students reported as of October 1, 2018 (Arizona Department of Education, 2018b). Arizona's population from the 2016 to 2017 school year was consistent with the population report from 2016 to 2017 when the U.S. Department of Education last released data. Because the data collected from Arizona during the 2018 to 2019 school year remained consistent

with the 2016 to 2017 school year, the statistics from the U.S. Department of Education were reported and assumed relatively consistent despite the last time data were released for 2017. Approximately 130,000 (U.S. Department of Education, 2017a) of 1.1 million students ages 3 to 17 and 5,000 students ages 18 to 21 (U.S. Department of Education, 2017a) in the state of Arizona received special education services. Of students ages 3 to 5, 27 received services for emotional disturbance (U.S. Department of Education, 2017b). The data regarding race of students who received special education services for emotional disturbance were collected for students age 6 to 21, and the data indicated 11.21% were African American, 8.45% were two or more races, 7.88% were Caucasian, 4.87% were Native American, and 4.03% were Hispanic (U.S. Department of Education, 2017c).

Gender differences. The gender of a teacher may affect students in many ways. Teachers may reinforce stereotypes by acting as role models and maintaining biases related to gender (Sansone, 2017). Student performance is influenced by the teacher's gender because the student may perform better with a same-sex teacher, and students internalize the expected negative stereotypes about their genders, thus causing their academic performances to fluctuate (Sansone, 2017). The teacher's own gender biases can affect how he or she treats and evaluates students. Pellegrini et al. (2011) suggested children's behaviors were stereotyped based on the child's gender; male children were reported as more aggressive than females. Additionally, Pellegrini et al. did not generally address the gender of the rater.

Friedman (1995) found male teachers were more sensitive to students with attention problems than female teachers who struggled more with disrespectful students.

In another study, researchers found male teachers to rate talking out of turn and uncooperative behavior more frequently, but there were minimal differences between the overall frequency of reported behavioral problems between male and female teachers (Caldarella et al., 2009). When considering the specifics of behavioral ratings, some patterns emerge. Alter, Walker, and Landers (2013) found that female teachers perceived verbal disruptions and off-task behaviors as more problematic than male teachers, whereas male teachers were found to perceive social isolation as more problematic than female teachers. Similarly, in a Chinese sample, female teachers rated inattention and overactive behaviors as being more serious (Caldarella et al., 2009). Caldarella et al. (2009) found male and female teachers rated students of the opposite genders as having more serious behaviors.

Another factor influencing behavioral observations may be sensitivity differences between male and female teachers. Klassen and Chiu (2010) found that female teachers perceived themselves as having greater stress from their workload, greater stress from classroom behaviors, and lower classroom management self-efficacy. Stress in these forms tends to foster hypersensitivity to all stimuli and over-responsiveness (Klassen & Chiu, 2010).

In a New Zealand population, Sargisson et al. (2016) found that teachers generally rated male students higher than female students among total difficulties, externalizing behaviors, and internalizing behaviors when using the SDQ. When Sargisson et al. broke down teachers' ratings of children, female students were rated lower on the scales of emotional problems, hyperactivity, conduct problems, and peer problems. Glock and

Kleen (2017) found that preservice teachers associated female students with positive behavior and had strong associations of negative behavior with male students. Preschool teachers for children ages 2 to 4 rated males as having significantly more problems than females except for ratings on the emotional symptoms subscale, and males had significantly lower scores on the prosocial behaviors scale than females (Gustafsson, Proczkowska-Björklund, & Gustafsson, 2017). Male students are more likely to be perceived as aggressive compared to female students (Alter et al., 2013; Beaman, Wheldall, & Kemp, 2006; Driessen, 2015). Additionally, Gibson and Gore (2015) identified that physical attractiveness influenced female perceptions of behaviors; less attractive males were perceived as having more negative behaviors or violating social norms than attractive males.

Overall, research has shown results indicating female students are rated as having fewer behavioral problems than male students, but females are rated higher in internalizing behaviors while males are rated higher in externalizing behaviors (Caldarella et al., 2009; Glock, 2016; Glock & Kleen, 2017; Sargisson et al., 2016; Zwirs et al., 2011). Researchers found that teachers rated females with fewer behavioral problems in comparison to the students' ratings of their parents (Munzer et al., 2018). However, in those females identified as having a conduct disorder, the symptoms are often diagnosed before age 10 (Keenan, Wroblewski, Hipwell, Loeber, & Stouthamer-Loeber, 2010).

Researchers have studied the consequences of teachers' perceptions because externalizing behaviors are punished while internalizing behaviors are given support

(Glock & Kleen, 2017). Furthermore, Noltemeyer, Kunesh, Hostutler, Frato, and Sarr-Kerman (2012) and Glock and Kleen (2017) associated males' externalizing behaviors with problems in the home. According to Lim and Meer (2017), male students do not academically benefit from having the same sex teacher, but females' academic performances increased by 8% of a standard deviation when taught by a female teacher.

Summary

Researchers have confirmed race and gender play a role in a person's perception of behavior (Downer et al., 2016; Pellegrini, 2011; Sansone, 2017). Most researchers have focused on the race and gender of the child being rated when using behavioral assessments and screening tools; despite the lack of research on the rater, the current literature has shown gender and racial differences between teachers and students can play an important role when rating behaviors (Zwirs et al., 2011). Additionally, the halo/horns effect indicates support for the claim that race and gender can influence the perceptions of behaviors because a negative or positive concept of someone else will influence how someone's behaviors are perceived and rated (Hoyt, 2000; Lasky, 2015). The following chapter provides a description of the research method used for this study.

Chapter 3: Research Method

Introduction

Race and gender may compromise the validity of assessment scores (Jaeger & Freijo, 1975; Mason et al., 2014). Additionally, there was a lack of research on rater bias when using SDQ-TF in the United States (Downs et al., 2012). Therefore, the purpose of this quantitative study was to investigate the effect of race and gender congruency on teacher ratings of problematic student behaviors, as measured by the total difficulties score using the SDQ-TF 4-10. I used a nonprobability convenience sample of 98 teachers teaching children age 4 to 10 in the greater Phoenix, Arizona area to investigate if and to what degree teacher and student racial and gender congruency impacted the total difficulties score using the SDQ-TF. The findings may assist the field of psychology by increasing the validity and accuracy of evaluations for children who might require interventions for behavioral issues, and recommendations for these interventions might be taken into stronger consideration.

This chapter includes the purpose and parameters of this research, procedures for data collection, a description of the SDQ-TF with an explanation regarding its reliability and validity, and a rationale for the use of two-way factorial ANOVA used to analyze the data collected. Finally, threats to validity, ethical considerations, and limitations related to this research are included.

Research Design and Rationale

This study involved a quantitative design. The SDQ total difficulties score was the dependent variable in this study. The first independent variable was identified as

teacher race, which had two levels: congruent with or incongruent with the rated student. The second independent variable, teacher gender, also had two levels: congruent with or incongruent with the rated student. A two-way ANOVA fit analysis of the quantitative data, as it was appropriate for comparing mean differences between groups and for determining the presence of interaction effects between two independent variables on a dependent variable (see Laerd Statistics, 2013). Using G*Power, the minimum sample size was 55 based on the parameters of an effect size of 0.5, alpha 0.05, and a 0.95 power factor. The medium effect size was supported in previous literature (see Tenenbaum & Ruck, 2007), and the alpha level was standard for most statistical analyses.

The collected data were analyzed using a two-way ANOVA, which included three *F* tests: (a) the main effect for race congruency; (b) the main effect for gender congruency; and (c) if the main effect is observed, an *F* test for the interaction effect was conducted. I used the Levene's test to ensure that my sample met the assumption of homogeneity, and skewness and kurtosis within each distribution were assessed for normality to ensure the assumptions for a two-way ANOVA were met. None of the assumptions were violated. The findings are reviewed in Chapter 4.

Population and Procedures

The target population for this research was teachers who worked with students ages 4 to 10 within the greater Phoenix, Arizona area. I recruited teachers through social media forums specific to teachers (IRB approval 05-03-19-0118927). Teachers outside of Arizona and teaching children age 11 or older were excluded from this research. I used

this subgroup of school-aged children to avoid potential variability and skewed results due to developmental differences present in wider student grades and age ranges.

I recruited 98 teachers in the greater Phoenix, Arizona area through social media and word-of-mouth. I made online posts in social media forums specific to teachers in Arizona. I provided a brief synopsis of myself, including my affiliated school, my contact information, and details regarding the participation criteria. The informed consent form and demographic form (Appendices B to E), as well as the SDQ-TF for 4- to 10-year-olds, were provided to participants through encrypted e-mail, and the completed forms were returned through e-mail. The SDQ forms are not provided in this document per the authorizations from Youthinmind, as the SDQ cannot be published or widely disseminated. Permission was granted to use only the form provided on http://www.sdqinfo.com, and no alterations were allowed.

Participants first contacted me with interest. In turn, I responded to each participant with an encrypted e-mail that explained the criteria for participation and outlined the directions to complete each of the documents. These documents were password-protected to prevent any alterations. Thus, a second e-mail was sent to the participant with the password for the documents. Participants were informed their information remained confidential, and their consent form would be separated immediately from the data to prevent identification between the consent form signature and data. Additionally, participants were informed that their e-mail addresses were saved and would only be used to send a summary of the results after the completion of the

study. E-mails were permanently deleted to ensure the confidentiality of the participants and data collection.

Participants signed the consent form and then completed one demographic form indicating if they were the same or different race as the student, the same or different gender of the student rated, the grade the participant taught, and how many years teaching the participant. The SDQ-TF 4-10 was also completed about one student whom they perceived to have problematic behaviors, and they had worked with for at least 6 months.

Regarding confidentiality, I ensured each participant was aware they were not to provide any identifying information about themselves or the student they rated. There was no identifying information about the student being rated or teacher analyzed in the context of this research. I downloaded the documents returned to me and placed those on a password-protected computer as well as backed those up on a password-protected hard drive to maintain the anonymity of participants and the data provided. The demographic form and SDQ-TF 4-10 were saved in a separate file on a password-protected computer and backed up on a password-protected hard drive.

Once the documents were received and downloaded, I assigned a number to each datum. Then, each SDQ-TF 4-10 was printed for the purpose of scoring. The scoring sheets provided by http://www.sdqinfo.com were printed onto transparency paper to use as a layover to score each subscale. The scores from the four subscales were added to provide the total difficulties score. The data were input into an Excel spreadsheet, and Excel formulas were used to ensure the total difficulties score was added correctly. After

all data were entered, the Excel form was placed into the IBM SPSS software for analysis. The printed SDQ forms were shredded after scoring.

Instrumentation

Strengths and Difficulties Questionnaire

Those who use the SDQ, developed by Goodman (1997) in Great Britain, screen the positive and negative behaviors of children ages 4 through 17 (Lane, Robertson Kalberg, Parks, & Carter, 2008). The SDQ consists of three forms: the parent form, teacher form, and self-form. The parent and teacher forms are the same forms, only identified as either parent or teacher for each age group. The age groups for each form are ages 2 to 4, 4 to 10, and 11 to 17. For this research, the teacher form for ages 4 to 10 was used. The SDQ consists of 25 questions and is scored using a 3-point Likert scale (0 = *Not True*, 1 = *Somewhat True*, and 2 = *Certainly True*). The 25 questions are divided into five subscales: conduct problems, emotional symptoms, hyperactivity, peer problems, and prosocial behaviors. The original three-based categorization scoring method was used in this research. The total difficulties score was the sum of all five subscales. Total difficulty scores ranging from 0 to 11 were scored as *normal*, 12 to 15 were scored as *borderline*, and scores 16 to 40 were scored as *abnormal*.

Validity. Based on a meta-analysis of 48 studies, the SDQ has been defined as valid and reliable (Stone et al., 2010). Further, regarding construct validity, 16 sets of researchers have concluded that satisfactory factor loadings were >0.40 - ≤0.70 (as cited in Stone et al., 2017). The highest loading on the teacher form was the prosocial subscale. According to Goodman et al. (2000, 2004), teachers' ratings were sensitive to psychiatric

disorders 82.2% of the time. In a meta-analysis, emotional symptoms were correctly identified in 36% of the cases, and 93% of children with conduct problems were correctly identified (Stone et al., 2010).

The DAWAB was developed for *ICD-10* and *DSM-IV* diagnoses of mental health disorders in children ages 5 to 17 (Goodman et al., 2000). The children identified as having a psychiatric disorder by the DAWAB were correctly classified as having a disorder in 77.3% of the cases using the SDQ algorithm (Goodman et al., 2000). Furthermore, when the DAWAB indicated a child had hyperactivity, 91% were rated as probable for a hyperactive disorder using the SDQ; of those identified as having a conduct-oppositional disorder by the DAWAB, 44% were rated as probable using the SDQ (Hysing et al., 2007).

In comparison to other behavioral screening instruments such as the CBCL, Becker, Woerner, et al. (2004) indicated the German CBCL, CBCL-TRF, and the SDQ teacher form could differentiate between children with and without clinically significant symptoms for clinical diagnosis; thereby supporting the validity. The SDQ total difficulties scores among teachers, parents, and the self-report form positively correlate with diagnostic criteria for a mental health disorder (Goodman, 1997; Goodman & Goodman, 2011; Goodman et al., 2003), with similar results found in Australia (Mathai et al., 2004). Furthermore, in a meta-analysis, the weighted correlation for teacher ratings between the SDQ and CBCL total was 0.76 (Stone et al., 2010).

When compared to the Rutter questionnaires (Rutter et al., 1974), Goodman (1997) found the total scores of the SDQ to correlate, indicating concurrent validity

highly. Goodman (1997) found a high correlation (r = 0.92) between the teacher form of the SDQ total difficulties score and the Rutter total deviance score. Gowers et al. (1999) found a moderate correlation (r = 0.46) between the SDQ teacher form and the Health of the Nation Outcome Scales for Children and Adolescents.

Reliability. Regarding the SDQ in the United States, Brown et al. (2006) collected normative data using the SDQ parent form for 9,878 children. Results indicated the total difficulties score was internally consistent (Cronbach's $\alpha = 0.83$), and the impairment scales were consistent (Cronbach's $\alpha = 0.80$; Brown et al., 2006). The subscales of conduct problems, emotional symptoms, hyperactivity, and prosocial behaviors also showed good internal consistency (Cronbach's $\alpha = 0.63$ to 0.77); however, the peer problems subscale presented with poor consistency (Cronbach's $\alpha = 0.46$). Goodman et al. (1998) indicated the internal consistency, Cronbach's α, for subscales and total difficulties scores as follows: 0.82 for total difficulties, 0.75 for emotional symptoms, 0.72 for conduct problems, 0.69 for hyperactivity, 0.65 for prosocial behaviors, and 0.61 for peer problems. However, Palmieri and Smith (2007) described a couple of the subscales as having uncertain levels of internal consistency. Downs et al. (2012) cited multiple studies that researched the psychometrics of the SDQ in Europe, Asia, the United States, and Australia, and the researchers concluded that the overall performance of the SDQ measured well when used by parents and teachers for children ages 4 to 16. Lastly, the publishers of the SDQ specify that I could only use a hard-copy of the SDQ-TF and its scoring sheets. The hard copy of the SDQ is available online for free at http://www.sdqinfo.com/.

Instructions and Demographic Questionnaire

I designed the instructions form and demographic questions (Appendices D to G) to provide directions to participants on how they could determine or select the specific student they would rate and to gather information about the sample. For racial and gender congruency, the form read as the following: (a) Complete this form for a student who is the same race and same gender as yourself. For racial congruency and gender dissimilarity, the form read as the following: (b) Complete this form for a student who is the same race and different gender as yourself. For racial dissimilarity and gender similarity, the form read as the following: (c) Complete this form for a student who is of different race and the same gender as yourself. For racial and gender incongruity, the form read as the following: (d) Complete this form for a student who is of different race and different gender than yourself. The purpose of specifying the student to rate allows for equal distribution of the data my research method requires. However, because variables were presented to the teacher, the teacher might have responded differently.

In addition to the instructions, I used the form to ask for basic demographic information: questions about the teacher's gender, race, the grade they currently teach, and the number of years teaching. There was no identifying information about the student analyzed in the context of this research. The goal for this researcher-designed demographic questionnaire was to gather participant demographics for descriptive purposes and to facilitate creating the comparison groups for statistical analysis.

Threats to Validity

A potential threat to the internal or research validity of this research was the experience of teachers as this factor fell outside the scope of this research. This research was a single event with each participant. Because teachers could choose not to participate, they did so by not sending the forms back to me for collection, and they occasionally emailed those forms to inform me of their refusal. The constructs of the SDQ were outside the scope of this research. The instrumentation for this research had been well established, and there was no indication the instrumentation is expected to change, as well as it could not be altered.

Because teachers were aware that the participants in a study, their responses on the SDQ teacher form could have been modified; for example, a teacher could respond more positively or negatively about a student they rated because of several confounding factors unbeknownst to this researcher. Furthermore, because I focused on the incongruency/congruency of the teacher and student, this focus also alters the teachers' responses. For example, a teacher having knowledge that the student they rate is either of the same or different race may cause the teacher to respond by rating the student as having more or less behavioral issues when using the SDQ.

The most significant threat to external validity was the ability to generalize these results to larger or different populations. I focused on teachers who taught in the greater Phoenix, Arizona area; thus, these research findings might not be generalized to other populations. The time at which data were collected might have impacted the results. For

example, the events of the day or week might have skewed the teachers' perceptions of the child's behaviors they rated at the time of data collection.

Although teachers were instructed on how to choose a student to rate, confounding factors might impact the data. Teachers might be overly critical or less critical of a student's behaviors based on their personal perspectives and experiences. Without my direct observation or the inclusion of parent reports and self-reports, there might be an inability to differentiate data that showed problematic behaviors versus unknown, unstated biases that had come into play; however, this limitation was accounted for as best as possible by providing specific instructions to the participants.

Ethical Procedures

I provided copies of Walden University's Institutional Review Board approval to any participant who asked; however, the institutional review board approval number was listed in the informed consent. Each teacher who agreed to participate received the informed consent form and was asked to sign. This study was confidential; therefore, no names or information were used outside the perimeters of this study. Likewise, teachers who chose to participate were instructed not to provide any identifying information about the student they rate. No school names or student names were recorded.

Participants were informed right away during the informed consent process that they could withdraw from the study at any time, for any reason, without consequence. The information gathered in this research was only used for the purpose of this current research. Identifying information about the school, teacher, or student was not collected at any time. All electronic data, informed consents, and email addresses will be kept for a

minimum of 5 years on a password-protected computer and backed up on a password-protected hard drive only accessible by me. No information gathered in this study was shared with anyone, and the information was not intended to diagnose or inform treatment.

Limitations of this Study

Teachers received instructions about the SDQ in relation to selecting the student they rated for this research. Therefore, I assumed teachers followed the correct instructions as it related to identifying a student with problematic behaviors and of the same or different race and gender as themselves. I did not verify the teachers' racial or gender congruity or incongruity. I did not observe the students' behaviors; instead, the teachers completed the SDQ based on their perceptions of the behavior. The results of this study were not intended to be generalized beyond the academic culture of Arizona schools because Arizona school leaders might deal with different cultures than do schools in other parts of the United States. I targeted only elementary school teachers in the greater Phoenix, Arizona area. This study included a nonprobability convenience sample, which was not representative of an entire population. The time of year at which data were collected might have impacted the results (see Gilliam et al., 2016).

Summary

In summary, I investigated the effect of race (congruent or incongruent) and gender (congruent or incongruent) on teacher ratings of problematic student behaviors, as measured by the total difficulties score using the SDQ. I used a two-way ANOVA to test race and gender congruency to SDQ total difficulties scores and determine if an

interaction effect exists. I added to the existing research regarding rater biases and behavioral ratings, as these biases could skew the treatment of children and the services.

Chapter 4: Results

Introduction

The purpose of this quantitative study was to investigate the effect of race and gender congruency on teacher ratings of problematic student behaviors, as measured by the total difficulties score using the SDQ-TF 4-10 for children ages 4 to 10. I used a 2X2 factorial ANOVA with teacher race and gender congruency as independent variables and the total difficulties score as the dependent variable. The first independent variable, teacher race, had two levels: different (incongruent) or same (congruent) with the rated student. The second independent variable, gender, had two levels: incongruent or congruent with the rated student. The three research questions with their respective hypotheses were the following:

RQ1: Does teachers' race congruency have a significant effect on rated students' SDQ total difficulties scores?

$H_0$1: Teachers' race congruency has no significant effect on rated students' SDQ total difficulties scores.

$H_1$1: Teachers' race congruency has a significant effect on rated students' SDQ total difficulties scores.

RQ2: Does teachers' gender congruency have a significant effect on rated students' SDQ total difficulties scores?

$H_0$2: Teachers' gender congruency has no significant effect on rated students' SDQ total difficulties scores.

$H_1$2: Teachers' gender congruency has a significant effect on rated students' SDQ total difficulties scores.

RQ3: Is there a significant interaction between teachers' race congruency and teachers' gender congruency, affecting rated students' SDQ total difficulties scores?

$H_0$3: There is no significant interaction between teachers' race congruency and teachers' gender congruency affecting rated students' SDQ total difficulties scores.

$H_1$3: There is a significant interaction between teachers' race congruency and teachers' gender congruency affecting rated students' SDQ total difficulties scores.

In this chapter, I describe the data collection process and descriptive statistics for the data. The results are organized by the research question and hypotheses. Tables are used to present the finding of the analysis.

Data Collection

Data collection for this study began in May 2019 and ended in October 2019. The proposed process of data collection did not change. Data were collected by making online posts in the social media forum, Facebook. I searched for teacher groups within Arizona, and I requested access to each of the groups by sending a letter of my intentions to the respective managing facilitators of the groups. I identified five groups as potential forums to reach participants. Once the managing facilitator granted me access to the group, I made online posts searching for teachers of children ages 4 to 10 to participate in the study. The post contained a shortened version of the consent form, which included my

role and contact information, my university affiliation, and the parameters for participation. Potential participants were instructed to e-mail, message, or call me with questions or with an expression of interest to participate.

All participants received encrypted e-mails with the SDQ-TF, demographic form, and consent form. Each participant e-mailed their completed forms to me. After the data were received, consent forms were kept separate from the data to prevent any identification of information through signatures on the consent form. Each datum was hand-scored per the SDQ procedure and entered into an Excel worksheet, which was later transferred to IBM SPSS 25 for analysis.

The sample consisted of 79 females and 19 males for a total of 98 participants. Of these 98 participants, 65 people identified as Caucasian, eight people identified as African American, four people identified as other, and one person did not list a race. This one datum was utilized because the teacher indicated they were the same race as the child they rated; therefore, this had no bearing statistically because demographic information was not separately analyzed.

Descriptive Statistics

Approximately two-thirds of the sample identified as Caucasian (66.30%), with another 20% identifying as Hispanic. Only one participant endorsed multiple race categories. Many participants were female (80.60%), which is roughly consistent with the existing literature about teacher gender identification in Arizona (National Center for Education Statistics, 2012). The average number of years teaching was 11. The average grade of the student rated was second grade approaching third grade. Slightly over half

(52%) of the sample was race congruent between the teacher and student. Table 1 shows this information.

Table 1

Frequencies and Percentages of Teacher Categorical Demographic Variables

Category	Frequency (%)
Race	
Caucasian	66.33
African American	8.16
Hispanic	20.41
Native American	0.00
Other	4.08
Not reported*	1.02
Gender	
Female	80.61
Male	19.31
Congruency	
Race	52.04
Gender	42.86

Note. One individual did not provide a racial identity.

The subscales of emotional symptoms, conduct problems, hyperactivity, and peer problems are summed to provide the total difficulties score. among the subscales, the largest observed mean obtained on the hyperactivity scale (7.81). Conduct problems were the second-highest observed mean (5.52). The total difficulties score does not include the prosocial behavior subscale; the prosocial behavior mean and standard deviation are included in Table 2 for completeness.

Table 2

Means and Standard Deviations of Strengths and Difficulties Subscales and Total Difficulties Score

Subscale	Mean	Std. deviation
Emotional Symptoms	3.10	2.44
Conduct Problems	5.52	2.55
Hyperactivity	7.81	2.31
Peer Problems	3.39	2.20
Prosocial Behaviors	4.76	2.35
Total Difficulties Score	19.82	6.28

Note. Two missing cases observed, and the cases removed pairwise. Prosocial scores have no impact on the total difficulties score.

Of the total difficulties scores, approximately three-fourths of the students rated fell outside the normal range on the SDQ-TF. This finding indicated 75.50% of the students rated had a high probability of meeting the criteria for a behavioral disorder. Fifteen percent of the total difficulties scores fell within the borderline range, and only about 9% fell within the normal range.

Table 3

Frequencies and Percentages of Categorical Total Difficulties Score

SDQ-TF Total Difficulties Severity	Frequency (%)
Normal	9.18
Borderline	15.31
Abnormal	75.51

Table 4

Mean and Standard Deviations of Student Continuous Demographic and Study Variables

Category	Mean	Std. deviation
Student variable		
Grade	4.43*	1.95
Teacher Variable		
Years Teaching	11.11	8.39

Note. Preschool was scored as 0, kindergarten as 1, and 1st grade as 2 (and so forth).

Statistical Assumptions

The planned analysis was a two-way ANOVA as the outcome variable was continuous, both predictors were categorical (with two levels), and the observations were independent. I examined the *Z*-scores to test for outliers. The largest Z-score equated to -0.87. No significant outliers were identified, and this assumption had been met. The homogeneity assumption was met as Levene's test was not significant ($p = 0.88$). I used a Shapiro-Wilk test to test the normality assumption (for each group). None of the Shapiro-Wilk tests were significant; meaning, the dependent variable was approximately normal for each group, as presented in Table 5.

Table 5

Shapiro-Wilk Tests

Group	W (*df*)
Race and Gender Congruent	0.96 (21) ^{ns}
Race Congruent/Gender Incongruent	0.95 (30) ^{ns}
Race Incongruent/Race Congruent	0.97 (21) ^{ns}
Race and Gender Incongruent	0.94 (26) ^{ns}

Note. ^{ns} not significant.

Two-Way ANOVA Findings

I used a 2x2 factorial ANOVA (a type of general linear model), with the outcome variable being the total difficulty score to explore the hypotheses. I used two predictor variables: race congruency (with the two levels being congruent or non-congruent) and gender congruency (with the two levels also being congruent or noncongruent). I did not use posthoc analyses because each predictor had only two levels. No additional options were employed beyond the default settings. Significance decisions were evaluated using

$\alpha = 0.05$. I analyzed data using IBM SPSS Statistics 25 Software. Table 6 shows the analysis summary results.

Table 6

Means and Standard Deviations for SDQ-TF by Group

Congruency	Mean	Confidence intervals
Gender Congruency		
Congruent	18.36	16.47 – 20.24
Incongruent	20.94	19.30 – 22.57
Race Congruency		
Congruent	20.24	18.50 – 21.97
Incongruent	19.06	17.27 – 20.85
Interaction		
Race/Gender Congruent	19.91	17.24 – 22.57
Race-Congruent Gender-Incongruent	20.57	18.34 – 22.80
Race-Incongruent Gender-Congruent	16.81	14.15 – 19.47
Gender/Race Non-Congruent	21.31	18.91 – 23.70

Research Question 1

I used a two-way ANOVA to test whether teacher race congruency had a statistically significant effect on student total difficulties scores. The main effect for race congruency was not significant, $F(1,94) = 0.88$, $p = 0.35$, $\eta^2 = 0.01$; therefore, the null hypothesis was retained, and the alternate hypothesis set aside. Based on the data from participants in this study, the results showed that teacher race congruency, the same race as the rated student compared with not the same race, had no significant effect on rated student SDQ total difficulties scores.

Research Question 2

A two-way ANOVA was used to test whether teacher gender congruency had a statistically significant effect on student total difficulties scores. The main effect for gender congruency was significant, $F(1,94) = 4.21$, $p = 0.043$. $\eta^2 = 0.04$; therefore, the

null hypothesis was rejected, and the alternate hypothesis accepted. Based on the data from participants in this study, the results affirmed teacher gender congruency, same gender as the rated student compared with not the same gender, had a significant effect on rated student SDQ total difficulties scores. More specifically, gender congruency was associated with lower total difficulties scores (M = 18.36, 6.56) than gender noncongruency (M = 20.91, SD = 5.88). These results are discussed in detail below.

Research Question 3

I used a two-way ANOVA to test whether the interaction between teachers' race congruency and teachers' gender congruency affected rated students' SDQ total difficult score. The interaction was not significant, $F(1,94) = 2.33$, $p = 0.13$, $\eta^2 = 0.02$; therefore, the null hypothesis was retained, and the alternate hypothesis set aside. Based on the data from participants in this study, the results affirmed that the interaction between teacher race congruency and gender congruency had no significant effect on the student rated SDQ total difficulties scores. Table 7 presents a summary of ANOVA results.

Table 7

ANOVA Summary Table

Variable	Sum of squares	*df*	Mean square	*F*	η^2
Race Congruency	33.18	1	33.18	0.88^{ns}	0.01
Gender Congruency	159.41	1	159.41	4.21^{*}	0.04
Interaction	88.11	1	88.11	2.33^{ns}	0.02
Error	3555.95	94	37.83		
Total	42304.00	98			

Note. Mean square and sum of square variables are identical because $df = 1$ for all variables. * $p < 0.05$, ^{ns} $p > 0.05$.

Summary

Based on the two-way ANOVA results, I did not reject the null hypothesis for Research Question 1. The results showed no significant differences occurred between the teachers' race congruency and total difficulties scores. The *F* tests for Research Question 2 were significant; therefore, I rejected the null hypothesis in favor of the alternative hypotheses. This finding indicated teacher gender congruency impacted the total difficulties score. Regarding Research Question 3, I did not reject the null hypothesis, indicating that no statistically significant interactions occurred between the two main effects of teacher race and gender congruency on total difficulties scores. I discuss these findings in Chapter 5.

Chapter 5: Discussion, Conclusions, and Recommendations

Introduction

The purpose of this quantitative research was to investigate the effect of race and gender congruency on teacher ratings of problematic student behaviors, as measured by the total difficulties score using the SDQ-TF 4-10 for children ages 4 to 10. Researchers have studied the SDQ in many countries, and it has been normed within the United States (Brown et al., 2006; Ruchkin et al., 2012). However, in some countries, the school population is relatively homogeneous, but in the United States, it is more diverse. Therefore, teachers' observations of their students may more often be influenced by gender or race differences between the teacher and the student.

The findings of this study indicated that the total difficulties score was influenced by the gender congruency of the teacher and student, even though the expected sample size was not achieved. The race congruency variable was not found to have a significant effect. There was no interaction effect between race and gender congruency on total difficulties scores.

Interpretation of the Findings

This study was guided by the implicit bias theory. The implicit bias theory indicates that humans unwittingly uphold biases of others (Ungvarsky, 2017), which can influence how people view and treat others. For instance, researchers have found that teachers treat their students differently based on race and gender (Okonofua & Eberhardt, 2015; Westerberg, 2016). The results of the present study indicated that teacher gender congruency influences total difficulties scores. When the teacher and student are of the

same gender, the total difficulties scores were lower. This finding was consistent with previous research that showed the gender of the teacher and gender of the student influenced behavioral ratings of the student. For example, Rong (1996) found that female teachers would rate female students more positively than male students, and female preschool teachers rated female students as having lower externalizing behaviors than males (Berg-Nielsen et al., 2012). Sansone (2017) also found that students might perform better with a teacher of the same gender. Additionally, researchers have thought of male and female teachers as more sensitive to different types of behaviors; therefore, they each perceive different behaviors differently among their male and female students (Alter et al., 2013; Caldarella et al., 2009; Friedman, 1995). Female teachers have perceived themselves to have greater stress from multiple sources, such as stress from the behaviors in the classroom, which may cause the teacher to foster hypersensitivity to all stimuli and be over-responsive (Klassen & Chiu, 2010).

Regarding the race congruency or incongruency between teachers and students, more favorable assessments of students' behaviors have occurred when the teacher and student are of the same race (Downey & Pribesh, 2004; Gilliam et al., 2016; Ouazad, 2014). The race of the student has shown to affect perceptions and rating of student behavior—more specifically, African American students are more likely to be poorly rated by their teachers and to experience harsher discipline when their teacher is of a different race (Horner et al., 2010; Wright, 2015); they are often rated as less disruptive when the teacher is of the same race (Chang & Demayan, 2007; Downer et al., 2016; Downey & Pribesh, 2004; McGrady & Reynolds, 2013; Munzer et al., 2018). Overall,

research indicates that Caucasian teachers rate African American students' behaviors poorly versus Caucasian students' behaviors (Horner et al., 2010). However, my findings did not show support for racial congruency or incongruency among teachers and students impacted the total difficulties scores on the SDQ.

Though the results were not significant for race as in previous research (see Zwirs et al., 2011), the race of this sample was consistent with existing statistics. In Arizona, during the 2011 to 2012 school year, approximately 80% of teachers were Caucasian, and approximately 13% were Hispanic. This study included approximately 66% Caucasian teachers and 20% Hispanic teachers. Additionally, in Arizona, approximately 45% of students were Hispanic, with a close following of 39% of students being Caucasian (National Center for Education Statistics, 2016). Research has shown that Caucasian teachers' perceptions of Hispanic students generally did not differ from their perceptions of Caucasian students; however, Caucasian, African American, and Hispanic students have better behavioral ratings when assessed by a teacher of the same race than by a teacher of a different race (McGrady & Reynolds, 2013).

One possible explanation for the lack of significance on the race congruency variable is that in Arizona, the presence of Hispanic students is long-standing. There was a time when Arizona and adjoining (now the United States) states were part of Mexico. Although the international boundary between Mexico and the United States changed due to the Mexican American War (1846–1848), the mixture of Caucasian and Hispanic residents of Arizona did not consequently change. This finding contrasted with the situation in, for example, the Netherlands, where researchers have found race congruency

significance (Zwirs et al., 2011). In the Netherlands, the population is largely Dutch at 79.3% of the total population, with 2.4% comprised of Turkish, 2.2% Moroccans, and 2.1% Surinamese, as shown on the website (Netherlands Population 2019, 2019). The Turkish population in the Netherlands did not occur until as recent as 1960. This finding means that the established Dutch population and the Turkish population do not have an extended history of living together. Additionally, the Dutch population has a history of discrimination and uneasiness toward accepting the Turkish and Moroccan populations into Dutch culture, as well as political and religious tensions (Crul & Doomernik, 2003). Potentially, the longer that different groups face fewer tensions, the more negative perceptions of others may decrease.

Researchers have defined the SDQ as reliable and valid (Goodman, 1997, 2001; Stone et al., 2010). This study was consistent with past research; 75.51% of total difficulties scores were within the abnormal range; therefore, approximately three-fourths of this sample rated children as likely having some type of behavioral disorder. Researchers have found that the SDQ-TF predicted a psychiatric disorder in children, 82.2% of the time (Goodman et al., 2000, 2004). More researchers concurred with existing research and found the SDQ-TF was 90% sensitive to those with a disorder (Goodman et al., 2003). This finding indicates the SDQ can serve as a screener as the current findings are relatively similar to the previous research.

The current results indicate the subscale of hyperactivity is the highest-rated subscale, as consistent with more recent research (see van den Heuvel et al., 2017). Among the subscales, emotional symptoms had the lowest mean score of 3.10, and peer

problems had a mean score of 3.39. This finding was relatively consistent with previous research. Some researchers found emotional symptoms and peer problems had lower scores than all other subscales (d'Acremont & Van der Linden, 2008), while others found those scales as higher (Goodman et al., 1998).

Limitations

This study was limited to a convenience sample; thus, results might not represent the general population, which might be partly related to the method of data collection. I found gathering participants online difficult because of teacher discomfort. Teachers were often disinterested because of seeing their names listed in the emails and having to sign the consent form. Many teachers opted not to participate despite affirmation that I would not associate or use their personal information.

Another limitation of this study was that only the teacher was used as an informant. Researchers have advised multiple informants to support mental health diagnoses because it solidifies the behaviors are consistent in multiple environments and not just singled out to one environment (Goodman et al., 2000). However, I did not study multiple informants as I did not intend to inform treatment or compare parents' and teachers' perceptions.

The number of years of teaching might also influence behavioral ratings, as well as the time of year that I collected data. Moreover, the teacher's years of experience could have impacted the behavioral ratings. For example, preservice teachers associated female students with positive behaviors and negative behaviors with male students (Glock & Kleen, 2017).

I gathered data toward the end of the school year. Gilliam et al. (2016) posited the timeframe within the school year could influence behaviors and teacher burnout. Gilliam et al. found Caucasian teachers rated African American students as having increased problematic behaviors by the end of the school year as compared to the beginning of the school year. Researchers have argued that African American teachers understood African American students culturally; thus, African American teachers did not perceive African American students' behaviors as negatively increasing throughout the school year (Downer et al., 2016; Gilliam et al., 2016).

Implications

The present findings were interesting. Results showed that gender congruency was a contributing factor to total difficulties scores, but race had no influence; moreover, no interaction effect occurred. In this sample, race had no effect compared to previous research that showed race influenced total difficulties scores using the SDQ in other countries. These results are helpful in learning about how the SDQ may be influenced in certain cultures and warrant further research to expand on this finding. If future researchers can add to this study, they may verify how much gender influences SDQ outcomes and confirm or deny racial influences. As stated throughout this research, behavioral assessments are generally influenced by race and gender differences; thus, professionals diagnosing or treating students with potential behavioral disorders should know the most current research about how outcomes on behavioral screeners or assessments may have been influenced and by what factors. Therefore, multiple informants should be used to ensuring consistency among raters and environments.

Recommendations

Future researchers can find other means for collecting data from teachers by using multiple informants to demonstrate the consistency of behavioral ratings. Future researchers can use the SDQ-TF in a more personal setting, potentially with the consent of a school district and parents to further the research of rater congruency. A future researcher may alleviate teacher discomfort and timeliness of receiving data from participants. Other factors worth researching can include the time of year a researcher collects data and teachers' years of teaching experience.

Future researchers should include an in-depth look at diversity and how members of specific cultures or races view each other in combination with their history together, politics, and religious factors compared to behavioral assessment outcomes. Future researchers may produce an understanding of potential biases and student behavioral assessments.

Conclusions

Although these results may not be generalizable, some remarkable consistencies are like previous research. The findings show the SDQ is valid and reliable with some racial and gender influences depending on the cultures. Race and gender may influence behavioral ratings of children; therefore, future researchers should research race and gender within U.S. subcultures. Not only should researchers expand this study, but existing research has also shown differences between cocultures and subcultures.

References

Allday, R. A., Duhon, G. J., Blackburn-Ellis, S., & Van Dycke, J. L. (2011). The biasing effects of labels on direct observation by preservice teachers. *Teacher Education and Special Education, 34*(1), 52–58. doi:10.1177/0888406410380422

Alter, P., Walker, J. N., & Landers, E. (2013). Teachers' perceptions of students' challenging behavior and the impact of teacher demographics. *Education and Treatment of Children, 36*(4), 51–69. Retrieved from https://www.jstor.org/stable/42900226

Álvarez-García, D., García, T., & González-Castro, P. (2014). Halo effect of student misbehavior on teacher ratings of impulsivity and other executive function deficits. In M. C. Olmstead (Ed.), *Psychology of emotions, motivations and actions. Psychology of impulsivity: New research* (pp. 21–31). New York, NY: Nova Science.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Arizona Department of Education. (2018a). *Disability categories.* Retrieved from http://www.azed.gov/specialeducation/disability-categories/

Arizona Department of Education. (2018b). *Enrollment data 2018-2019 School Year.* [Data file]. Retrieved from https://www.azed.gov/accountability-research/data/

Avant, N. D., Weed, E., Connelly, C., Hincapie, A. L., & Penn, J. (2018). Qualitative analysis of student pharmacists' reflections of Harvard's Race Implicit Association Test. *Currents in Pharmacy Teaching & Learning, 10*(5), 611–617.

doi:10.1016/j.cptl.2018.02.002

Beaman, R., Wheldall, K., & Kemp, C. (2006). Differential teacher attention to boys and girls in the classroom. *Educational Review*, *58*(3), 339–366. doi:10.1080/00131910600748406

Becker, A., Hagenberg, N., Roessner, V., Woerner, W., & Rothenberger, A. (2004). Evaluation of the self-reported SDQ in a clinical setting: Do self-reports tell us more than ratings by adult informants? *European Child & Adolescent Psychiatry, 13*(Suppl2), II17–II24. doi:10.1007/s00787-004-2004-4

Becker, A., Woerner, W., Hasselhorn, M., Banaschewski, T., & Rothenberger, A. (2004). Validation of the parent and teacher SDQ in a clinical sample. *European Child & Adolescent Psychiatry, 13*(Suppl2), II11–II16. doi:10.1007/s00787-004-2003-5

Berg-Nielsen, T. S., Solheim, E., Belsky, J., & Wichstrom, L. (2012). Preschoolers' psychosocial problems: In the eyes of the beholder? Adding teacher characteristics as determinants of discrepant parent–teacher reports. *Child Psychiatry & Human Development, 43*(3), 393–413. doi:10.1007/s10578-011-0271-0

Björnsdotter, A., Enebrink, P., & Ghaderi, A. (2013). Psychometric properties of online administered parental strengths and difficulties questionnaire (SDQ), and normative data based on combined online and paper-and-pencil administration. *Child and Adolescent Psychiatry and Mental Health, 7*(1), Article 40. doi:10.1186/1753-2000-7-40

Blake, J. J., Smith, D. M., Marchbanks, M. P., Seibert, A. L., Wood, S. M., & Kim, E. S.

(2016). Does student–teacher racial/ethnic match impact Black students' discipline risk? A test of the cultural synchrony hypothesis. In *Inequality in school discipline* (pp. 79–98). doi:10.1037/e626432013-001

Bradshaw, C. P., Mitchell, M. M., O'Brennan, L. M., & Leaf, P. J. (2010). Multilevel exploration of factors contributing to the overrepresentation of black students in office disciplinary referrals. *Journal of Educational Psychology, 102*(2), 508–520. doi:10.1037/a0018450

Brown, J. D., Ciara, Z., Bartlett, E., & Horn, I. (2006). Parent and teacher mental health ratings of children using primary care services: Inter-rater agreement and implications for mental health screening. *Ambulatory Pediatrics, 6*(6), 347–351. doi:10.1016/jambp.2006.09.004

Caldarella, P., Shatzer, R. H., Richardson, M. J., Shen, J., Zhang, N., & Zhang, C. (2009). The impact of gender on Chinese elementary school teachers' perceptions of student behavior problems. *New Horizons in Education, 57*(2), 17–31. doi:10.1080/01443410802654909

Capron, C., Thérond, C., & Duyme, M. (2007). Psychometric properties of the French version of the self-report and teacher Strengths and Difficulties Questionnaire (SDQ). *European Journal of Psychological Assessment, 23*(2), 79–88. doi:10.1027/1015–5759.23.2.79

Center for Parent Information & Resources. (2017). *Emotional disturbance*. Retrieved from https://www.parentcenterhub.org/emotionaldisturbance/#conduct

Chang, D. F., & Demyan, A. L. (2007). Teachers' stereotypes of Asian, Black, and White

students. *School Psychology Quarterly, 22*(2), 91–114. doi:10.1037/1045–3830.22.2.91

Chang, D. F., & Sue, S. (2003). The effects of race and problem type on teachers' assessments of student behavior. *Journal of Consulting & Clinical Psychology, 71*(2), 235–242. doi:10.1037/0022-006x.71.2.235

Cheng, S., Keyes, K. M., Bitfoi, A., Carta, M. G., Koç, C., Goelitz, D., . . . Kovess-Masfety, V. (2018). Understanding parent–teacher agreement of the Strengths and Difficulties Questionnaire (SDQ): Comparison across seven European countries. *International Journal of Methods in Psychiatric Research, 27*(1), e1589. doi:10.1002/mpr.1589

Crul, M., & Doomernik, J. (2003). The Turkish and Moroccan second generation in the Netherlands: Divergent trends between and polarization within the two groups. *International Migration Review, 37*(4), 1039–1064. doi:10.1111/j.1747-7379.2003.tb00169.x

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 12*(2), 163–170. doi:10.1111/1467-9280.00328

Cushman, P. (2010). Male primary school teachers: Helping or hindering a move to gender equity? *Teaching and Teacher Education, 26*(5), 1211–1218. doi:10.1016/j.tate.2010.01.002

d'Acremont, M., & Van der Linden, M. (2008). Confirmatory factor analysis of the Strengths and Difficulties Questionnaire in a community sample of French-

speaking adolescents. *European Journal of Psychological Assessment, 24*(1), 1–8. doi:10.1027/1015-5759.24.1.1

De Los Reyes, A. (2011). Introduction to the special section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child & Adolescent Psychology, 40*(1), 1–9. doi:10.1080/15374416.2011.533405

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin, 141*(4), 858–900. doi:10.1037/a0038498

Dirks, M. A., De Los Reyes, A., Briggs-Gowan, M., Cella, D., & Wakschlag, L. S. (2012). Annual research review: Embracing not erasing contextual variability in children's behavior–theory and utility in the selection and use of methods and informants in developmental psychopathology. *Journal of Child Psychology and Psychiatry, 53*(5), 558–574. doi:10.1111/j.1469-7610.2012.02537.x

Downer, J. T., Goble, P., Myers, S. S., & Pianta, R. C. (2016). Teacher-child racial/ethnic match within pre-kindergarten classrooms and children's early school adjustment. *Early Childhood Research Quarterly, 37*, 26–38. doi:10.1016/j.ecresq.2016.02.007

Downey, D. B., & Pribesh, S. (2004). When race matters: Teachers' evaluations of students' classroom behavior. *Sociology of Education, 77*(4), 267–282. doi:10.1177/003804070407700401

Downs, A., Strand, P. S., Heinrichs, N., & Cerna, S. (2012). Use of the teacher version of the Strengths and Difficulties Questionnaire with German and American preschoolers. *Early Education and Development, 23*(4), 493–516. doi:10.1080/10409289.2010.532082

Driessen, G. (2015). Teacher ethnicity, student ethnicity, and student outcomes. *Intercultural Education*, *26*(3), 179–191. doi:10.1080/14675986.2015.1048049

Du, Y., Kou, J., & Coghill, D. (2008). The validity, reliability and normative scores of the parent, teacher and self report versions of the Strengths and Difficulties Questionnaire in China. *Child and Adolescent Psychiatry and Mental Health*, *2*(1), Article 8. doi:10.1186/1753-2000-2-8

Egalite, A. J., Kisida, B., & Winters, M. A. (2015). Representation in the classroom: The effect of own-race teachers on student achievement. *Economics of Education Review*, *45*, 44–52. doi:10.1016/j.econedurev.2015.01.007

Epstein, J. N., March, J. S., & Conners, C. K. (1998). Racial differences on Conners Teacher Rating Scale. *Journal of Abnormal Child Psychology*, *26*(2), 109–118. doi:10.1023/a:1022617821422

Fält, E., Wallby, T., Sarkadi, A., Salari, R., & Fabian, H. (2018). Agreement between mothers', fathers', and teachers' ratings of behavioural and emotional problems in 3–5-year-old children. *PloS One*, *13*(11), e0206752. doi:10.1371/journal.pone.0206752

Fenning, P., & Rose, J. (2007). Overrepresentation of African American student in exclusionary discipline – the role of school policy. *Urban Education, 42*(6), 536–

559. doi:10.1177/0042085907305039

Fergusson, D. M., Boden, J. M., & Horwood, L. J. (2009). Situational and generalised conduct problems and later life outcomes: Evidence from a New Zealand birth cohort. *Journal of Child Psychology and Psychiatry, 50*(9), 1084–1092. doi:10.1111/j.1469-7610.2009.02070.x

Foster, G., & Ysseldyke, J. (1976). Expectancy and halo effects as a result of artificially induced teacher bias. *Contemporary Educational Psychology, 1*(1), 37–45. doi:10.1016/0361-476x(76)90005-9

Friedman, I. A. (1995). Student behavior patterns contributing to teacher burnout. *The Journal of Educational Research, 88*(5), 281–289. doi:10.1080/00220671.1995.9941312

Gender. (n.d.). In *Merriam-Webster's collegiate dictionary* (10th ed.). Springfield, MA: Merriam-Webster.

Gender identity. (n.d.). In *Merriam-Webster's collegiate dictionary* (10th ed.). Springfield, MA: Merriam-Webster.

Gibson, J. L., & Gore, J. S. (2015). You're OK until you misbehave: How norm violations magnify the attractiveness devil effect. *Gender Issues, 32*(4), 266–278. doi:10.1007/s12147-015-9142-5

Gilliam, W. S., Maupin, A. N., Reyes, C. R., Accavitti, M., & Shic, F. (2016). A research study brief: Do early educators' implicit biases regarding sex and race relate to behavior expectations and recommendations of preschool expulsions and suspensions? *Yale Child Study Center*. Retrieved from

https://medicine.yale.edu/childstudy/zigler/publications/Preschool%20Implicit%20Bias%20Policy%20Brief_final_9_26_276766_5379_v1.pdf

Glaser, B. A., Kronsnoble, K. M., & Forkner, C. W. (1997). Parents and teachers as raters of children's problem behaviors. *Child & Family Behavior Therapy, 19*(4), 1–13. doi:10.1300/J019v19n04_01

Glock, S. (2016). Stop talking out of turn: The influence of student's gender and ethnicity on preservice teachers' intervention strategies for student misbehavior. *Teaching and Teacher Education, 56,* 106–114. doi: 10.1016/j.tate.2016.02.012

Glock, S., & Kleen, H. (2017). Gender and student misbehavior: Evidence from implicit and explicit measures. *Teaching and Teacher Education, 67,* 93–103. doi:10.1016/j.tate.2017.05.015

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38,* 581–586. doi:10.1111/j.1469-7610.1997.tb01545.x

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(11), 1337–1345. doi:10.1097/00004583-200111000-00015

Goodman, R., & Goodman, A. (2011). Population mean scores predict child mental disorder rates: Validating SDQ prevalence estimators in Britain. *Journal of Child Psychology and Psychiatry, 52*(1), 100–108. doi:10.1111/j.1469-7610.2010.02278.x

Goodman, R., Ford, T., Corbin, T., & Meltzer, H. (2004). Using the Strengths and

Difficulties Questionnaire (SDQ) multi-informant algorithm to screen looked-after children for psychiatric disorders. *European Child & Adolescent Psychiatry, 13*(Supplement 2), 25–31. doi:10.1007/s00787-004-2005-3

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2003). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *International Review of Psychiatry*, *15*(1–2), 166–172. doi:10.1080/0954026021000046128

Goodman, R., Meltzer, H., & Bailey, V. (1998). The strengths and difficulties questionnaire: A pilot study on the validity of the self-report version. *European Chand and Adolescent Psychiatry, 7*, 125–130. doi:10.1007/s007870050057

Goodman, R., Renfrew, D., & Mullick, M. (2000). Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Child & Adolescent Psychiatry, 9*(2), 129–134. doi:10.1007/s007870050008

Goodman, R., & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? *Journal of Abnormal Child Psychology, 27*(1), 17–24. doi:10.1023/a:1022658222914

Gowers, S. G., Harrington, R. C., Whitton, A., Lelliott, P., Beevor, A., Wing, J., & Jezzard, R. (1999). Brief scale for measuring the outcomes of emotional and behavioural disorders in children. Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA). *British Journal of Psychiatry, 174*(5), 413–416. doi:10.1192/bjp.174.5.413

Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review, 94*(4), 945–967. doi:10.2307/20439056

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,* 1464–1480. doi:10.1037/0022-3514.74.6.7464

Gustafsson, B. M., Proczkowska-Björklund, M., & Gustafsson, P. A. (2017). Emotional and behavioural problems in Swedish preschool children rated by preschool teachers with the Strengths and Difficulties Questionnaire (SDQ). *BioMed Central Pediatrics, 17*(1), 110. doi:10.1186/s12887-017-0864-2

Halberstadt, A. G., Castro, V. L., Chu, Q., Lozada, F. T., & Sims, C. M. (2018). Preservice teachers' racialized emotion recognition, anger bias, and hostility attributions. *Contemporary Educational Psychology, 54,* 125–138. doi:10.1016/j.cedpsych.2018.06.004

Hawes, D. J., & Dadds, M. R. (2004). Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. *Australian and New Zealand Journal of Psychiatry, 38*(8), 644–651. doi:10.1111/j.1440-1614.2004.01427x

He, J. P., Burstein, M., Schmitz, A., & Merikangas, K. R. (2013). The Strengths and Difficulties Questionnaire (SDQ): The factor structure and scale validation in US adolescents. *Journal of Abnormal Child Psychology, 41*(4), 583–595. doi:10.1007/s10802-012-9696-6

Hodgins, S., Larm, P., Ellenbogen, M., Vitaro, F., & Tremblay, R. E. (2013). Teachers'

ratings of childhood behaviours predict adolescent and adult crime among 3016 males and females. *The Canadian Journal of Psychiatry, 58*(3), 143–150. doi:10.1177/070674371305800304

Horner, S. B., Fireman, G. D., & Wang, E. W. (2010). The relation of student behavior, peer status, race, and gender to decisions about school discipline using CHAID decision trees and regression modeling. *Journal of School Psychology, 48*(2), 135–161. doi:10.1016/j.jsp.2009.12.001

Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods, 5*(1), 64–86. doi:10.1037/1082-989X.5.1.64

Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology*, *3*, 29–51. doi:10.1146/annurev.clinpsy.3.022806.091419

Hysing, M., Elgen, I., Gillberg, C., Lie, S., & Lundervold, A. J. (2007). Chronic physical illness and mental health in children. Results from a large-scale population study. *Journal of Child Psychology and Psychiatry*, *48*(8), 785–792. doi:10.1111/j.1469-7610.2007.01755.x

Jaeger, R. M., & Freijo, T. D. (1975). Race and sex as concomitants of composite halo in teachers' evaluative rating of pupils. *Journal of Educational Psychology, 67*(2), 226–237. doi:10.1037/h0076996

Jolls, C., & Sunstein, C. R. (2006). The law of implicit bias. *California Law Review, 94*, 969. doi:10.2307/20439057

Kappes, A., Harvey, A. H., Lohrenz, T., Montague, P. R., & Sharot, T. (2020).

Confirmation bias in the utilization of others' opinion strength. *Nature Neuroscience*, *23*(1), 130–137. doi:10.1038/s41593-019-0549-2

Katz, I. (2017). In the eye of the beholder: Motivational effects of gender differences in perceptions of teachers. *Journal of Experimental Education*, *85*(1), 73–86. doi:10.1080/00220973.2015.1101533

Keenan, K., Wroblewski, K., Hipwell, A., Loeber, R., & Stouthamer-Loeber, M. (2010). Age of onset, symptom threshold, and expansion of the nosology of conduct disorder for girls. *Journal of Abnormal Psychology, 119*(4), 689–698. doi:10.1037/a0019346

Kempf, A. (2020). If we are going to talk about implicit race bias, we need to talk about structural racism: Moving beyond ubiquity and inevitability in teaching and learning about race. *Taboo: The Journal of Culture & Education*, *19*(2), 115–132. Retrieved from Complementary Index. (Accession No. 141855658)

Klassen, R. M., & Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of Educational Psychology*, *102*(3), 741–756. doi:10.1037/a0019237

Kozlowski, K. P. (2015). Culture or teacher bias? Racial and ethnic variation in student-teacher effort assessment match/mismatch. *Race and Social Problems, 7*(1), 43–59. doi:10.1007/s12552-014-9138-x

Kulinna, P. H. (2008). Teachers' attributions and strategies for student misbehavior. *Journal of Classroom Interaction, 42*, 21–30. Retrieved from https://files.eric.ed.gov/fulltext/EJ829002.pdf

Laerd Statistics. (2013). *Two-way ANOVA in SPSS statistics*. Retrieved from https://statistics.laerd.com/spss-tutorials/two-way-anova-using-spss-statistics.php.

Lane, K. L., Robertson Kalberg, J., Parks, R. J., & Carter, E. W. (2008). Student Risk Screening Scale: Initial evidence for score reliability and validity at the high school level. *Journal of Emotional and Behavioral Disorders, 16*(3), 178–190. doi:10.1177/1063426608314218

Lasky, J. (2015). Halo effect. In *Salem press encyclopedia of health*. Retrieved from https://www.salempress.com/

Lim, J., & Meer, J. (2017). The impact of teacher–student gender matches random assignment evidence from South Korea. *Journal of Human Resources, 52*(4), 979–997. doi:10.3368/jhr.52.4.1215-7585R1

Liu, S. K., Chien, Y. L., Shang, C. Y., Lin, C. H., Liu, Y. C., & Gau, S. S. F. (2013). Psychometric properties of the Chinese version of Strength and Difficulties Questionnaire. *Comprehensive Psychiatry, 54*(6), 720–730. doi:10.1016/j.comppsych.2013.01.002

Mason, B. A., Gunersel, A. B., & Ney, E. A. (2014). Cultural and ethnic bias in teacher ratings of behavior: A criterion-focused review. *Psychology in the Schools, 51*(10), 1017–1030. doi:10.1002/pits.21800

Mathai, J., Anderson, P., & Bourne, A. (2004). Comparing psychiatric diagnoses generated by the Strengths and Difficulties Questionnaire with diagnoses made by clinicians. *Australian & New Zealand Journal of Psychiatry, 38*(8), 639–643. doi:10.1080/j.1440-1614.2004.01428.x

McFadden, A. C., Marsh, G. E., Price, B. J., & Hwang, Y. (1992). A study of race and gender bias in the punishment of school children. *Education and Treatment of Children, 15*(2), 140–146. Retrieved from https://www.jstor.org/stable/42900466

McGrady, P. B., & Reynolds, J. R. (2013). Racial mismatch in the classroom: Beyond Black-White differences. *Sociology of Education, 86*(1), 3–17. doi:10.1177/0038040712444857.

Mieloo, C. L., Bevaart, F., Donker, M. C., van Oort, F. V., Raat, H., & Jansen, W. (2014). Validation of the SDQ in a multi-ethnic population of young children. *The European Journal of Public Health*, *24*(1), 26–32. doi:10.1093/eurpub/ckt100

Mortenson, B. C. (2018). *The role of teacher implicit bias in the racial achievement gap* (Doctoral dissertation). Retrieved from https://repository.library.georgetown.edu/handle/10822/1050845

Munzer, T. G., Miller, A. L., Brophy-Herb, H. E., Peterson, K. E., Horodynski, M. A., Contreras, D., . . . Lumeng, J. C. (2018). Characteristics associated with parent-teacher concordance on child behavior problem ratings in low-income preschoolers. *Academic Pediatrics*, *18*, 452–459. doi:10.1016/j.acap.2017.10.006

National Center for Education Statistics. (2012). *Schools and Staffing Survey (SASS).* Retrieved from https://nces.ed.gov/ccd/elsi/tableGenerator.aspx?savedTableID=72529

National Center for Education Statistics. (2016). *Schools and Staffing Survey (SASS).* Retrieved from https://nces.ed.gov/ccd/elsi/tableGenerator.aspx?savedTableID=72529

Netherlands Population 2019. (2019). Retrieved from http://worldpopulationreview.com/countries/netherlands-population/

Niclasen, J., Skovgaard, A. M., Andersen, A. N., Sømhovd, M. J., & Obel, C. (2013). A confirmatory approach to examining the factor structure of the Strengths and Difficulties Questionnaire (SDQ): A large scale cohort study. *Journal of Abnormal Child Psychology, 41*(3), 355–365. doi:10.1007/s10802-012-9683-y

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. Retrieved from http://www.ccsenet.org/journal/index.php/es

Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes: Race and the disciplining of young students. *Psychological Science*, *26*(5), 617–624. doi:10.1177/0956797615570365

O'Neil, J. M., Helms, B., Gable, R., David, L., Wrightsman, L., Pederson, E. L., & Vogel, D. L. (2007). Gender Role Conflict Scale--I. *Journal of Counseling Psychology, 54*(4), 373–384. doi:10.1037/0022-0167.54.4.373

Ouazad, A. (2014). Assessed by a teacher like me: Race and teacher assessments. *Education Finance and Policy, 9*(3), 334–372. doi:10.1162/EDFP_a_00136

Palmieri, P. A., & Smith, G. C. (2007). Examining the structural validity of the Strengths and Difficulties Questionnaire (SDQ) in a US sample of custodial grandmothers. *Psychological Assessment, 19*(2), 189–198. doi:10.1037/1040-3590.19.2.189

Pellegrini, A. D., Van Ryzin, M. J., Roseth, C., Bohn-Gettler, C., Dupuis, D., Hickey, M.,

& Peshkam, A. (2011). Behavioral and social cognitive processes in preschool children's social dominance. *Aggressive Behavior*, *37*(3), 248–257. doi:10.1002/ab.20385

Race. (n.d.). In *Merriam-Webster's collegiate dictionary* (10th ed.). Springfield, MA: Merriam-Webster.

Rasmussen, K. (2008). Halo effect. *Encyclopedia of Educational Psychology, 1*, 458–460. doi:10.4135/9781412963848.n125

Riley, T. A. (2014). Boys are like puppies, girls aim to please: How teachers' gender stereotypes may influence student placement decisions and classroom teaching. *Alberta Journal of Educational Research, 60*(1), 1–21. doi:10.1080/02607476.2013.864020

Rong, X. L. (1996). Effects of race and gender on teachers' perception of the social behavior of elementary students. *Urban Education, 31*(3), 261–290. doi:10.1177/0042085996031003003

Ruchkin, V., Koposov, R., Vermeiren, R., & Schwab-Stone, M. (2012). The Strength and Difficulties Questionnaire: Russian validation of the teacher version and comparison of teacher and student reports. *Journal of Adolescence, 35*(1), 87–96. doi:10.1016/j.adolescence.2011.06.003

Rutter, M., Yule, W., Berger, M., Yule, B., Morton, J., & Bagley, C. (1974). Children of West Indian immigrants. Rates of behavioural deviance and of psychiatric disorder. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 15*(4), 241–262. doi:10.1111/j.1469-7610.1974.tb01250.x

Saft, E. W., & Pianta, R. C. (2001). Teachers' perceptions of their relationships with students: Effects of child age, gender, and ethnicity of teachers and children. *School Psychology Quarterly, 16*(2), 125–141. doi:10.1521/scpq.16.2.125.18698

Sansone, D. (2017). Why does teacher gender matter? *Economics of Education Review, 61*, 9–18. doi:10.1016/j.econedurev.2017.09.004

Sargisson, R. J., Stanley, P. G., & Hayward, A. (2016). Multi-informant scores and gender differences on the Strengths and Difficulties Questionnaire for New Zealand children. *New Zealand Journal of Psychology (Online), 45*(2), 4. doi:10.1017/edp.2014.1

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Buehner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *European Journal of Research Methods for the Behavioral and Social Sciences, 6*, 147–151. doi:10.1027/1614-2241/a000016.

Scott, T. M., Gage, N., Hirn, R., & Han, H. (2019). Teacher and student race as a predictor for negative feedback during instruction. *School Psychology, 34*(1), 22–31. doi:10.1037/spq0000251

Shibata, H., Cattaneo, L. R., Leach, A. M., & Galloway, J. N. (2014). First approach to the Japanese nitrogen footprint model to predict the loss of nitrogen to the environment. *Environmental Research Letters, 9*(11), Article 115013. doi:10.1088/1748-9326/9/11/115013

Shifrer, D. (2013). Stigma of a label: Educational expectations for high school students labeled with learning disabilities. *Journal of Health and Social Behavior, 54*(4),

462–480. doi:10.1177/0022146513503346

Snowden, J. L. (2005). *Explicit and implicit bias measures: Their relation and utility as predictors of criminal verdict tendency*. Retrieved from https://libres.uncg.edu/ir/uncw/f/snowdenj2005-1.pdf

Staats, C. (2016). Understanding implicit bias: What educators should know. *American Educator, 39*(4), 29. Retrieved from https://www.aft.org/ae

Steffens, M. C. (2004). Is the implicit association test immune to faking? *Experimental Psychology, 51*(3), 165–179. doi:10.1027/1618-3169.51.3.165

Stone, L. L., Otten, R., Engels, R. C., Vermulst, A. A., & Janssens, J. M. (2010). Psychometric properties of the parent and teacher versions of the Strengths and Difficulties Questionnaire for 4-to 12-year-olds: A review. *Clinical child and Family Psychology Review, 13*(3), 254–274. doi:10.1007/s10567-010-0071-2

Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology, 99*(2), 253–273. doi:10.1037/0022-0663.99.2.253

Thomas, D. E., Coard, S. I., Stevenson, H. C., Bentley, K., & Zamel, P. (2009). Racial and emotional factors predicting teachers' perceptions of classroom behavioral maladjustment for urban African American male youth. *Psychology in the Schools, 46*(2), 184–196. doi:10.1002/pits.20362

Thys, S., & Van Houtte, M. (2016). Ethnic composition of the primary school and educational choice: Does the culture of teacher expectations matter? *Teaching and Teacher Education, 59,* 383–391. doi:10.1016/j.tate.2016.06.011

Tobia, K., Buckwalter, W., & Stich, S. (2013). Moral intuitions: Are philosophers experts? *Philosophical Psychology, 26*(5), 629–638. doi:10.1080/09515089.2012.696327

Ungvarsky, J. (2017). Implicit bias. In *Salem press encyclopedia*. Retrieved from https://www.salempress.com/encyclopedia_global_resources_2019

U.S. Department of Education. (2017a). *Number of children and students served under IDEA, Part B, by age group and state: 2016-17*[1] [Data file]. Retrieved from https://www2.ed.gov/programs/osepidea/618-data/static-tables/index.html

U.S. Department of Education. (2017b). *Number of children ages 3 through 5 served under IDEA, Part B, by disability and state: 2016-17* [Data file]. Retrieved from https://www2.ed.gov/programs/osepidea/618-data/static-tables/index.html

U.S. Department of Education. (2017c). *Number and percent of children in race/ethnicity category ages 6 through 21 with disabilities served under IDEA, Part B, by disability category and state* [Data file]. Retrieved from https://www2.ed.gov/programs/osepidea/618-data/static-tables/index.html

Van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal, 47*(2), 497–527. doi:10.3102/0002831209353594

van den Heuvel, M., Jansen, D. E., Stewart, R. E., Smits-Engelsman, B. C., Reijneveld, S. A., & Flapper, B. C. (2017). How reliable and valid is the teacher version of the Strengths and Difficulties Questionnaire in primary school children? *PloS*

One, 12(4). doi:10.1371/journal.pone.0176605

Westerberg, D. (2016). Understanding and dealing with implicit bias and discipline in early care and education. *Brown University Child & Adolescent Behavior Letter, 32*(10), 1–6. doi:10.1002/cbl.30155

Whitford, D. K., & Emerson, A. M. (2019). Empathy intervention to reduce implicit bias in pre-service teachers. *Psychological Reports*, *122*(2), 670–688. doi:10.1177/0033294118767435

Wright, A. C. (2015). *Teachers' perceptions of students' disruptive behavior: The effect of racial similarity and consequences for school suspension* (Doctoral dissertation). Retrieved from https://aefpweb.org.

Yates, T. M., & Marcelo, A. K. (2014). Through race-colored glasses: Preschoolers' pretend play and teachers' ratings of preschooler adjustment. *Early Childhood Research Quarterly, 29*(1), 1–11. doi:10.1016/j.ecresq.2013.09.003

Zwirs, B., Burger, H., Schulpen, T., Vermulst, A. A., HiraSing, R. A., & Buitelaar, J. (2011). Teacher ratings of children's behavior problems and functional impairment across gender and ethnicity: Construct equivalence of the strengths and difficulties questionnaire. *Journal of Cross-Cultural Psychology*, *42*(3), 466–481. doi:10.1177/0022022110362752

Appendix A: Strengths and Difficulties Questionnaire-Teacher Form 4-10 Permission

Re: The use of the SDQ in a Dissertation- Permission

Youthinmind [redacted]

Mon 2/8/2016, 1:40 AM

Amanda Haas [redacted]

Dissertation

Dear Amanda,

Thank you for your interest in the SDQ.

If the dissertation is printed (i.e. on paper) and will not be distributed, it's fine to include the SDQ. It will have to appear exactly as on sdinfo.com. You will need to include columns and rows just like the paper SDQ – a slavish copying of everything (including title, preamble and copyright notice).

However, even if the dissertation is on paper, then if it is going to be published or widely disseminated, this is not allowed.

Best wishes
Helena Hamilton
YIM

Appendix B: Instruction Form for Same Race and Same Gender

INSTRUCTIONS

Thank you for your participation. For this study, please complete one SDQ-TF on one student who has persistent behavioral issues and you judge as actively demonstrating problematic behaviors. You must have worked with this student for a minimum of 6 months and have first-hand knowledge about their behaviors.

*Please complete this form for a student who is the **same race** and **same gender** as yourself.*

Please answer the following questions:

*What is your gender?*_____
*What is your race?*_____
*What grade do you currently teach?*_____
*How many years have you taught?*_____

Appendix C: Instruction Form for Same Race and Different Gender

INSTRUCTIONS

Thank you for your participation. For this study, please complete one SDQ-TF on one student who has persistent behavioral issues and you judge as actively demonstrating problematic behaviors. You must have worked with this student for a minimum of 6 months and have first-hand knowledge about their behaviors.

*Please complete this form for a student who is the **same race** and **different gender** as yourself.*

Please answer the following questions:

*What is your gender?*_____
*What is your race?*_____
*What grade do you currently teach?*_____
*How many years have you taught?*_____

Appendix D: Instruction Form for Different Race and Same Gender

INSTRUCTIONS

Thank you for your participation. For this study, please complete one SDQ-TF on one student who has persistent behavioral issues and you judge as actively demonstrating problematic behaviors. You must have worked with this student for a minimum of 6 months and have first-hand knowledge about their behaviors.

*Please complete this form for a student who is the **different race** and **same gender** as yourself.*

Please answer the following questions:

*What is your gender?*_____
*What is your race?*_____
*What grade do you currently teach?*_____
*How many years have you taught?*_____

Appendix E: Instruction Form for Different Race and Different Gender

INSTRUCTIONS

Thank you for your participation. For this study, please complete one SDQ-TF on one student who has persistent behavioral issues and you judge as actively demonstrating problematic behaviors. You must have worked with this student for a minimum of 6 months and have first-hand knowledge about their behaviors.

*Please complete this form for a student who is the **different race** and **different gender** as yourself.*

Please answer the following questions:

*What is your gender?*_____
*What is your race?*_____
*What grade do you currently teach?*_____
*How many years have you taught?*_____