2019

Teacher Biases as an Influence on Early Childhood Assessments

Crystal Mason
Walden University

Walden University

College of Education

This is to certify that the doctoral dissertation by

Crystal Mason

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. Patricia Anderson, Committee Chairperson, Education Faculty
Dr. Andrew Alexson, Committee Member, Education Faculty
Dr. Jennifer Lapin, University Reviewer, Education Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2019

Abstract

Teacher Biases as an Influence on Early Childhood Assessments

by

Crystal Mason

MA, Walden University, 2012

BS, CSUSB, 2006

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Education

Walden University

August 2019

Abstract

Teacher perspectives and judgments of students' race and gender are known to influence their assessment of primary and secondary students' academic achievements. However, little is known about the effect on children's academic achievement of preschool teacher perspectives and judgments of students' race and gender, which forms the basis for this study. The purpose of this study was to analyze teacher assessment of preschool children's mathematics and science skills on the Desired Results Developmental Profile (DRDP) and Teacher Strategies GOLD (TSG), along with teacher comments written in preparation for each child's parent-teacher conference, to determine if there was a relationship between preschool teachers' assessment and comments and the race and gender of the child. Wason's theory of confirmation bias formed the theoretical foundation of this study. The research questions addressed the relationship between preschool teacher assessments recorded on the DRDP and TSG regarding children's mathematics and science skill and teacher comments coded from Racasens linguistic model and those children's race and gender. Archival data from 2 Head Start centers in a western and southwestern state were analyzed using the Mann-Whitney U test, and the point-biserial Pearson correlation. The Mann-Whitney U test found no statistically significant differences in DRDP and TSG scores by students' race and gender. The point-biserial Pearson correlation found no statistically significant correlation between race or gender and teacher comments. This study contributes to positive social change by confirming observational assessments to be free from teacher bias, supporting their continued use with preschool children to promote their learning and development.

Teacher Biases as an Influence on Early Childhood Assessments

by

Crystal Mason

MA, Walden University, 2012

BS, CSUSB, 2006

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Education

Walden University

August 2019

Dedication

This study is dedicated to Mrs. Heidi, my high school home economics teacher, who fostered my passion for Early Childhood Education. Also, to Mrs. Connie, a former preschool cook, whose encouragement and positive attitude got me to pursue my doctorate in Early Childhood Education.

Acknowledgments

First and foremost, I like to acknowledge my mother and godmother for teaching me that you can achieve anything through hard work and determination. Secondly, I would like to acknowledge Dr. Patricia Anderson, who, as my advisor for my KAMS and committee chairperson, has never given up on me, when I wanted to give up on myself. Lastly, I would like to acknowledge my closest friends, Chris, Anna, and Susan, for being there for me throughout this dissertation journey.

Table of Contents

List of Tables

Chapter 1: Introduction to the Study

Early education teachers are trained to be excellent observers of children (Casbergue, Bedford & Burstein, 2014). Through their observations teachers can identify their students' actual abilities and plan curriculum to meet students' needs for intellectual and social development. Such observations are central to assessment in early childhood, when standardized paper-and-pencil assessments are inappropriate.

However, observation is open to bias. According to Brawley and Stormont (2014), observations of young children are affected by teacher bias. For example, Gilliam Maupin, Reyes, Accavitti and Shic (2016) asked early education teachers to watch short videos of four young children, including an African American boy, African American girl, Caucasian boy, and Caucasian girl. Each time the teachers saw a behavior they believed was a challenging behavior, they were asked to press the enter key on the external keypad. Gilliam et al. (2016) concluded that the teachers watched the African American children more closely than they did the Caucasian children and that more focus was placed on the African American boy than on any other of the children, even though all four children's behavior was similar. Gilliam et al. discovered that teachers' presumptions of the likely source of challenging behavior were racially biased.

Ishmine and Tayler (2014) found that early childhood teachers' biases influence student achievement and student/teacher relationships. However, research is lacking on how preschool teachers' biases might influence outcomes of early childhood assessment tools such as the Desired Result Developmental Profile (DRDP) and Teaching Strategies GOLD (TSG). The DRDP is a qualitative assessment that relies on teacher observation in

assigning scale values to children's accomplishment of benchmark skills and behaviors. It includes narrative comments by teachers in support of their numerical assignments. Similar to the DRDP, the TSG is a qualitative assessment that links observable behaviors to essential early learning requirements (Burts & Kim, 2014). In this study, I will explore the extent to which teachers' biases influence assessments on the DRDP and TSG.

In this chapter, I describe the background of my study, lay out the problem statement, explain the purpose of the study, and introduce the research question. I describe my theoretical framework that was based on Wason's idea of confirmation bias. Also, I clarify the nature of the study including definitions, assumptions, scope, delimitations, and limitations. Lastly, I further justify the significance of this study and conclude the chapter with a summary.

Background

Teachers' perspectives have been known to influence children's perspectives of their own abilities (Upadyaya & Eccles, 2015). Upadyaya and Eccles (2015) found that when teachers believed that their students' had innate abilities in mathematics and reading, the students' self-concept of their own mathematics and reading abilities was high. In contrast, when teachers believed that their students' mathematics and reading abilities were low, students' self-concept of their own mathematics and reading abilities declined. Upadyaya and Eccles (2015) investigated the extent in which primary teachers' perspectives of their students' abilities and effort predicted developmental changes in children's mathematics and reading abilities. Upadyaya and Eccles (2015) found that teachers' perspectives of their students' abilities affected the students' self-concept with

regard to their own academic abilities in mathematics and reading. Upadyaya and Eccles found that teachers' perspectives of their students' abilities affected the students' self-concept with regard to their own academic abilities in mathematics and reading.

Teachers' judgments have been known to predict adult intelligence and further key life outcomes such as educational attainment and socioeconomic achievement (Fischbach, Baudson, Preckel, Martin, & Brunner, 2013). Fischbach et al. (2013) addressed three questions: whether teacher judgments of 12-year-old students affect those students' scores on an intelligence test; whether there are long-term effects of teacher judgments on student development in 12-year-old children; and whether the effects of teacher judgments of 12-year-old children generalize across a broader range of key life outcomes. They concluded that teacher judgments of intelligence reflected student achievement rather than intelligence. They also found that teacher judgments predicted important life outcomes (including intelligence) across a time span of 40 years, and that the predictive power of teacher judgments remained even after controlling for childhood intelligence. In short, Fischbach et al. found that teachers' judgments of 12-year-old students do in fact predict adult intelligence.

Teachers who rely on their perspectives and judgments of children often underestimate some children's cognitive abilities, which hinders children's motivation (Mega, Ronconi, & De Beni, 2014). A key component of successfully obtaining academic achievement is to set realistic goals (Komarraju & Nadler, 2013). However, teachers who rely on their perspectives and judgments of children often set goals for children either too high or low (Rubie-Davis, Peterson, Sibley, & Rosenthal, 2015).

Teacher perspectives and judgments affect teacher and student relationships (McCormick & O'Connor, 2015). If the teacher has negative perspectives and judgments of a child, that teacher is more likely to treat the child negatively, which prevents the child from establishing a positive relationship with the teacher (McCormick & O'Connor, 2015).

Teachers' perspectives and judgments of a child are often based on ascriptive characteristics such as race and gender (Mason, Gunersel, & Ney, 2014). In a study involving assessment of secondary education children, Weisman (2012) found that students' race, gender, and extraneous factors, including whether the teacher liked the student, influenced assessments. McGrady and Reynolds (2013) discovered that teachers' biases showed favoritism towards 10th grade Caucasian students compared to 10th grade African American students. Teachers' perspectives and judgments of a child are also based on the student's gender. According to Cornell, Mustard, and Parys (2013), kindergarten teachers' grades for boys were lower than boys' actual test scores, indicating an assessment based on bias.

Most of the literature about the influence of teacher perspectives and judgments of students' race and gender have focused on primary, secondary, and high-school students' academic achievement (Südkamp, Kaiser, & Möller, 2012). What was unknown is if teacher perspectives and judgments of students' race and gender also affect their assessment of preschool children's academic achievement. Because the researchers have indicated that academic achievement among public education students has been influenced by teachers' perspectives and judgments, the influence of teachers'

perspectives and judgments on preschool assessments such as the DRDP and TSG was investigated.

Problem Statement

Children's assessments are a vital component of any early education program, but many early educators do not assess students' actual abilities. Instead, teachers assess students by what they think the child can achieve (Krolak-Schwerdt, Böhmer, & Gräsel, 2013). Teachers rely on not only what they perceive the child can do, but also their judgments of the students' abilities (Campbell, 2015). When teachers rely on perspectives and judgments of students' abilities, the students' academic performance can be affected (Südkamp, Kaiser, & Möller, 2012).

Teachers' perspectives and judgments in assessments have had a profound influence in connection to students' ascriptive characteristics such as race and gender (Campbell, 2015; Gilliam, 2016; Ouazad, 2014). To determine if elementary teachers assess same-race children more favorably than they do children of a different race from themselves, Ouazad (2014) compared subjective assessments by teachers and children's test scores from kindergarten to fifth grade. He concluded that African American and Hispanic children who were assessed by Caucasian teachers were graded significantly lower than the same children graded by teachers of the same race (Quazad, 2014). In another study, using a sample of 5,000 7-year-old children, Campbell (2015) reviewed assessments of children according to students' race and economic status. She concluded that inequalities in the children's assessments reflected stereotypical attitudes regarding race and household income (Campbell, 2015).

Like race, students' gender can influence teachers' perspectives and judgments in assessments. Falch and Naper (2013) investigated gender gaps in teachers' assessments of eighth to 10th grade students. They concluded that girls were awarded higher grades than boys when the same skills were assessed by their teacher (Falch & Naper, 2013). Krkovic, Greiff, Kupiainen, Vainikainen, and Hautamäki (2014) also concluded that teachers tended to evaluate sixth grade girls more positively than sixth grade boys in first language performance and as having higher potential for success in education, without corroborating evidence from objective tests.

Previously, researchers on the effect of teacher perspectives of the students' race and gender on student assessment focused on students in primary and high school (Ouazad, 2014). However, education does not start at kindergarten, but at the preschool level. Also, most of the research on race and gender in student assessments only analyzed standardized test scores of the children (Campbell, 2015). However, in most preschools, teachers use anecdotal assessment such as the DRDP and TSG (Brawley & Stormont, 2014). The possible effect of teacher perception on preschool children's DRDP and TSG assessment formed the problem of this study.

Purpose of the Study

The purpose of this study was to analyze teacher assessment of children's mathematics and science skills on the DRDP and TSG along with teacher comments written in preparation for each child's parent-teacher conference to determine if there is a relationship between teachers' assessment and comments and the race and gender of the child.

Research Questions and Hypotheses

Two research questions (RQs) guided this quantitative study:

RQ1: Is there a statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's race?

$H_0$1: There is no statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's race?

H_a1: There is a statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's race?

RQ2: Is there a statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's gender?

$H_0$2: There is no statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's gender.

H_a2: There is a statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's gender.

Theoretical Foundation

In this study, I examined the relationship between preschool children's race and gender and teachers' assessment of the children's success in mathematics, as indicated on existing DRDP and TSG preschool assessment records. As noted above, it is possible that biased perspectives affect teacher assessments, in that predetermined expectations based on race or gender may influence what the teacher notices and remembers in support of those expectations. Similar to the findings Gilliam et al. (2016) reported in the study described above of teachers' observations of children's behavior, confirmation bias may be a factor in assessment on qualitative measures like the DRDP and TSG. Confirmation bias is a type of cognitive bias and a systematic error of inductive reasoning in which an individual gathers information that confirms preexisting beliefs or hypotheses (Wason, 1968).

In order to demonstrate the phenomenon of confirmation bias, Wason conducted a series of experiments in the 1960s to demonstrate that people are indeed biased towards confirming their existing beliefs. In one experiment, Wason presented subjects with four cards each labelled with a letter on one side and a number on the other side (A, D, 4, 7). The purpose of the experiment as it was explained to participants was to test the hypothesis that if a card has a vowel on one side, then it has an even number on the other side. Wason began by having the subjects flip over the cards until they believed the hypothesis was proven or rejected. The majority of the participants only flipped over an even numbered card (4) and the vowel card (A) to see what was on the other side. Only a few of the subjects correctly chose the (A) and (7) cards. Wason's explanation for this

occurrence was that the subjects were trying to confirm the statement that if a card has a vowel on one side, then it has an even number on the other side. This explanation lead Wason to conclude that most people confirmed the statement, when in fact the task was to determine if the statement was false (Wason & Johnson, 1970).

The belief that people immediately favor information that validates their preconceptions, hypotheses, and personal beliefs is supported by research of the implicit biases of teachers. Jacoby-Senghor, Sinclair, and Shelton (2016) examined the role implicit biases have on teachers' expectations of students' academic achievements. The researchers began by recruiting teachers and having them take part in one of two studies. In Study 1, the researchers assigned participants to the role of an instructor and asked each to present a short lesson to a learner; these lessons were videotaped. In Study 2, the participants viewed one of the videos taken from Study 1 and then were asked to complete a measure of explicit bias based on the video they watched. The researchers were able to conclude from both studies that the instructors' implicit bias predicted low expectation for the African American students. These findings suggested that underperformance by minorities in education may be driven by the effect implicit racial bias has on educators (Jacoby-Senghor et al., 2016). A more detailed explanation will be pretended in Chapter 2. The theory of confirmation bias provided the theoretical foundation for this study because, even though seeking out information that confirms one's beliefs comes naturally, to make valid assessments teachers must look for evidence that is true even it if contradicts their beliefs. Once teachers can look for instances that

prove them wrong, they will no longer look to confirm their biases when assessing children.

Nature of the Study

In this quantitative study, I used an analysis of pre-existing DRDP and TSG assessment along with teacher comments written in preparation for each child's parent-teacher conference to determine if a relationship exists between preschool teacher assessment regarding children's mathematics and science skill and those children's race and gender, and to determine if a relationship exists between teachers' reports regarding children's mathematics and science skill and those children's race and gender. Although a correlational study cannot determine any cause-and-effect relationship, I assumed in both analyses that race and gender form two independent variables that influence teachers' assessment of children's mathematics and science skill as indicated on the DRDP and TSG in both child ratings and teachers' comments, which form the dependent variables. Existing data from DRDPs filed by three to four preschool teachers for their classes of 14 to 21 children (N = 50 to 60) was examined. Also, existing data from TSG filed by three to four preschool teachers for their classes of 20 to 21 children (N = 50 to 60) were examined using a linguistic model for detecting biased language (Recasens, Danescu-Niculescu-Mizil & Jurafsky, 2013). A Mann-Whitney U test and point-biserial Pearson correlation are the statistics that were applied to determine any relationship or relationship between achievement labels from the DRDP, TSG, teachers' comments, and children's race or gender.

Definitions

Bias: The unequal assessment between two alternatives, which typically puts one option in a favorable position and the other in an unfavorable one (Navarro, 2019).

Desired Results Developmental Profile (DRDP): Instrument designed for teachers to observe, document, and reflect on the learning, development, and progress of children, birth through 12 years of age (California Department of Education, 2016).

Teachers Strategies GOLD (TSG): An authentic, ongoing observational system for assessing children from birth through kindergarten (Lambert, Kim, & Burts, 2015).

Assumptions and Limitations

I assumed that teachers speak and write English as a condition of employment as lead teachers, so that their use of English in comments used in this study represents familiarity with the nuances of that language. This study was limited to two organizations that focus on early education, so the results were not generalizable to the entire population of early childhood programs. Also, the study focus was on only the mathematics and science section of the assessment tool, which restricted generalizability of results in this study to other domains. Another limitation of this study was that the analysis done with the DRDP did not include the use of the linguistic model for detecting biased language (Recasens et al., 2013). By not using the linguistic model for detecting biased language for the DRDP assessment, biased language that was detected was only limited to the TSG assessment.

Scope and Delimitations

The scope of this study included analysis of teacher assessment of preschool children's mathematics skills on the DRDP and the mathematics and science skills on the TSG along with teachers' comments written in preparation for parent-teacher conferences to determine the strength of the relationship between teachers' assessment and comments and the race and gender of the child. This study was delimited to include the mathematics section only of 67 DRDPs and the mathematics and science section only of 58 TSGs and teachers' comments from two early childhood development centers in a major city in the western United States and southwestern United States.

Significance

In this study, I analyzed teacher assessment of preschool children's mathematics and science skills on the DRDP and TSG along with teachers' comments to determine any relationship between teachers' ratings of children's mathematics and science achievement and children's race and gender, and the strength of the relationship between teachers' comments about children's mathematics and science achievement and the race and gender of children. The results of this study provided insight by which to evaluate the validity of the DRDP, TSG, and similar assessments that are based on teacher perspectives, and therefore may improve the quality of child assessment and interpretation of assessment results.

Summary

In this chapter the background of this study was presented in which teachers' perspectives and judgments were addressed along with the effect they have on students'

assessments, leading to the problem of preschool teachers' perspectives and judgments in regards to the race and gender of the students on a qualitative assessment like the DRDP and TSG. The purpose of this study was to analyze teacher assessment of children's mathematics and science skills on the DRDP and TSG, along with teacher comments written in preparation for each child's parent-teacher conference, to determine if there is a relationship between teachers' assessment and comments and the race and gender of the child. Wason's ideas of confirmation bias formed the theoretical foundation.

Chapter 2: Literature Review

The possible effect of teacher perception on preschool children's DRDP and TSG assessment formed the problem of this study. The purpose of this study was to analyze teacher assessment of children's mathematics and science skills on the DRDP and TSG along with teacher comments written in preparation for each child's parent-teacher conference to determine if there is a relationship between teachers' assessment and comments and the race and gender of the child. To facilitate this task, I present research on teachers' perspectives and the influences they have on teachers' judgment and the effect of these factors on student achievement and student assessments. In this chapter, I include my literature search strategies, the conceptual foundation of the study, and a review of current research pertinent to the purpose of this paper.

Literature Search Strategy

I accessed databases searched for this literature review using the Walden University Library and included Sage Premier, ERIC, PsycNET, Google Scholar, EBSCO Host, Yahoo, and Theses at Yale. I limited the search to peer-reviewed journal articles published within the past six years. I also used government and agency websites such as those of the National Association for the Education of Young Children (NAEYC) and the California Department of Education to supply information about early childhood assessments. Search terms included: *teacher perceptions, teacher judgments, teacher bias, biases in preschool, bias in assessments, implicit biases, preschool teacher implicit biases, racial bias, gender bias, gender biases in preschool,* and *gender and teacher biases in preschool*.

Theoretical Foundation

The theory of confirmation bias by Wason (1968) provides the foundation for this study. Before Wason's psychological research on confirmation bias, the occurrence of confirmation bias had been observed throughout history. The Greek historian Thucydides believed that it was, "the habit of an individual to entrust in careless hope to gain what is wanted while using sovereign reason to set aside what he does not fancy" (as cited in Crawley,1914, p.316). In the 1620 work, *Novum Organum Scientiarum* Bacon (as cited in Burtt, 1939, p.36), stated that once an individual adopts an opinion, "the individual draws everything into confirming and supporting that opinion, even if the greater number of instances are found to support elsewhere." These historical references indicate that individuals' tendency to continue to believe what they already believe is a commonplace impulse in Western culture.

The term, *confirmation bias*, however, was first coined by Wason (1968). According to Wason, once an individual has formed a belief about a hypothesis, the individual will seek out information that supports that hypothesis. To demonstrate the notion that people are biased towards confirming their existing beliefs, Wason conducted an experiment known as the four card task. In Wason's four card task, participants were presented with four cards that were laid flat on a table. The first two cards displayed had numbers face up while the other two cards had colors face up. Wason then asked the participants which card must be turned over to test the hypothesis that if a card shows an even number on one face, then its opposite face is blue. Most of the participants only flipped over an even card and the blue card. Wason's explanation for this occurrence was

that the participants were trying to confirm the statement when in fact the task was to determine if the statement was false (Wason & Johnson, 1970).

Wason's next experiment that explained why people make certain consistent mistake in their reasoning is known as the THOG problem. In the THOG problem a participant is shown four designs: black diamond, white diamond, black circle, and white circle. The participant is then told to assume that the researcher has written down one of the shapes and one of the colors. Now, the participant is asked to read the following rule carefully, "if and only if, any of the designs includes either the color that I wrote down or the shape I wrote down, but not both, this is called a THOG" (Wason & Brooks, 1979, p. 80). The researcher then stated that "the black diamond is a THOG. Which one of the other designs is also THOG?" (Wason & Brooks, 1979, p.80). More than half the participants who attempt this problem get it wrong. Most people state that the white circle is not a THOG and that the other two designs are THOGS, however; the correct answer is the white circle. Wason's explanation for this occurrence was that universally people choose and arrange events in an order they assume is correct, regardless of the fact that the rule tells otherwise (Wason & Brooks, 1979).

Wason's theory of confirmation bias helped demonstrate that people favor information that gives support for their perspectives or hypotheses, regardless of whether those ideas hold true (Wason & Johnson, 1970). The belief that people favor information that gives support for their perspectives or hypotheses is further supported by research of Jacoby-Senghor et al. (2016) on the implicit biases of individuals. To explore implicit biases of individuals, the researchers engaged undergraduate participants in two types of

studies. Prior to the start of the studies, participants completed a cognitive flexibility task that in reality was a subliminal priming task that assessed implicit racial bias. In Study 1, participants were assigned the role of instructor and were asked to present a short lesson to a learner who was either Caucasian or African American. In Study 1 researchers found that greater instructor implicit bias predicted lower test performance for African American learners but not for Caucasian learners. To rule out the alternative hypothesis that lower performance of African American learners in Study 1 was driven by those learners' concerns about being targets of prejudice, in Study 2 another group of participants included African American observers who watched the videotaped lessons were asked to report any apparent bias. Study 2 showed that the African American observers also perceived instructors' implicit bias (Jacoby-Senghor et al., 2016). Knowing that teachers do hold implicit bias of their students' makes it likely that those biases may affect their assessments of the students.

Confirmation bias not only affects how people gather information, but also influences their interpretation of the gathered information (Wason, & Johnson, 1970). Knowing that confirmation bias can affect both the gathering of information and the interpretation of the information helps explain why people hold biases. The effect of confirmation bias on the gathering and people's renditions of the gathered information is relevant to this study because both are vital components in making reliable preschool assessments (Goldstein & Flake, 2016). However, to the extent that teachers seek to confirm preexisting biases in their assessment of children, the reliability of the children's assessment scores diminishes (Goldstein & Flake, 2016). The theory of confirmation bias

provides the foundation of this study because the impulse to favor information that supports preexisting beliefs appears to be universal (Wason & Johnson, 1970). In the following review of the current literature, I will address the influences teachers' perspectives and judgments may have on their students' academic success. In addition, I will consider in this literature review the possibility that teachers' perspectives and judgments are related to student characteristics, specifically to race and gender.

Teacher Perspectives Influence Judgments of Students

Numerous researchers have found that teachers' perspectives influence their judgments of their students' (Hansen, 2016; Robinson-Cimpian, Lubienski, Ganley, & Copur-Gencturk, 2014a). Pas and Bradshaw (2014) used quantitative analysis with a sample of 702 general education teachers from 42 elementary schools across three school years to discover that teachers who reported fewer concentration problems, emotional dysfunctional, and aggressive behaviors among their students' rated students more highly for pro-social behavior. Also, these teachers had more favorable perspectives of their school environment (Pas & Bradshaw 2014). Similarly, O'Brennan, Bradshaw, and Furlong (2014), in a large-scale study, found that teachers' perspectives of their school's climate were significantly related to their perspectives of students' behaviors (O'Brennan et al., 2014).

Teachers' relationships with the students are affected by teachers' perspectives. For example, Baker, Tichovolsky, Kupersmidt, Voegler-Lee, and Arnold (2015) found in a study of 123 teachers, that teachers who severely underestimated their preschool students' assessed abilities had considerably weaker relationships with their students'.

Also, the researchers found that teachers both under- and overestimated the academic abilities of their preschoolers compared to the objective assessments of their skills, based on teachers' perspectives of factors such as the child's gender and social skills (Baker et al., 2015). For example, girls frequently were described as being proficient in reading and writing while boys frequently were described as being proficient in science and mathematics. In addition, socially competent students' academic abilities were often exaggerated by their teachers (Baker et al., 2015).

In discovering that preschoolers' abilities are either under- or overestimated, Baker et al. (2015) did not mention the characteristics of the teachers. One characteristic known to be related to teachers' judgments is the teachers' professional experiences. Baudson, Fischbach, and Preckel (2016) found, in a study of 1774 primary school children from first to third grade and 95 of their teachers, that child and family variables, such as household income and ethnicity and teachers' professional experiences, including their years of teaching, affected teacher judgments of students' cognitive ability. Mayer, Wiley, Wiley, Dees, and Raiford (2016) found that experienced teachers demonstrated more fairness in their expectations of students' abilities than less experienced teachers (Mayer et al., 2016).

Teachers with similar characteristics as their students are more likely to rely on their perspectives of the students, not students' abilities as demonstrated on formal assessments (McCormick & O'Connor, 2015). Rausch, Karing, Dorfler, and Artelt (2016) found that teachers who shared similarities with their students were more likely to judge the students more favorably than they were to judge favorably students who were

different from themselves. Gershenson, Holt, and Papageorge (2016) also found that teachers who are female tended to judge their female students more favorably than their male students. Caucasian teachers have been found to favor Caucasian students over African American students (Egalite, Kisida & Winters, 2015; Ouazad, 2014). This is problematic because teacher perspectives influence their judgments of students (Campbell, 2015; Fischbach et al., 2013).

In a series of three studies, Kaiser, Retelsdorf, Sudkamp and Moller (2013) found a correlation between sixth grade teachers' judgments of their students' achievement and engagement and their students' actual achievement and engagement. The students who were judged positively by their teachers were noted as being high achieving students who were more engaged in their school work, suggesting a virtuous cycle between teacher approval and student performance (Kaiser et al., 2013). Teachers' judgments have been known to affect students' motivation which then affects their performance in school. Urhahne (2015), in a study of 246 sixth grade students and 13 of their teachers, students who were described by their teachers as low performing showed lower motivation than students who were described as high performing. It is known that teachers' perspectives influence their judgments of students (Hansen, 2016; Robinson-Cimpian, Lubienski, Ganley, & Copur-Gencturk, 2014a). This is especially true for student attributes such as the students' race and gender.

Teachers' Perspectives and Students' Race

Specific characteristics such as a students' race influence teachers' perspectives which then influence their judgments (Blanchard & Muller, 2015; Glock, Krolak-

Schwerdt, & Cate, 2015). Faulkner, Stiff, Marshall, Nietfeld and Crossland (2014), in a study of 11,260 students from third to fifth grade and 3,055 students from eighth grade, found that African American students' odds of being placed in advanced mathematics courses such as algebra were reduced by two-thirds to two-fifths compared to the chances of Caucasian students. Since teachers are less likely to place African American students in advanced courses such as algebra, this makes it highly likely that their expectation for African American academic success is low. Similarly, Burgess and Greaves (2013) found that for pupil's ages 5 to14 Black Caribbean, Black African, Pakistani, and Bangladeshi students' academic abilities were underestimated in comparison to Indian and Chinese students' abilities. Along with the students' race, the students' age, family socioeconomic status, and gender have been found to influence teachers' perspectives (McGrath & Van-Bergen, 2015).

Teachers' perspectives of African American students' makes it more likely that they will be labeled as troublemakers. Okonofua and Eberhardt (2015), in a study of 57 female teachers from K-12 school districts, found that African American students were significantly more likely than Caucasian students to be labeled a trouble maker. They also reported that African American students' misbehaviors were more likely than Caucasian students' misbehaviors to be based on an apparent tendency of teachers to report some students repeatedly for the same offenses. In addition, Anyon et al. (2014) showed that when comparing students with similar characteristics, African American students remain at a much higher risk of getting a disciplinary referral than are Caucasian students. Finn and Servoss (2014), in a study of 8,775 10th grade students, showed that African

American students were suspended at a higher rate than Caucasian students. Gilliam and Reyes (2005) found that preschoolers are three times more likely to be expelled compared to K-12 students. Even more concerning, Gregory et al. (2016) discovered racial disparities in expulsion rates. He found that African American students were two times more likely than Caucasian and Hispanic students to be expelled (Gregory et al., 2016). Multiple researchers have come to the conclusion that a student's race influences teachers' perspectives (Blanchard & Muller, 2015; Glock et al., 2015).

Teachers' Perspectives and Students' Gender

Students' gender also has been found to influence teachers' perspectives and their judgments of students' (Falch & Naper, 2013; Krkovic, Greiff, Kupiainen, Vainikainen, & Hautamäki, 2014). McGrath and Van-Bergen (2015) found, in a meta-analysis of 92 studies, that students' gender places some at heightened risk of experiencing a negative student-teacher relationship. The researchers found that girls have higher quality relationships with their teachers compared to boys. Boys, especially older boys, had more conflicting relationships with their teachers and were given less academic support than girls.

Retelsdorf, Schwartz, and Asbrock (2015), in a study of 54 fifth and sixth grade teachers and 1,358 of their students, found that teachers' stereotypical perspectives negatively affected teachers' judgment of boys' reading achievement compared to their actual performance more than did teachers' perspectives of the reading abilities of girls. One explanation for this occurrence is because boys are perceived by teachers as being less competent in the areas of reading and writing (Wolter, Braun, & Hannover, 2015).

Rimm-Kaufman, Baroody, Larsen, Curby, and Abry (2014), in a similar study of 387 fifth grade students, found that girls were judged to be more behaviorally engaged than boys in mathematics and that girls reported higher cognitive and social engagement than boys in mathematics classrooms. One possible explanation for this occurrence is that teacher support is given primarily to girl students due to the perception that boys are more competent in the areas of mathematics and science than are girls (Baker et al., 2015), in opposition to the motivation presumed for teachers differential treatment of boys and girls in reading. Together, these findings suggest that teachers perceive girls as more deserving of their attention than boys, regardless of boys' and girls' need for help in the subject at hand.

Teacher Biases in Assessments

Students' assessments are often based on what the teachers observe (Ishimine & Tayler, 2014). However, this can lead to teachers relying on their biases when assessing students' abilities (Brawley & Stormont, 2014). Research has shown that teachers' rating of their students has a lower association to school performance than the students' actual observed skills (Fortin, Oreopoulos, & Phipps, 2015; Sprietsma, 2013). Robinson-Cimpian, Lubienski, Ganley, and Copur-Gencturk (2014b) found, in a study with kindergarten through 8th grade students, that when both boys and girls demonstrate equivalent mathematics test performance, girls are scored less proficient in mathematics than boys. Similarly, Banjong (2014), in a study of 129 students ranging from 4th to 7th grade, found that there was no significant relationship in mathematics performance level

between boy and girl students, but that girls felt their mathematics abilities were limited due to teachers who scored them below average.

Teachers who rely on their perspectives to influence their judgments of their students' abilities are less effective in enhancing students' academic abilities. Soleimani and Rahmanian (2014) found, in a sample of 90 English-as-a-foreign-language (EFL) female students ages 17 to 19, that teacher assessment was not as effective as self-assessment and peer-assessment in terms of enhancing female students' writing proficiency. Similarly, Graham, Hebert and Harris (2015) discovered, in a meta-analysis of 35 formal assessment studies of students in grades 1st to 8th, that formal assessments alone are not effective in enhancing students' academic abilities. However, along with daily feedback to the students, formal assessments can be effective in enhancing the students' academic abilities (Graham et al., 2015). This raises concerns because assessments rely on the teacher accurately assessing the students' abilities in order to use those assessments effectively, but, as research has shown, teachers rely on their biases when assessing students' abilities (Brawley & Stormont, 2014).

Most studies have focused on teacher perspectives and student outcomes in the public school system, including primarily the upper grades. Yet teachers' perspectives may also affect assessment of children even in the preschool, where assessment is often less standardized and more anecdotal than it is in elementary.

Early Education Assessments

The earliest significant interest in understanding the development of preschool children was shown by Gesell at the Yale Clinic for Child Development in 1916 (Thelen

& Adolph, 1992). Believing that growth and development were biologically predetermined, Gesell argued for a maturational perspective that incorporated age and stages. To test his assumption, Gesell conducted both a psychological examination and an observational study at the child's home with 50 children at each of 10 age levels beginning at birth and ending at age five (Thelen & Adolph, 1992). Since little attention was put forth into the precise methodology of the study, the results were presented as a developmental schedule that consisted of 150 items based on what was considered the four core areas of child development. These areas included motor development, language development, adaptive behaviors, and personal-social behaviors. Following Gesell's study, several assessment tools for infants and preschool children began to emerge for use with young children (Thelen & Adolph, 1992). Examples include the Palmer Scale of Mental Tests, the Minnesota Preschool Scale, and the Iowa Test for Young Children (Bracken, 2004; "Guide to Assessment in Early Childhood; Infancy to Age Eight," 2008). Although the reliability and validity of these early scales would be considered questionable by today's standards (Bracken, 2004), the formation of these early childhood assessments paved the way for further research on other types of assessment tools.

In early education, assessment methods are either informal or formal. Informal methods are what teachers use every day to evaluate the progress and the comprehension abilities of each student. These assessments come in many forms such as observations, portfolios, educator ratings, and parent ratings (Mertler, 2016). According to Mertler (2016), observations are the first option teachers use because they can be done with

minimal intrusion into children's daily activities. Portfolios are another option for teachers because they allow the teacher to collect records of the students' work over a period of time. Educator ratings are used as an option for teachers because they allow the teacher to link their ratings to other methods of assessments such as standardized testing (Mertler, 2016). Similarly, parents' ratings are used as an option because they too can be linked to other methods of assessments such as standardized testing. Also, parents' ratings can be a vital component in informal assessment methods because they help detect target milestones that may have been overlooked by teachers (Mertler, 2016).

Formal methods of assessment measure how well students have mastered specific learning outcomes (Black, 2015). These assessments may take the form of a questionnaire or a standardized test. Questionnaires are used as an option because they help the teacher gain feedback on effective center practices. Standardized testing is often described as excellent objective indictor of student performance (Mertler, 2016).

Early childhood assessments are more likely to fall into two systems. The first system includes program-developed child assessment tools, which are made to align with a specific program's philosophy and curriculum (McLachlan, Fleer, & Edwards, 2018). The second system includes child assessment tools which are independent of any particular program and are used to measure specific areas in child development (McLachlan et al., 2018). Examples of these second, program-independent assessment tools include the Preschool Program Quality Assessment (PQA), the DRDP (Baneerjee & Luckner, 2013) and TSG.

Unlike program-developed child assessment tools, program-independent child assessment tools were developed and empirically tested with the intention of creating a reliable tool for assessing children's development. However, the reliability of program-independent child assessment tools such as the PQA, DRDP and TSG are questionable due to their reliance on teachers' observations. The PQA is a rating instrument designed to evaluate the quality of early childhood programs as well as identify staff training needs (Keys et al., 2013). The components that the PQA examines are the physical characteristics of the setting, the nature of adult-child interactions, and program management (Keys et al., 2013). There are two parts of the PQA that are labeled form A and form B. Form A primarily consists of teachers' observations as the support for the first four domains of the assessment. Form B is gathered primarily through interviews and is the support for the last three domains of the assessment. The DRDP is an observation tool teachers use to document and reflect on the learning, development, and progress of young children in four desired results areas (Sutter et al., 2017). These four desired results areas include personal and social competence, effective learning abilities, physical and motor competence, and personal safety and wellness. The TSG is an observation-based assessment that is grounded in 38 research-based objectives that are aligned with state early learning standards (Lambert et al., 2015). These 38 objectives are broken into 10 domains which include social-emotional, physical, language, cognitive, literacy, mathematics, science and technology, social studies, the arts, and English language acquisition. Similar to the PQA and DRDP, TSG relies heavily on teachers' observations

Summary and Conclusions

Teachers' perspectives of students' abilities are known to have an effect on students' academic success. This is especially true for students of a specific race and gender. For instance, African American students, especially preschool students, are more likely to be negatively affected by teachers' perspectives than are Caucasian or Hispanic students (Gilliam & Reyes, 2018). Boys are more likely to be perceived by their teachers as being more competent in mathematics and science (Baker et al., 2015), while girls are more likely to be perceived by their teachers as being more competent in reading and writing (Wolter et al., 2015), yet girls receive more teacher support in both subject areas (Baker et al., 2015). These differences are concerning because they suggest that teachers may not accurately assess students' abilities, but rely on their perspectives of the students' abilities to make judgments (Hansen, 2016; Robinson-Cimpian et al., 2014a). More concerning is that most assessments, especially preschool program-independent child assessment tools, such as the PCA, DRDP, and TSG rely heavily on teachers' observations (Keys et al., 2013; Lambert et al., 2015; & Sutter et al., 2017).

Chapter 2 included an evaluation of the literature on ways in which teacher perspective influence their judgments of students' and the student characteristics that seem to have an influence on teachers' perspectives. Also, in this chapter I presented a discussion of early childhood assessments and biases in assessments. In Chapter 3, I discuss the methodology and data collection process.

Chapter 3: Research Method

The purpose of this study was to analyze teacher assessment of children's mathematics and science skills on the DRDP and TSG along with teacher comments written in preparation for each child's parent-teacher conference to determine if there is a relationship between teachers' assessment and comments and the race and gender of the child. The study involved examining the relationship between pre-existing DRDP, TSG assessments, and teacher comments written in preparation for each child's parent-teacher conference, utilizing the linguistic model for detecting biased language suggested by Recasens et al. (2013). The results help provide information by which to evaluate the fairness of the DRDP, TSG, and similar assessments that are based on teacher perspectives, and therefore lead to improvement in the quality of child assessments.

In this chapter I address the design of this research study and my reasoning for selecting a quantitative correlational method. This chapter includes a description of the research questions, research design and rationale, my role as the researcher, and methods by which I select participants and gather and analyze data. I also address any threats to the validity of my study results and means by which I ensure ethical protection of participants.

Research Questions

Two research questions (RQs) guided this quantitative study:

RQ1: Is there a statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's race?

$H_0$1: There is no statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's race?

H_a1: There is a statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's race?

RQ2: Is there a statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's gender?

$H_0$2: There is no statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's gender.

H_a2: There is a statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's gender

Research Design and Rationale

Determining if early childhood teachers' assessments of children and their comments show significant relationships to the race and gender of each child required quantitative analysis of available data. Because bias is unconscious (Benaji & Greenwald, 2013), a qualitative research method by which teacher assessments and comments are solicited through interviews would be unlikely to detect bias that is unnoticed by the teachers themselves. Qualitative research using observation of teacher interactions with

children to assess possible racial or gender bias that affect teacher assessments and comments also would be vulnerable to unconscious bias, this time of the observer, since all individuals harbor unconscious bias (Benaji & Greenwald, 2013). For these reasons, I rejected a qualitative design. An experimental approach, in which teachers might be randomly assigned to teach classes organized by race or gender, was rejected because of the unworkability of this arrangement in conventional preschool settings, and because of ethical issues surrounding the use of children as an independent variable. A correlational method within a quantitative framework was selected for this study because it was the appropriate way to determine the strength of the relationship between data from preschool students' DRDP, TSG, and teacher comments written in preparation for each child's parent-teacher conference and helped to establish if there was an association between the data and students' race and gender. My selection of a correlational approach, instead of a descriptive or experimental approach, helped me to recognize trends and patterns in the data but not determine their cause.

Role of the Researcher

In my role as a professional in early education, I have assisted and taught at a federally funded Head Start center for over five years, as well as volunteered at a few Head Start centers. While I was working at a Head Start center, I was required to complete DRDP assessments of children ages three to five and write notes to be used as the basis for parent-teacher conferences about each student. When I was volunteering at a Head Start center, my task was to listen with preschool teachers as they were conducting parent-teacher conferences. My professional experiences have made me aware of the

possibility for bias in teacher assessment of children using the DRDP, TSG, and in teacher comments prepared to be shared with parents. Currently, I am not affiliated with any of the centers from which data for this study was collected, nor did I have a professional or personal relationship with individuals who work at any of the centers.

Methodology

Participant Selection Logic

I collected existing data in the form of DRDP assessments for approximately 60 children ages three to five. Also, I collected existing data in the form of TSG assessment of children and their comments written for parent-teacher conferences for approximately 60 children ages three to four. Approximately 120 student records were chosen because the RaoSoft sample size calculation determine that the minimum number of usable student records for a confidence level of 95% is 92 (RaoSoft, 2004), a likely result from the proposed data collection method. The existing data came from teachers of children ages three to five who are employed in two centers in both a mid-sized city in a western state and a mid-sized city in a southwestern state. I gathered data from federally funded Head Start programs, because such centers often have a larger group of students in each class than other types of centers (up to 24 children) and each class is staffed by one fully qualified teacher who holds a professional teaching certificate and has completed 12 units in early childhood development, and one or two aides. Teachers in federally funded Head Start programs also meet the criterion for this study of requiring teachers to complete early childhood assessments such as the DRDP and TSG. Federally funded Head Start

centers are also required to provide parents with written comments for parent-teacher conferences to discuss their students' progress.

Selection of the centers that assess children using the DRDP was based on the Child Care Resource Center (CCRC) website that allows individuals to search licensed child care providers in counties near enough to me to facilitate in-person interviews. The selection of the centers that assess children using the TSG was based on a colleague who worked with state licensed child care centers in the southwestern United States. The pool of possible centers met the following criteria: be designated as federally funded Head Start centers, and provided comments written for parent-teacher conferences for the same children. Also, the centers enrolled similar numbers of children who are from different races. The first two organizations that allowed their preschool centers to provide me access to their archival data were accepted as data sources. Since this quantitative study was based on the students' archival data, approximately 60 students' DRDP and 60 students' TSG assessments and their comments written for parent-teacher conferences were needed.

Instrumentation

The first instrument in this study consisted of the DRDP preschool assessments for approximately 60 preschool children. The DRDP is a formative assessment instrument developed by the California Department of Education that is used to assess young children in eight domains (Sutter et al., 2017). The DRDP offers no validity statistics but, according to the California Department of Education (2013), has been found to correlate with other types of early childhood assessments such as the Battelle

Developmental Index (BDI), which is a comprehensive developmental assessment tool used with infants and young children. The strength of the relationship between each domain of the DRDP and each domain of the BDI ranges from moderate to strong. For example, there is a strong correlation between the DRDP's cognitive domain and the BDI's cognitive abilities domain, which was statistically significant ($r = .56$, $p = .005$) (California Department of Education, 2013).

In this study, the cognitive domain that includes mathematics was the domain analyzed on the DRDP. On the DRDP, the mathematics domain includes seven rating areas on a 9-point scale. These rating areas included spatial relationships, classification, number sense of quantity, number sense of mathematics operations, measurement, patterning, and shapes (Sutter et al., 2017). Child ratings on the mathematics domain formed ordinal data.

In the DRDP, achievement in each domain was assessed in four categories, from least to most developed: responding, exploring, building, and integrating (Sutter et al., 2017). The first category, *responding*, included knowledge and behaviors that are developed through the senses. The second category of *exploring* included knowledge and behaviors based on the child's active exploration of the world around them. The third category, *building*, included knowledge and behaviors based on the child's understanding of how people and objects relate. The final category, *integrating*, included knowledge and behaviors based on the child's ability to connect strategies in order to solve problems (Sutter et al., 2017).

The second instrumentation in this study consisted of TSG preschool assessments for approximately 60 preschool children and their comments written for parent-teacher conferences for the same children. The TSG is an ongoing, observation based assessment tool that is grounded in 38 research-based objectives for development and learning. When analyzing the validity of each assessment item's fit within the six areas of development, statistics such as the root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), and a comparative fit index (CFI) were reported (Burts & Kim, 2014). The RMSEA=.066, with a cut off value <0.06 for good fit. The SRMR=. 033, with a cut off value <0.08 for good fit. The CFI= .931 with a cut off value of ≥.90 for good fit. All of these analyses were statistically significant at $p < .001$. Several analyses were done to determine the TSG internal consistency reliability (Burts & Kim, 2014). The internal consistency reliability estimates ranged from .957 for the physical scale to .980 for the cognitive scale, which represents extremely high internal consistency reliability. Interrater reliability between the teachers' rating and the master teachers' rating, correlations was high at .80 (Burts & Kim, 2014).

In this study, mathematics and science were the only dimensions analyzed on the TSG. On the TSG, the mathematics dimensions included four subsections that were broken into 12 rating areas on a 9-point scale (Lambert et al., 2015). These rating areas include counts, quantifies, connects numerals with their quantities, understands and uses place values and base ten, applies properties of mathematical operations and relationships, applies number combinations and mental number strategies in mathematical operations, understands spatial relationships, understands shapes,

understands objects, measures time and money, represents and analyzes data, and demonstrates knowledge of patterns. On the TSG, the science dimensions include four rating areas on a 9-point scale (Lambert et al., 2015). These rating areas included uses scientific inquiry skills, demonstrates knowledge of the characteristics of living things, demonstrates knowledge of the physical properties of objects and materials, and demonstrates knowledge of earth's environment.

In TSG, achievement in each domain was assessed in four categories, from least to most developed: beginning, processing, increasing, and advancing. The first category, *beginning*, included knowledge and behaviors expected for children birth to age two. The second category of *processing* included knowledge and behaviors expected for children ages two to three. The third category, *increasing,* included knowledge and behaviors expected for children ages three to four. The final category, *advancing,* included knowledge and behaviors expected for children ages four to five (Lambert et al., 2015).

In completing the DRDP and TSG, teachers indicate the race and gender of each child. Gender is designated as either "male" or "female." Race on the assessments includes 18 choices along with a 19th choice "intentionally left blank." To simplify analysis, the 18 choices were organized into four choices, including Black or African American, White, Latino, and Other Ethnicity. These teacher determinations were used as categorical variables in organizing data from the DRDP and TSG. This information is vital for this study because characteristics such as a student's race influence teachers' perspectives (Blanchard & Muller, 2015; Glock et al., 2015). Students' gender also has

been found to influence teachers' perspectives (Falch & Naper, 2013; Krkovic et al., 2015).

Along with the DRDP and TSG preschool assessment, teachers' comments written for the parent-teacher conferences were used in this study. Parent-teacher conferences often consist of a description of the strengths of the child, the child's current developmental focus, and the child's need for targeted support, if any (Kroth & Edge, 2017). Comments were analyzed using the linguistic model for detecting biased language developed by Recasens et al. (2013). Comments made in reference to each child were reviewed for epistemological and framing bias by three independent raters, yielding a decimal score reflecting the proportion of biased words to the total number of words in each comment.

Procedures for Recruitment, Participation, and Data Collection

In this quantitative study, the recruitment process began with a list of centers provided by the CCRC website that allowed individuals to search licensed child care providers in counties near me. Directors of early childhood organizations were called and asked if they met the criteria described above. Once the organizations confirmed that it met the criteria, an explanation of the purpose of the study was given followed by a request for permission to analyze their students' DRDPs or TSBs and comments written for parent-teacher conferences. In doing so, the directors of the organizations were informed that all identifying information about the center, teachers, and students would remain unknown to me because a professional not associated with the study would collect the data. These professionals, one for each center, signed a confidentiality agreement

prior to collecting data that included keeping confidential all information that could be used to identify a specific center, teacher, or student.

Once I received permission for my colleagues to retrieve the archival data from the directors of the early childhood organizations, my colleagues went to the centers. One colleague collected the DRDP preschool assessments and the other colleague collected the TSG preschool assessments and comments written for parent-teacher conferences.

One colleague made photocopies of the following: first page of the DRDP, the page that indicates child ratings on the cognitive domain that includes mathematics. The second colleague made photocopies of the following: first page of the TSG, the page that indicates child ratings on the mathematics and science, and the student's parent-teacher conferences. On the first page of the DRDP and TSG assessments, the only information that was needed about the student was the student's race and gender. Other information, on that page and on other collected pages that could identify the child, teacher, or center were covered so this information did not appear on photocopies made of the assessments. My colleagues made photocopies of these pages and returned the pages to the center director. All photocopied pages for individual students were stapled, and my colleague labeled each stapled set with a number to serve as a student identifier. These numbers were assigned in simple numerical order, with no additional indicator that might connect the pages to any particular child, teacher, or center. This process took approximately one hour for each colleague to complete data gathering of approximately 60 student records for the DRDP and TSG.

Data Analysis Plan

The first step in data analysis was to calculate a composite rating for each child in the cognitive domain that included mathematics understanding, using teacher assessment on each child's DRDP and TSG assessment. DRDP child ratings in the cognitive domain that included mathematics, child gender, and child race were entered into a spreadsheet. A rating on each of the seven factors in this domain ranged from 1 to 8, so the composite rating on all factors for each child in the cognitive domain that included mathematics ranged from 7 to 56. Next, the TSG child ratings in mathematics and science, child gender, and child race were entered into the spreadsheet. A rating on each of the 12 factors in the mathematics domain ranged from 1 to 9, so the composite rating on all factors for each child in the mathematics domain ranged from 12 to 108. A rating on each of the four factors in the science domain ranged from 1 to 3, so the composite rating on all factors for each child in the science domain ranged from four to 12.

The second step in the data analysis was to use the Mann-Whitney U test to determine the relationship between each child's composite rating in the DRDP (7-56) and TSG (12-108) mathematics domain by the child's race (Black or African American, White, Latino, and Other Ethnicity). A second Mann-Whitney U test was run to determine the relationship between each child's composite rating in the DRDP (7-56) and TSG (12-108) mathematics domain by the child's gender (boy or girl). To determine the relationship between each child's composite rating in the TSG (4-12) science domain by the child's race (Black or African American, White, Latino, and Other Ethnicity), another Mann-Whitney U test was used, followed with a Mann-Whitney U test to determine the

relationship between each child's composite rating in the TSG (4-12) science domain by the child's gender (boy or girl). Since race and gender are nominal factors, and the composite ratings were ordinal data, a Mann-Whitney U test was employed to determine the degree of relationship between ratings in mathematics for the DRDP and TSG and science for the TSG.

Following initial analysis of the DRDP and TSG data, I analyzed the comments made per child by teachers in anticipation of parent-teacher conferences, to determine any evidence of bias in these comments. To do this, I relied on linguistic modeling described by Recasens et al. (2013). Biased language, according to Recasens et al., falls under two major classes: epistemological biases and framing biases. Epistemological bias is evident when propositions are assumed to be true or false. Epistemological bias is divided into four subtypes: factive verbs, entailments, assertive statements, and hedges (Recasens et al, 2013). Factive verbs are verbs that, whether stated as positive or negative, imply a difference in the meaning of an embedded presupposition. For instance, "Cathy knew Josh could count to 10" and "Cathy didn't know Josh could count to 10" convey the same meaning but imply, in the second sentence, the possibility of surprise at Josh's ability to count and so might indicate bias concerning Josh's accomplishment. Entailments refers to a relationship between two words in which the second word must be true if the first word is true. For example," the teacher was coerced into accepting the bribe." Assertive verbs are those whose complement clauses assert a proposition. For example, "she clearly stated that Susan is a good reader." Hedges involve reducing one's commitment to the truth of a proposition. For example, "David may have taken Billy's

toy car." (Recasens et al., 2013). Another major class of biases includes framing biases that occur when subjective or one-sided words are used. They are divided into two subtypes: intensifiers and one-sided terms. Intensifiers are adjectives or adverbs that add force to the meaning (Recasens et al., 2013). For example, "Amber accurately placed the square into the right area." One-sided terms are reflections that give only one part of an issue. For example, "Ms. Smith said that David was freely given the choice to sit at circle time."

Teacher comments were transcribed into a comment file, with comments for each child labeled with the child identifying number coupled with a comment number, so that comment 1 for child 1 appeared as 1-1 and comment 2 for child 1 appeared as 1-2. Word counts per comment were calculated using the word count tool of Microsoft Word. Three identical copies of this comment file were printed. A spreadsheet was prepared to indicate child number, child gender, child race, and word count per comment with additional columns available for inserting biased comment counts and biased comment averages per comment.

Because linguistic analysis is open to bias, several steps were taken to preserve the integrity of the process. First, comments pertaining to each child were entered into a new spreadsheet on which only the child's identifying number indicated their connection to other data on the original spreadsheet. The child's gender, race, and assessment composite score did not appear on the linguistic analysis spreadsheet or on the three copies of the comment file. In addition, I enlisted the help of two other professional persons to serve as raters, so that linguistic analysis was performed three times using

comments for each child. Each rater received a copy of the comment file and instructions for making the linguistic analysis. For each comment, the types of linguistic bias detected in each comment was coded by the three independent raters and the number of biasing words counted (Appendix A). When all three raters completed their analysis of the comments included on the comment file, I and the two additional raters met and compared our bias scores for each child's comments. In the case of a difference of opinion about the level of bias on a particular comment, we three raters discussed the comment and arrived at a consensus score for that comment.

Finally, the number of biasing words were entered into the second spreadsheet in the columns reserved for this purpose and the ratio of biasing words to total word count calculated, to arrive at a bias score for each child's comments. At this point the data were organized to permit a determination of the relationship between bias and gender and bias and race. I used a point-biserial Pearson correlation on SPSS to determine whether there was a statistical association between teachers' comments and the race and gender of the child. A point-biserial Pearson correlation was used for this study because it is suitable for nominal dichotomous variables such as race (Black and White, Black and Latino, and Latino and White) and gender (Male and Female) and ratio data such as decimal fractions from the teacher comments.

Threats to Validity

In this study there were threats to validity that may have made this study less effective than it might otherwise be. With reference to internal validity, maturation is a threat to this present study. As children continue in their education, they become better

educated, which is shown in the DRDP and TSG data from the entire year. However, the parent-teacher conferences are not for the entire year. I cannot be sure that what the teachers said in their comments was actually based on the students' assessment scores. In regard to external validity, this study had a small sample size so the results were not generalizable to the larger population of early childhood programs. Also, the study focused on just the mathematics and science section of the assessment tools, which restricts generalizability of results in this study to include the other domains in the DRDP and TSG.

Ethical Protection of Participants

I obtained the approval from Walden's IRB before gathering the data (11-16-18-0061400). A data agreement letter was signed by an authorized member of each participating organization prior to data collection. My colleagues who photocopied data at each center, and the two professionals who served as independent raters, each signed a confidentiality agreement to preserve the confidentiality of material they accessed. My colleagues who photocopied the data covered identifying information on each page so that no information that identifies the centers, teachers, or students was visible to me or anyone else associated with the study. Once entered into spreadsheets or transcribed into the comment file, the original data were stored in a locked file cabinet to which only I have access. All data containing the students' identifying number, race, and gender were stored on my computer's external thumb drive. Following review of comments by myself and two independent raters, the comment files were retrieved from the two independent

raters and shredded, leaving only a single file used to record the final biasing comments rating following inter-rater consensus. This file also was stored in the locked file cabinet

Summary

For this quantitative study, the purpose of examining the relationship between pre-existing DRDP and TSG assessments in the mathematics, and science domain and teacher comments written in preparation for each child's parent-teacher conference, were discussed. Chapter 3 included a detailed description of the quantitative correlational method and the reasoning for selecting this method. This chapter also included a description of the research questions, research design, and rationale. My role as the researcher, and the methods by which I selected participants and gathered and analyzed data were also discussed. Lastly, any threats to the validity of my study results and means by which I ensured ethical protection of participants concluded this chapter. In Chapter 4, I present the findings from this study.

Chapter 4: Results

The purpose of this study was to analyze teacher assessment of children's mathematics and science skills on the DRDP and TSG along with teacher comments written in preparation for each child's parent-teacher conference to determine if there is a relationship between teachers' assessment and comments and the race and gender of the child. The RQs were the following:

RQ1: Is there a statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's race?

$H_0$1: There is no statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's race?

H_a1: There is a statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's race?

RQ2: Is there a statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's gender?

$H_0$2: There is no statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's gender.

H_a2: There is a statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's gender

After gathering the archival data from two early childhood centers, I coded and analyzed the collected data. I then organized these data according to students' mathematics and science scores on each assessment, including teacher comments where available. In this chapter, I present the data collection process and my analysis of the data organized by research question.

Data Collection

The existing data came from preschool teachers who are employed to work with three- to five-year-old children in two centers in a mid-size city in a western state (DRDP data) and a mid-sized city in a southwestern state (TSG data). I gathered data from centers that were federally funded Head Start programs because such centers often have a larger group of students in each class than other types of centers (up to 24 children) and each class is staffed by one fully qualified teacher who holds a professional teaching certificate and has completed 12 units in early childhood development, and one or two aides. Federally funded Head Start centers also meet the criterion for this study of requiring teachers to complete early childhood assessments such as the DRDP and TSG along with providing parents with written comments for parent-teacher conferences to discuss their students' progress.

Once I received permission from the directors of the two early childhood organizations included in this study to retrieve the archival data, one colleague collected

the DRDP preschool assessments from one center and the other colleague collected the TSG preschool assessments and comments written for parent-teacher conferences from the second center.

The first colleague made photocopies of the following pages: the first page of the DRDP, and the DRDP cognitive domain, which indicated child ratings in mathematics. The second colleague made photocopies of the first page of the TSG and the TSG mathematics and science objectives. The second colleague also made photocopies of the students' parent-teacher conference comments. On the first pages of the DRDP and TSG assessments, the only information that was needed about the student was the student's race and gender. Other information, on that page and on other collected pages that could reveal the identity of the child, teacher, or center was covered so this information did not appear on photocopies made of the assessments. My colleagues then made photocopies of these pages and returned the pages to the center. All photocopied pages for individual students were stapled, and my colleagues labeled each stapled set with a number to serve as the student identifier. These numbers were assigned in simple numerical order, with no additional indicator that connected the pages to any particular child, teacher, or center. This process took approximately one hour for each colleague to complete data gathering.

When collecting the data for the early childhood assessments, the initial plan was to collect 120 DRDP assessments in the mathematics domain and the science domain. However, many of the early childhood organizations that assessed children using the DRDP would not permit outside individuals to access their archival data. The organization that allowed me access to their data used a form of the DRDP assessment

that did not include the science domain. This organization provided a total of 67 DRDP children's mathematics assessments. Parent conference comments did not accompany the DRDP assessment data because of restrictions imposed by the center policies. I chose TSG data as my second data set because the assessment is similar to the DRDP and a director at a center that used TSG agreed to participate in the study. My colleague collected 55 TSG children's mathematics and science assessments and 46 written parent-teacher conference comments. Parent-teacher conference comments were not collected for nine children because parents of those nine children had not participated in a parent-teacher conference.

Data Analysis

The first step in data analysis was to calculate a composite rating for each child in the cognitive domain that included mathematics understanding, using teacher assessment on each child's DRDP and TSG assessment. DRDP child information in the mathematics cognitive domain and child gender and child race were entered into a spreadsheet. A rating on each of the seven factors in this domain ranged from 1 to 8, so the composite rating on all factors for each child in the cognitive domain that included mathematics ranged from 7 to 56. Next, the TSG child ratings in mathematics and science, along with child gender and child race, were entered into a spreadsheet. A rating on each of the 12 factors in the mathematics domain ranged from 1 to 9, so the composite rating on all factors for each child in the mathematics domain ranged from 12 to 108. A rating on each of the four factors in the science domain ranged from 1 to 3, so the composite rating on all factors for each child in the science domain ranged from four to 12.

To account for missing math and science measures that were apparent on the TSG data set, the multiple imputation method was used. The multiple imputation method was the best method to choose instead of the series means because more than 5% of the data was missing. Using the series means for a percent higher than five could lead to biases in the data analysis.

The second step in the data analysis was to use the Mann-Whitney U test to determine the relationship between each child's composite rating in the DRDP (7-56) and TSG (12-108) mathematics domain by the child's race (Black or African American, White, Latino, and Other Ethnicity), followed with a second Mann-Whitney U test to determine the relationship between each child's composite rating in the DRDP (7-56) and TSG (12-108) mathematics domain by the child's gender (boy or girl). To determine the relationship between each child's composite rating in the TSG (4-12) science domain by the child's race (Black or African American, White, Latino, and Other Ethnicity), a Mann-Whitney U test also was used, followed with a second Mann-Whitney U test to determine the relationship between each child's composite rating in the TSG (4-12) science domain by the child's gender (boy or girl). Since race and gender are nominal factors, and the composite ratings were ordinal data, a Mann-Whitney U test was employed to determine the degree of relationship between ratings in mathematics for the DRDP and TSG and science for the TSG.

Following initial analysis of the TSG data, I analyzed the comments made by teachers in parent-teacher conferences, to uncover any evidence of bias in these comments. To do this, I relied on linguistic modeling described by Recasens et al. (2013).

Teacher comments were first transcribed into a comment file, with comments for each child labeled with the child identifying number coupled with a comment number, so that comment 1 for child 1 appeared as 1-1 and comment 2 for child 1 appeared as 1-2. Word count per comment was calculated using the word count tool of Microsoft Word. Three identical copies of these comment files were printed. A spreadsheet was prepared to indicate child number, child gender, child race, and word count per comment with additional columns available for inserting biased comment counts and biased comment averages per comment.

Because linguistic analysis is open to bias, several steps were taken to preserve the integrity of the process. First, comments pertaining to each child were entered into a new spreadsheet on which only the child's identifying number indicated their connection to other data on the original spreadsheet. The child's gender, race, and composite scores for mathematics and science did not appear on the linguistic analysis spreadsheet or on the three copies of the comment file. In addition, I enlisted the help of two other professional persons to serve as raters, so that linguistic analysis was performed three times using comments for each child. Each rater received a copy of the comment files and instructions for making the linguistic analysis. For conference comments that were written in Spanish, I enlisted a colleague to translate the documents into English. This person signed a confidentially agreement: she was not told that the translated comments would be analyzed for bias and she did not take part in the bias coding process. For each comment, the types of linguistic bias detected in each comment were coded by the three independent raters and the number of biasing words selected from the word list was

counted (Appendix A). When all three raters completed their analysis of the comments included on the comment file, I and the two raters met and compared our bias scores for each child's comments. In the case of a difference of opinion about the level of bias on a particular comment, we three raters discussed the comment and arrived at a consensus score for that comment.

Finally, the number of biasing words was entered into the second spreadsheet in the columns reserved for this purpose and the ratio of biasing words to total word count calculated, to arrive at a bias score for each child's comments. At this point the data were organized to permit a determination of the relationship between bias and gender and bias and race. Since gender and race are nominal factors, and the bias scores were decimal fractions, a point-biserial Pearson correlation was used to determine the strength of the relationship between teachers' comments and the race and gender of the child.

Results

A Mann-Whitney U test was used to determine whether there is a statistically significant relationship between teachers' assessment and the race and gender of the child in the DRDP and TSG early childhood assessment. The Mann-Whitney U test was used because gender and race are nominal data and the composite ratings are ordinal data.

Results for RQ1

RQ 1 asked, "Is there a statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's race?" The independent variable was race. I organized the analysis for six pair-wise comparisons of assessed skills between children identified as Black and Latino, Black and Other Ethnicity, Latino and Other Ethnicity, White and Black, White and Latino, and White and Other Ethnicity. The dependent continuous variable was mean scores in mathematics and science where scores ranged from 16% to 95% for mathematics and 33% to 100% for science. The null hypothesis stated that no statistically significant relationship existed between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's race. If the *p* value of the Mann-Whitney U test was greater than 0.05, then there were no significant differences between the groups and the null hypothesis was to be accepted.

A Mann-Whitney U test was run to determine the strength of the relationship between mathematics scores and race of children identified as Black and Latino. Distributions of the mathematics scores for children identified as Black and Latino were similar, as assessed by visual inspection. Mathematics scores were not statistically significantly different between Blacks (*Mdn* = .54) and Latinos (*Mdn* = .47), resulting in $U = 1013$, $z = -.534$, $p = .593$.

Distributions of the mathematics scores for each of these race groups (Black and Other Ethnicity, Latino and Other Ethnicity, Latino and White, Black and White, Other

Ethnicity and White) were not similar, as assessed by visual inspection. Mathematics scores for Black students (mean rank =18.97) and Other Ethnicity students (mean rank = 24.00) were not statistically significantly different ($U = 92$ $z = -1.12$, $p = .265$). Mathematics scores for Latino students (mean rank = 38.18) and Other Ethnicity students (mean rank =51.06) were not statistically significantly different ($U = 188$, $z = -1.53$, $p = .127$). However, mathematics scores for Latino students (mean rank = 44.70) were statistically significantly higher than the mathematics scores for White students (mean rank = 27.46, $U = 266$, $z = -2.37$, $p = .018$). Mathematics scores for Black students (mean rank = 25.47) were statistically significantly higher than the mathematics scores for White students (mean rank = 15.42), $U = 110$, $z = -2.37$, $p = .018$). Also, mathematics scores for Other Ethnicity students (mean rank = 15.25) were statistically significantly higher than the mathematics scores for White students (mean rank = 8.38, $U = 18$, $z = -2.46$, $p = .014$). See Table 1.

Table 1.

Difference Between Math Scores for Race

Score	Sample	Mean Rank	Median	Z score	P value
Math Scores					
Black	31	53.34	.536	-.534	.593
Latino	70	49.96	.446		
Black	31	18.97	.536	-1.12	.265
Other Ethnicity	8	24.00	.580		
Latino	70	38.18	.446	-1.53	.127
Other Ethnicity	8	51.06	.580		
Black	31	25.47	.536	-2.37	.018
White	13	15.42	.265		
Latino	70	44.70	.446	-2.37	.018
White	13	27.46	.265		
Other Ethnicity	8	15.25	.580	-2.46	.014
White	13	8.38	.265		

A Mann-Whitney U test was run to determine the strength of relationships between race and science scores in pair-wise comparisons of children identified as Black and Latino, Black and Other Ethnicity, Latino and Other Ethnicity, White and Black, White and Latino, and White and Other Ethnicity. Distributions of the science scores for each of these race groups were not similar, as assessed by visual inspection. Science scores for Latino student's (mean rank = 24.84) were statistically significantly higher than the science scores for Black (mean rank = 16.25, U =117, z = -2.00, p = .045). No statistically significant differences were found for science scores in any of the pairs:

Black (mean rank = 7.04) and Other Ethnicity (mean rank = 6.50, $U = 6$, $z = -.136$, $p = .891$); Latino (mean rank = 17.28) and Other Ethnicity (mean rank =8.00, $U = 7$, $z = -957$, $p = .339$); White (mean rank = 11.80) and Black (mean rank = 11.25, $U = 57$, $z = -.204$, $p = .839$); White (mean rank =15.90) and Latino (mean rank = 23.25, $U = 104$, $z = -1.68$, $p = .093$); and White (mean rank =6.05) and Other Ethnicity (mean rank= 5.50, $U = 5$, $z = -165$, $p = .869$). See Table 2.

Table 2.

Difference Between Science Scores for Race

Score	Sample	Mean Rank	Median	*p* value
Science Scores				
Black	12	16.25	.592	.045
Latino	32	24.84	.667	
Black	12	7.04	.592	.891
Other Ethnicity	1	6.50	.583	
Latino	32	17.28	.667	.339
Other Ethnicity	1	8.00	.583	
Black	12	11.25	.592	.839
White	10	11.80	.583	
Latino	32	23.25	.667	.093
White	10	15.90	.583	
Other Ethnicity	1	6.05	.583	.869
White			.583	
	10	5.50		

A point-biserial Pearson correlation was run between teacher comments and race in the same pair-wise comparisons as before. Preliminary analyses showed that (a) there

were no outliers, as assessed by boxplot for each dichotomous race group; (b) teacher comments were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$); and (c) there was homogeneity of variances, as assessed by Levene's test for equality of variances for each dichotomous race group. There was no statistically significant correlation between race and teacher comments, $r_{pb}(17) = .268$, $p = .268$, with Black students having more biased words used than White students (.093 ± .043 versus .072 ± .036). Race accounted for 7.2% of the variability in teacher comments for Black and White students. There was no statistically significant correlation between race and teacher comments, $r_{pb}(35) = .142$, $p = .403$, with Black students having more biased words used than Latino students (.093 ± .043 versus .082 ± .031). Race accounted for 2% of the variability in teacher comments for Black and Latino students. There was no statistically significant correlation between race and teacher comments, $r_{pb}(34) = .141$, $p = .411$, with Latino students having more biased words used than White students (.082 ± .031 versus .072 ± .036). Race accounted for 2% of the variability in teacher comments for Latino and White students (see Table 3).

Table 3.

Point-biserial Pearson Correlation between Race Groups and Teacher Comments

	Black & White		Black & Latino		Latino & White	
	r	*p*	*r*	*p*	*r*	*p*
Teacher Comments	.268	.268	.142	.404	.141	.411

In summary, a Mann-Whitney U test was used to determine whether there was a statistically significant relationship between teachers' assessment and the race of the child in the mathematics and science domains of the DRDP and TSG early childhood

assessments. The results showed for mathematics that Black, Latino, and Other Ethnicity student scores were statistically significantly higher than for White students. For science, Black and Latino student scores were statistically significantly different. However, for the pairs Black and Other Ethnicity, Latino and Other Ethnicity, White and Black, White and Latino, and White and Other Ethnicity, student scores were not statistically significantly different.

To determine whether there was a statistical association between teachers' comments and the race of the child, a point-biserial Pearson correlation was run between race and teacher comments. There was no statistically significant correlation between race and teacher comments in each of the dichotomous race groups (Black & White, Black & Latino and Latino & White). The results for RQ1 indicate the need to reject the null hypothesis for specific subsets of children identified by race.

Results for RQ2

RQ2 asked, "Is there a statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's gender?" The independent variable was gender, identified as male and female. The dependent continuous variable was mean scores in mathematics and science where scores ranged from 16% to 95% for mathematics and 33% to 100% for science. The null hypothesis stated that there was no statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's gender. If the *p* value of the Mann-Whitney U test was

greater than 0.05, than there were no significant differences between the groups and the null hypothesis was accepted.

A Mann-Whitney U test was run to determine if there were relationships in mathematics scores between children identified as male and female. Distributions of the mathematics and science scores for Male and Female were similar, as assessed by visual inspection. Mathematics scores were not statistically significantly different between Males (*Mdn* = .51) and Females (*Mdn* = .38, *U* = 1763, *z* = -.500, *p* = .617). Science scores were not statistically significantly different between Males (*Mdn* = .67) and Females (*Mdn* = .67, *U* = 357, *z* = -.300, *p* = .764). See Table 4.

Table 4.

Difference between Math and Science Scores for Gender

Scores	Sample	Mean Rank	Median	Z score	*p* value
Math Scores					
Male	60	63.13	.509	-.500	.617
Female	62	59.94	.382		
Science Scores					
Male	60	27.30	.667	-.300	.764
female	62	28.58	.667		

A point-biserial Pearson correlation was run between gender and teacher comments. Preliminary analyses showed that (a) there were no outliers, as assessed by boxplot for each gender group; (b) teacher comments were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$); and (c) there was homogeneity of variances, as

assessed by Levene's test for equality of variances. There was no statistically significant correlation between gender and teacher comments, r_{pb} (44 = .148, p = .325), with Male students having more biased words used than Female students (.088 ± .038 versus .078 ± .032). Gender accounted for 2% of the variability in teacher comments (see Table 5).

Table 5

Point-biserial Pearson Correlation between Gender Groups and Teacher Comments

	Male & Female	
	r	*p*
Teacher Comments	.148	.325

In summary, a Mann-Whitney U test was used to determine whether there was a statistically significant relationship between teachers' assessment and the gender of the child in the DRDP and TSG early childhood assessment. The gender results showed for mathematics and science that male and female scores were not statistically significantly different. To determine whether there was a statistical association between teachers' comments and the gender of the child, a point-biserial Pearson correlation was run between gender and teacher comments. The results showed no statistically significant correlation between gender and teacher comments.

Summary

The purpose of this study was to analyze teacher assessment of children's mathematics and science skills on the DRDP and TSG, along with teacher comments written in preparation for each child's parent-teacher conference, to determine if there is a relationship between teachers' assessment and comments and the race and gender of the child. The results showed for mathematics that Black, Latino, and Other Ethnicity student

scores were statistically significantly higher than for White students. For science, pairs Black and Latino were statistically significantly different while student score pairs Black and Other Ethnicity, Latino and Other Ethnicity, White and Black, White and Latino, and White and Other Ethnicity were not statistically significantly different. Also, there was no statistically significant correlation between race and teacher comments in each of the dichotomous race groups (Black & White, Black & Latino, and Latino & White). The results showed for mathematics and science that male and female scores were not statistically significantly different. Also, there was no statistically significant correlation between gender and teacher comments.

In this chapter, I presented the research questions, data collection process, results, and data analysis of my study. In the final chapter, the interpretation of findings, limitations of study, recommendation for future research, implications concerning positive social change, and the conclusion will be presented.

Chapter 5: Discussion, Conclusions, and Recommendations

The purpose of this study was to analyze teacher assessment of children's mathematics and science skills on the DRDP and TSG, along with teacher comments written in preparation for each child's parent-teacher conference, to determine if there is a relationship between teachers' assessment and comments and the race and gender of the child. This study utilized a Mann-Whitney U nonparametric test and a point-biserial Pearson correlation parametric test to determine the differences and strength of the relationship between data from preschool students' DRDP or TSG assessment, and teacher comments written in preparation for each child's parent-teacher conference. Presented in this chapter are the conclusions resulting from this study and the recommendations for further research.

Interpretation of the Findings

Interpretation of RQ1

RQ 1 asked, "Is there a statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's race?" Both the Mann-Whitney U nonparametric test and a point-biserial variation of the Pearson correlation parametric test were used to address this question. The results indicated from the Mann-Whitney U for mathematics that Black, Latino, and Other Ethnicity student scores were statistically significantly higher than for White students. However, for pair-wise comparisons of children identified as Black and Latino, Black and Other Ethnicity, and

Latino and Other Ethnicity, students' mathematics scores were not statistically significantly different.

For Black, Latino, and Other Ethnicity students, mathematic scores were statistically significantly higher than for White students; this result may be because of racial matching of students and teachers. The data came from two federally funded Head Start programs. Head Start is a low-income based child care program that provides education, health, nutrition, and parent involvement services to children ages birth to 5. Unlike elementary and secondary education where White teachers are predominant, staff in early education programs such as Head Start are more diverse (Whitebook, McLean, & Edwards, 2018). The students in Head Start programs are also diverse. As of 2016-2017 the student racial composition for Head Start was 42 % Black, 38 % Latino, and 25% White children (Child Trends Databank, 2015). Researchers have shown that racial matching has an influence on a teacher's perception of a student (Yarnell & Bohrnstedt, 2018). Wright, Gottfried, and Vi-Nhuan (2017) used the most recent nationally representative data and found that kindergarten students whose teacher's race was the same as their own had more favorable rating of eternalizing behaviors than they did when there was a difference in race between a student and a teacher. Grissom, Rodriguez, and Kern (2017) found that schools with larger number of Black and Latino teachers had a higher number of Black and Latino students enrolled in gifted programs.

For the Latino students, science scores were found to be statistically significantly higher than for Black students. One possible explanation for these results is the fact that Black students nationally perform at significantly lower levels in STEM areas of science,

technology, engineering and math (Camera, 2015). In recent years researchers found disparities in science achievement across race in which Latino students' scored higher than Black students in the science area of the Early Childhood Longitudinal Study, Kindergarten (ECLS-K) (Curran & Kellogg, 2016). Supporting the findings, Morgan, Farkes, Hillemeier, and Maczuga (2016) discovered disparities in science that showed children who are Black scored .62 standard deviations lower than children who were White, while Latino children scored .29 of a standard deviation lower than children who are White.

The results indicated that there was no statistically significant correlation between race and teacher comments in each of the dichotomous race groups (Black & White, Black & Latino and Latino & White). There are many factors besides a student's race that may influence how the teacher perceives students (Ali, Khan, & Hussain, 2018). Several teacher comments associated excellent students with those students' affinity for being in large groups, however, other teachers associated children who preferred to work alone and were quiet in large groups with being good students. Teachers who wrote the conference comments may have been drawn to students whose personalities were like their own. Rausch et al. (2016) found that students whose personality is similar to their teachers are judged more positively than students who are dissimilar, even when students' test performance is controlled .Gehlbach et al. (2016) found that teachers who received feedback about their similarities with specific students perceived better relationships with those students, and those students earned higher course grades.

A deeper understanding of these findings comes from Wason's theory of confirmation bias. Confirmation bias suggests that individuals do not perceive circumstances objectively, but rather pick out information that confirms their prejudices (Wason, 1968). This can possibly explain why for Black, Latino, and Other Ethnicity students' mathematic scores were statistically significantly higher than for White students, and why White students' scores were significantly lower. The scores on the assessments are based on teachers' observation of the students' behavior. While observing the child's behavior, the teacher documents all the instances that support her belief about a certain child and then records it on the assessment. Instead of basing her findings on what the child can actually do, she may have based her findings on what she believes a child can or cannot do. These beliefs may be influenced by affinities with race or personality type. However, it is also possible that White children objectively underperformed Black and Latino children, pointing to possible differences in children's experiences outside the classroom or in their prior experience.

Interpretation of RQ2

RQ 2 asked, "Is there a statistically significant relationship between preschool teachers' assessments and comments recorded on preschool assessments regarding children's mathematics and science skills and those children's gender?" Both the Mann-Whitney U nonparametric test and a point-biserial Pearson correlation parametric test were used to address this question. The results from the Mann-Whitney U indicated that for mathematics and science, male and female scores were not statistically significantly different, which supports the null hypothesis that there is no difference in math scores

between male and female students. Scheiber, Reynolds, Hajovsky, and Kaufman (2015) investigated developmental gender differences in academic achievement areas on the Kaufman Test of Educational Achievement - Second Edition, and concluded that a small but consistent advantage was identified for females in reading, but that there were no gender differences in mathematics. Chen, Yang, and Hsiao (2016) found that although females and males in secondary education showed different topic interest in mathematics courses such as pre-calculus, they performed equally well. Curran and Kellogg (2016) found no gender gap in science achievement during early education but that a gap begins to emerge as students move through the first few years of schooling. Similarly, Morgan et al. (2016) found that large science achievement gaps are first evident in third grade and are persistent to at least the eighth grade. The results for RQ2 confirm these prior findings from the literature.

The point-biserial Pearson correlation parametric test indicated that there also was no statistically significant correlation between gender and teacher comments. In the teacher comments there were several instances where the teacher used the word *amazing* to describe students who paid attention, listened, and actively participated. The word *fine* was used to describe students who were easily distracted, lacked participation, and had difficulty following directions. Research has shown that teachers perceive students more favorably when they display desirable behaviors (McGrath & Bergen, 2015). Cho (2016) found that students who demonstrate high achievement, good behaviors, and good communication skills were identified as a teacher's favorite students, but for these teachers and these preschool students, even children less-skilled than others in self-

regulation of behavior were recorded as *fine*. These distinctions in teacher comments were similar for all students, regardless of gender.

The gender of teachers who recorded comments used in this study is unknown, but given the large proportion of early childhood teachers who are female, I assume that most or all of the teachers were female. Confirmation bias suggests that female teachers might view female students more favorably than male students, or that a pre-existing gender bias towards male achievement might cause teachers to view male achievement more favorably than female achievement or excuse male misbehavior more than female misbehavior. Given that there were no significant differences found by gender, any action of confirmation bias in children's assessments is unknown.

Limitations of the Study

This study was limited to only two federally funded Head Start programs that provided preschool services for low-income families with young children, so the results cannot be generalizable to the entire population of early childhood programs. Second, the number of participants by race was not evenly distributed because enrollment in the selected programs included a higher population of Latino students compared to all other racial groups. In addition, the science component of the DRDP assessment was not present because the cooperating organization used a form of the DRDP that does not include a science component. By not having as many science scores as math scores, the results found for science are less reliable than are the results for mathematics. On the TSG assessment, several of the student assessments had missing data and so a multiple imputation method was used in place of missing data, which may have affected the

results. Finally, the data from parent-teacher conference comments were limited because the first organization did not allow outsiders access to the conference comments and the second organization allowed parents to opt out of participating in conferences. By not having a larger number of conference comments to analyze, the relationship between teacher comments and the students' race and gender remains unclear.

Recommendations

Although, there were no difference found for minority students and students of a specific gender, White students' mathematics scores were found to be statistically significantly lower than the mathematics scores of Black, Latino, and Other Ethnicity students. Since differences were found in math assessment scores for White students, recommendations for further research can come from this study. The first recommendation is to investigate why White students' mathematics scores were statistically significantly lower than the mathematics scores of Black, Latino, and Other Ethnicity students. Since research has shown that mathematics scores are often higher for White students than for minority students (Lawler, 2016; Sonnenschein, & Sun, 2017), the finding of lower scores for White students in this study is puzzling. Investigating this occurrence could help determine if this difference is a reflection of teacher bias, or if there are important barriers to these students' success that could be addressed in preschool.

Another recommendation for further research is an examination of the language domain of the DRDP and TSG. Recent studies have discovered differences for student factors of race and gender within the language domain (Brey et al., 2019; Reilly,

Neumann, & Andrews, 2018). The last recommendation for further research is to replicate the study using centers that are not part of the federally funded Head Start program, because it is possible that teachers in other programs might be trained differently in conducting objective assessments.

Implications

In this study, there were no significantly significant differences found for minority students' assessment scores in math and science. Also, there were no differences found between boys' and girls' assessment scores in math and science. Since, there were no differences found in math and science assessment scores for certain student characteristics, several recommendations for practice can be suggested for this study. The first recommendation is for directors to continue giving support and guidance to their teachers so that they may continue to assess students objectively. The second recommendation arising from this finding is to continue providing early education teachers with anti-bias resources so that they can continue to avoid minority and gender differences that are a result of biased assessment practices. A final recommendation is for federally funded Head Start programs to continue assessing children using valid and reliable observational assessment tools such as the DRDP and TSG, since these assessment tools have shown no minority and gender differences in assessment practices.

The findings from this study support teachers' use of observation-based assessment tools, because this study found no differences in assessment outcomes by race or gender. The results from this study support the contention that observational assessment tools such as the DRDP and TSG are valid and reliable, and that minority

students and students of a specific gender are accurately accessed by such assessment tools. This study established implications for positive social change by supporting teachers' continued use of observational assessments to access students' abilities. Observational assessments are appropriate for preschool children, because they are unobtrusive and based on children's demonstrated abilities in natural contexts. Continued use of such assessments over more intrusive and decontextualized standardized assessments may contribute to children's learning success, by providing to their teachers and their parents an authentic perspective on children's abilities and challenges.

Conclusion

The purpose of this study was to analyze teacher assessment of children's mathematics and science skills on the DRDP and TSG, along with teacher comments written in preparation for each child's parent-teacher conference, to determine if there is a relationship between teachers' assessment and comments and the race and gender of the child. A Mann-Whitney U test was conducted for race and gender and it determined that the mathematics scores were statistically significant higher for Blacks, Latinos, and Other Ethnicity students, compared to White students. However, for science scores, scores of Latino students were statistically significantly higher than scores for Black students. For gender, no difference was found in mathematics or science assessment scores. A point-biserial Pearson correlation parametric test was conducted on teacher comments and determined that there was no statistically significant correlation between race or gender and teacher comments. These results indicate that no differences in assessment practices for minority students or for students of a specific gender were inherent in observational

assessment of preschool children in Head Start centers in two regions of the United States. Based on these findings, early educators may continue to assess student abilities using observational assessment tools, confident in the fairness of the outcomes. Teachers are empowered to use their skills in observing children, and in rating and commenting on children's abilities and challenges, to help young children, regardless of race and gender, to develop their abilities in mathematics and science.

References

Ali, A., Khan, D. M., & Hussain, M. (2018). Causes of teacher's favoritism and its effects on the university students: A case study. *Global Social Sciences Review, 3*(2), 371-386. doi:10.31703/gssr.2018(III-II).21

Anyon, Y., Jenson, J., Altschul, I., Farrar, J., McQueen, J., Greer, E., & Simmons, J. (2014). The persistent effect of race and the promise of alternatives to suspension in school discipline outcomes. *Children and Youth Services Review*, *44*, 379–386. doi:10.1016/j.childyouth.2014.06.025

Baker, C. N., Tichovolsky, M. H., Kupersmidt, J. B., Voegler-Lee, M. E., & Arnold, D. H. (2015). Teacher (mis) perceptions of preschoolers' academic skills: Predictors and associations with longitudinal outcomes. *Journal of Educational Psychology*, *107*(3), 805-820. doi:10.1037/edu0000008

Banerjee, R., & Luckner, J. L. (2013). Assessment practices and training needs of early childhood professionals. *Journal of Early Childhood Teacher Education, 34*(3), 231-248. doi:10.1080/10901027.2013.816808

Banjong, D. N. (2014). Same performance but different perception: Female stereotypes in mathematics emerge in fifth grade. *International Online Journal of Educational Sciences*, *6*(2), 258-268. doi:10.15345/iojes.2014.02.001

Baudson, G.T, Fischbach, A & Preckel, F. (2016). Teacher judgments as measures of children's cognitive: A multilevel analysis*, Learning and Individual Relationship, 52,* 148-156. doi:10.1016/j.lindif.2014.06.001

Benaji, R.M., & Greenwald, G.A. (2013). *Blindspot: Hidden Biases of Good People.* New York City, NY: Bantam Books.

Black, P. (2015). Formative assessment–an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice*, *22*(1), 161-177. doi:10.1080/0969594X.2014.999643

Blanchard, S., & Muller, C. (2015). Gatekeepers of the American dream: How teachers' perceptions shape the academic outcomes of immigrant and language-minority students. *Social science Research, 51*, 262-275. doi:10.1016/j.ssresearch.2014.10.003

Bracken, B. A. (2004). *The psychoeducational assessment of preschool children*. Mahwah, NJ: Lawrence Erlbaum Associates.

Brawley S., Stormont M. (2014). Investigating reported data practices in early childhood: An exploratory study. *Journal of Positive Behavior Interventions, 16*(2), 102-11. doi:10.1177/1098300713480838

Brey, D.C., Musu, L., McFarland, J., Wilkinson-Flicker, S., Diliberti, M., Zhang, A., Branstetter, C., & Wang, X. (2019). Status and trends in the education of racial and ethnic groups 2018. Retrieved from https://nces.ed.gov/pubs2019/2019038.pdf

Burgess, S., & Greaves, E. (2013). Test scores, subjective assessment, and stereotyping of ethnic minorities. *Journal of Labor Economics*, *31*(3), 535-576. doi:10.1086/669340

Burts, D. C., & Kim, D. H. (2014). The TSG assessment system: measurement properties and use. *NHSA Dialog*, *17*(3), 122-135. Retrieved from https://journals.uncc.edu/dialog/article/viewFile/170/290

Burtt, A.E. (1939). *The English Philosophers from Bacon to Mill*. New York: The Modern Library. Random House.

California Department of Education. (2016). Introduction to Desired Results. Retrieved from http://www.cde.ca.gov/sp/cd/ci/desiredresults.asp

California Department of Education. (2013). California Desired Results Developmental Profile© (2010) Technical Report. Retrieved from https://www.desiredresults.us/sites/default/files/docs/resources/DRDP%C2%A9%20%282010%29%20Technical%20ReportAug15_13%281%29.pdf

Camera, L. (2015). African American students lagging far behind. Retrieved from http://www.madeirapr.com/uploads/2/6/9/8/26989365/us_news_and_world_report.pdf

Campbell, T. (2015). Stereotyped at seven? Biases in teacher judgement of pupils' ability and attainment. *Journal of Social Policy, 44*(3), 517-547. doi:10.1017/S0047279415000227

Casbergue, R. M., Bedford, A. W. & Burstein, K. (2014). CLASS reliability training as professional development for preschool teachers. *Journal of Research in Childhood Education 28*(4), 426-440. doi:10.1080/02568543.2014.944724

Chen, S. C., Yang, S. J., & Hsiao, C. C. (2016). Exploring student perceptions, learning outcome and gender differences in a flipped mathematics course. *British Journal of Educational Technology*, *47*(6), 1096-1112. doi:10.1111/bjet.12278

Child Trends Databank. (2015). *Head Start*. Retrieved from https://www.childtrends.org/indicators/head-start

Cho, S. (2016). Domain-specific identification of talented students in the Republic of Korea. *Turkish Journal of Giftedness & Education*, *6*(1) 24-34. Retrieved from http://www.tuzed.org/publications/cilt6/tuzed_6_1/tuzed_2016_6_1/tuzed_6_1_cho.pdf

Cornwell, C., Mustard, D. B., & Van Parys, J. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources*, *48*(1), 236-264. doi:10.3368/jhr.48.1.236

Curran, F. C., & Kellogg, A. T. (2016). Understanding science achievement gaps by race/ethnicity and gender in kindergarten and first grade. *Educational Researcher, 45*(5), 273-282. doi:10.3102/0013189X16656611

Egalite, J.A., Kisida, B., & A.M. Winters. (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

Falch, T., & Naper, L.R. (2013). Educational evaluation schemes and gender gaps in student achievement. *Economics of Education Review, 36*, 12-25. doi:10.1016/j.econedurev.2013.05.002

Faulkner, V. N., Stiff, L. V., Marshall, P. L., Nietfeld, J., & Crossland, C. L. (2014). Race and teacher evaluations as predictors of algebra placement. *Journal for Research in Mathematics Education*, *45*(3), 288-311. doi:10.5951/jresemathematicseduc.45.3.0288

Finn, J. D., & Servoss, T. J. (2014). Misbehavior, suspensions, and security measures in high school: Racial/ethnic and gender relationships. *Journal of Applied Research on Children: Informing Policy for Children at Risk*, *5*(2), 1-50. Retrieved from http://digitalcommons.library.tmc.edu/cgi/viewcontent.cgi?article=1211&context=childrenatrisk

Fischbach, A., Baudson, T. G., Preckel, F., Martin, R., & Brunner, M. (2013). Do teacher judgments of student intelligence predict life outcomes? *Learning and Individual Relationships, 27*, 109-119.doi:10.1016/j.lindif.2013.07.004

Fortin, N. M., Oreopoulos, P., & Phipps, S. (2015). Leaving boys behind gender disparities in high academic achievement. *Journal of Human Resources, 50*(3), 549-579. doi:10.3386/w19331

Gehlbach, H., Brinkworth, M. E., King, A. M., Hsu, L. M., McIntyre, J., & Rogers, T. (2016). Creating birds of similar feathers: Leveraging similarity to improve teacher-student relationships and academic achievement. J*ournal of Educational Psychology, 108*(3), 342-352.

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review*, *52,* 209-224. doi:10.17848/wp15-231

Gilliam, W. S., & Reyes, C. R. (2018). Teacher decision factors that lead to preschool expulsion: scale development and preliminary validation of the preschool expulsion risk measure. *Infants & Young Children*, *31*(2), 93-108. doi:10.1097/IYC.0000000000000113

Gilliam, S.W., Maupin., N.A., Reyes, R.C., Accavitti, M., & Shic, F. (2016). Do early educators' implicit biases regarding sex and race relate to behavior expectations and recommendations of preschool expulsions and suspensions? *Yale University Child Study Center*, 1-18. Retrieved from http://ziglercenter.yale.edu/publications/Preschool%20Implicit%20Bias%20Policy%20Brief_final_9_26_276766_5379.pdf

Gilliam, W. S. (2005). *Prekindergarteners left behind: Expulsion rates in state prekindergarten systems*. New York, NY: Foundation for Child Development.

Glock, S., Krolak-Schwerdt, & Cate, M.P.I. (2015). Are school placement recommendation accurate? The effect of students' ethnicity on teachers' judgments and recognition memory. *European Journal Of Psychology of Education, 30*(2), 169-188. doi:10.1007/s10212-014-0237-2

Goldstein, J., & Flake, K.J. (2016). Towards a framework for the validation of early childhood assessments. *Educational Assessment, Evaluation And Accountability*, 28(3), 273-293. doi:10.1007/s11092-015-9231-8

Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal*, *115*(4), 523-5. doi:10.1086/681947

Gregory, A., Hafen, C. A., Ruzek, E., Mikami, A. Y., Allen, J. P., & Pianta, R. C. (2016). Closing the racial discipline gap in classrooms by changing teacher practice. *School Psychology Review*, *45*(2), 171-191. doi:10.17105/SPR45-2.171-191

Grissom, J. A., Rodriguez, L. A., & Kern, E. C. (2017). Teacher and principal diversity and the representation of students of color in gifted programs: Evidence from national data. *The Elementary School Journal*, *117*(3), 396-422. doi:10.1086/690274

Guide to Assessment in Early Childhood; Infancy to Age Eight. (2008). Retrieved from http://www.k12.wa.us/EarlyLearning/pubdocs/assessment_print.pdf

Hansen, K. (2016). The relationship between teacher perceptions of pupil attractiveness and academic ability. *British Educational Research Journal, 42*(3), 376–98. doi:10.1002/ berj.3227

Ishimine, K., & Tayler, C. (2014). Assessing quality in early childhood education and care. *European Journal of Education, 49*(2), 272-290.doi:10.1111/ejed.12043

Jacoby-Senghor, D. S., Sinclair, S., & Shelton, J. N. (2016). A lesson in bias: The relationship between implicit racial bias and performance in pedagogical contexts. *Journal of Experimental Social Psychology*, *63*, 50-55. doi:10.1016/j.jesp.2015.10.010

Kaiser, J., Retelsdorf, J., Südkamp, A., & Möller, J. (2013). Achievement and engagement: How student characteristics influence teacher judgments. *Learning and Instruction*, *28,* 73-84. doi:10.1016/j.learninstruc.2013.06.001

Keys, T. D., Farkas, G., Burchinal, M. R., Duncan, G. J., Vandell, D. L., Li, W., & Howes, C. (2013). Preschool center quality and school readiness: Quality effects and variation by demographic and child characteristics. *Child Development*, *84*(4), 1171-1190. doi:10.1111/cdev.12048

Komarraju, M., & Nadler, D. (2013). Self-efficacy and academic achievement: Why do implicit beliefs, goals, and effort regulation matter? *Learning and Individual Relationships*, *25*, 67-72. doi:10.1016/j.lindif.2013.01.005

Krkovic, K., Greiff, S., Kupiainen, S., Vainikainen, M.-P., & Hautamäki, J. (2014). Teacher evaluation of student ability: what roles do teacher gender, student gender, and their interaction play? *Educational Research, 56*, 243–256. doi:10.1080/00131881.2014.898909

Krolak-Schwerdt, S., Böhmer, M., & Gräsel, C. (2013). The impact of accountability on teachers' assessments of student performance: A social cognitive analysis. *Social Psychology of Education*, *16*(2), 215-239. doi:10.1007/s11218-013-9215-9

Kroth, R.L., & Edge, D. (2017). Parent-teacher conferences. *Focus On Exceptional Children, 40*(2), 1-8. Retrieved from http://journals.ku.edu/FOEC/article/download/6828/6182

Lambert, R. G., Kim, D. H., & Burts, D. C. (2015). The measurement properties of the TSG® assessment system. *Early Childhood Research Quarterly*, *33*, 49-63. doi:10.1016/j.ecresq.2015.05.004

Lawler, B. R. (2016). Persistent iniquities: A twenty-year perspective on "race, sex, socioeconomic status, and mathematics." *The Mathematics Educator, 1,* 29-46.

Retrieved from http://tme.journals.libs.uga.edu/index.php/tme/article/viewFile/329/275

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

Mayer, F.A.,Wiley,W.E., Wiley, P.L.,Dees, C.D., & Raiford, A.S.. (2016). Teacher and school characteristics: Predictors of student achievement in Georgia public schools. *Georgia Educational Researcher, 13*(1), 86- 112. doi:10.20429/ger.2016.130103

McCormick, M. P., & O'Connor, E. E. (2015). Teacher-child relationship quality and academic achievement in elementary school: Does gender matter? *Journal of Educational Psychology*, *107*(2), 502. doi:10.1037/a0037457

McGrady, P. B., & Reynolds, J. R. (2013). Racial mismatch in the classroom beyond black-white relationships. *Sociology of Education*, *86*(1), 3-17. doi:10.1177/0038040712444857

McGrath, K. F., & Van Bergen, P. (2015). Who, when, why, and to what end? Students at risk of negative student-teacher relationships and their outcomes. *Educational Research Review, 14,* 1- 17. doi:10.1016/j.edurev.2014.12.001

McLachlan, C., Fleer, M., & Edwards, S. (2018). *Early childhood curriculum: Planning, assessment and implementation*. Cambridge, England: Cambridge University Press.

Mega, C., Ronconi, L., & De Beni, R. (2014). What makes a good student? How emotions, self-regulated learning, and motivation contribute to academic achievement. *Journal of Educational Psychology*, *106*(1), 121-131. doi:10.1037/a0033546

Mertler, C.A. (2016). *Classroom assessment: A practical guide for educators*. New York, NY: Routledge.

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Science achievement gaps begin very early, persist, and are largely explained by modifiable factors. *Educational Researcher, 45*(1), 18-35. doi:10.3102/0013189X16633182

Navarro, R.J. (2019). *Unconscious Bias*. Retrieved from https://diversity.ucsf.edu/resources/unconscious-bias

O'Brennan, L. M., Bradshaw, C. P., & Furlong, M. J. (2014). Influence of classroom and school climate on teacher perceptions of student problem behavior. *School Mental Health*, *6*(2), 125-136. doi:10.1007/s12310-014-9118-8

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

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

Pas, E. T., & Bradshaw, C. P. (2014). What affects teacher ratings of student behaviors? The potential influence of teachers' perceptions of the school environment and experiences. *Prevention Science*, *15*(6), 940-950. doi:10.1007/s11121-013-0432-4

RaoSoft. (2004). *Sample size calculator.* Retrieved from http://www.raosoft.com/samplesize.html

Rausch, T., Karing, C., Dorfler, T., & Artelt, C. (2016). Personality similarity between teachers and their students influences teacher judgement of student achievement. *Educational Psychology, 36*(5), 863-878. doi:10.1080/01443410.2014.998629

Recasens, M., Danescu-Niculescu-Mizil, C., & Jurafsky, D. (2013). Linguistic models for analyzing and detecting biased language. *The Association for Computer Linguistics, 1,* 1650-1659. Retrieved from http://www.aclweb.org/anthology/P13-1162

Reilly, D., Neumann, D. L., & Andrews, G. (2018). Gender differences in reading and writing achievement: Evidence from the national assessment of educational progress (NAEP). *American Psychologist*, 1-21 doi:10.1037/amp0000356

Retelsdorf, J., Schwartz, K., and Asbrock, F. (2015). "Michael can't read!" Teachers' gender stereotypes and boys' reading self-concept. *Journal of Education Psychology, 107*, 186–194. doi:10.1037/a0037107

Rimm-Kaufman, S. E., Baroody, A. E., Larsen, R. A., Curby, T. W., & Abry, T. (2014). To what extent do teacher–student interaction quality and student gender contribute to fifth graders' engagement in mathematics learning? *Journal of Educational Psychology, 107*(1), 170. doi:10.1037/a0037252

Robinson-Cimpian, J. P., Lubienski, S. T., Ganley, C. M., & Copur-Gencturk, Y. (2014a). Teachers' perceptions of students' mathematics proficiency may

exacerbate early gender gaps in achievement. *Developmental Psychology, 50*(4), 1262-1281. doi:10.1037/a0035073

Robinson-Cimpian, J. P., Lubienski, S. T., Ganley, C. M., & Copur-Gencturk, Y. (2014b). Are schools shortchanging boys or girls? The answer rests on methods and assumptions: Reply to Card (2014) and Penner (2014). *Developmental Psychology, 50*(6), 1840 –1844. doi:10.1037/a0036693

Rubie-Davies, C. M., Peterson, E. R., Sibley, C. G., & Rosenthal, R. (2015). A teacher expectation intervention: Modelling the practices of high expectation teachers. *Contemporary Educational Psychology*, *40*, 72-85. doi:10.1016/j.cedpsych.2014.03.003

Scheiber, C., Reynolds, M. R., Hajovsky, D. B., & Kaufman, A. S. (2015). Gender differences in achievement in a large, nationally representative sample of children and adolescents. *Psychology in the Schools*, *52*(4), 335-348. doi:10.1002/pits.21827

Soleimani, H., & Rahmanian, M. (2014). Self-, peer-, and teacher-assessments in writing improvement: A study of complexity, accuracy, and fluency. *Research in Applied Linguistics*, *5*(2), 128-148. Retrieved from http://rals.scu.ac.ir/article_11016_994f2efb124ba6e4ccbddf3a1b517656.pdf

Sonnenschein, S., & Sun, S. (2017). Racial/ethnic differences in kindergartners' reading and math skills: Parents' knowledge of children's development and home□based activities as mediators. *Infant and Child Development*, *26*(5), 1-21. doi:10.1002/icd.2010

Sprietsma, M. (2013). Discrimination in grading: Experimental evidence from primary school teachers. *Empirical Economics*, *45*(1), 523-538. doi:10.1007/s00181-012-0609-x

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*(3), 743.doi:10.1037/a0027627

Sutter, C., Ontai, L. L., Nishina, A., Conger, K. J., Shilts, M. K., & Townsend, M. S. (2017). Utilizing the desired results developmental profile as a measure of school readiness: evaluating factor structure and predictors of school readiness. *Early Child Development and Care*, *187*(9), 1433-1445. doi:10.1080/03004430.2016.1174933

Thelen, E., & Adolph, K.E. (1992). Arnold L Gesell: The paradox of nature and nurture. *Developmental Psychology, 28*, (3), 357-387. doi:10.1037/0012-1649.28.3.368

Upadyaya, K., & Eccles, J.S. (2015). Do teachers' perceptions of children's mathematics and reading related ability and effort predict children's self-concept of ability in mathematics and reading? *Educational Psychology*, *35*(1), 110-127. doi:10.1080/01443410.2014.915927

Urhahne, D. (2015). Teacher behavior as a mediator of the relationship between teacher judgment and students' motivation and emotion. *Teaching and Teacher Education, 45*, 73-82. doi:10.1016/j.tate.2014.09.006

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 20*, 273-281. doi:10.1080/14640746808400161

Wason, P. C., & Johnson, P. N. (1970). A conflict between selecting and evaluating information in an inferential task. *British Journal of Psychology, 61(*4), 509-515. doi:10.1111/j.2044-8295.1970.tb01270.x

Wason, P.C., & Brooks, P.G. (1979). THOG: The anatomy of a problem. *Psychological Research, 41*(1), 79-90. doi:10.1007/BF00309425

Weisman, B (2012). Student motivation and the alignment of teacher beliefs. *The Clearing House, 85,* 102-108. doi:10.1080/00098655.2011.

Whitebook, M., McLean, C., Austin, L. J., & Edwards, B. (2018). Early Childhood Workforce Index 2018. *Center for the Study of Child Care Employment, University of California at Berkeley*.

Wolter, I., Braun, E., & Hannover, B. (2015). Reading is for girls!? The negative impact of preschool teachers' traditional gender role attitudes on boys' reading related motivation and skills. *Frontiers in Psychology*, 1267 (6), 1-11. doi:10.3389/fpsyg.2015.01267

Wright, A., Gottfried, M. A., & Le, V. N. (2017). A kindergarten teacher like me: The role of student-teacher race in social-emotional development. *American Educational Research Journal*, *54*(1), 78S-101S. doi:10.3102%2F0002831216635733

Yarnell, L. M., & Bohrnstedt, G. W. (2018). Student-teacher racial match and its association with black student achievement: An exploration using multilevel structural equation modeling. *American Educational Research Journal*, *55*(2), 287-324. doi:10.3102%2F0002831217734804

Appendix A: Detecting Biased Language Training Sheet

1. Count the number of propositions that appear to be *assumed to be true or false*. Usually these are indicated by verbs that imply a relationship in the meaning of an embedded presupposition.

Cathy knew Josh could count to 10
compared to
Cathy didn't know Josh could count to 10

2. Count the number of times in which the verb implies a bias or judgment.

the teacher was coerced into accepting the bribe
compared to
the teacher accepted the bribe

3. Count the number of times in which *an adjective or adverb implies bias or judgment* that is otherwise unstated.

she clearly stated that Susan is a good reader
compared to
she stated that Susan is a good reader

4. Count the number of times in which *a conditional verb* suggests alternatives to the observed event, reducing the speaker's commitment to the truth of a proposition.

David may have taken Billy's toy car
compared to
David took Billy's toy car.

5. Count the number of times that *one side of an event is presented* as if it were the only perspective or opinion or as if the observer did not personally observe the event.

Ms. Smith said that David was given the choice to sit at circle time
compared to
David was given the choice to sit at circle time.

I **Biased Elements Quick List**

1. Surprise or lack of surprise is evident and implies bias
2. The verb implies bias
3. An adjective or adverb implies bias
4. Use of a conditional verb implies lack of commitment to the statement's truth
5. A statement is attributed to someone other than the writer

Example Paragraph

At circle time I noticed(1) she was forced (2) to sit next to a boy the teacher knew (5) she did not like. She was often (3) very (3) verbal about how she did not want to sit next to him, but the teacher told her to sit next to him. She did as she was told and then started yelling (2) at the boy saying, "I do not like you and my mom knows it."

Biased language found= 6
Total words used in statement= 76
Level of bias detected = 6/76= 0.07