AN ANALYSIS OF MULTIPLE MEASURES TO REDUCE DISPROPORTIONALITY IN ACCELERATED PROGRAM IDENTIFICATION

BY

JON STADLER

A dissertation submitted to City University of Seattle

in partial fulfillment of the requirements for the degree of

DOCTOR OF EDUCATION

2021

SIGNATURE PAGE

This dissertation has been examined and approved.

[signature]	April 12, 2021
Craig Schieber, Ed.D., Committee Chair	Date
Michele C. Weber	April 13, 2021
Michele Weber, Ph.D., Committee Member	Date
Eunjeong Choi	04-13-2021
Eunjeong Choi, Ph.D., Committee Member	Date
Vicki Butler, EdD	4/14/2021
Vicki Butler, Ed.D., Dean	Date

DEDICATION

This paper is dedicated to everyone who never thought they could, those who did not know they should, and those who wondered if they ever would. Yes, you can.

ACKNOWLEDGMENTS

There are many people that have helped me through this journey, and I could not have accomplished this without them. First, thank you to my wife and partner, Carey. Your encouragement, support, and patience never stopped. You were there every step of the way and I cannot imagine having done this without you. Taryn and Cameron, your willingness to let me work whenever I needed to work and your ability to celebrate the small steps always brought me encouragement. I hope you always remember to work hard and never give up because you can accomplish anything! Alex, thank you for your continuous check-ins and encouragement.

To my Mom and Dad, I know you have always believed in me. Thank you for never wavering in that belief in anything I have done. Amy, you have always been my biggest supporter and never wanted anything but the best for me, thank you. Darrin, the many miles we have spent talking about school and giving one another perspective has helped make getting to this point possible.

To my colleagues who I work with, I have appreciated all your encouragement, guidance, and willingness to let me talk whenever I needed to about this journey. David, your willingness to help me think through problems, proofread papers, and desire to push me when I needed it, you have no idea how much it has meant to me; I could not have done this without you. Damien, thank you for getting me started. I never thought completing this degree was for me, but you made sure I knew that I could do it and you believed in me. Thank you all!

Finally, to the City University staff, thank you for all the support you have provided. Dr. Ferry, you were available and knowledgeable with each and every question.

You knew when I needed encouragement and when to push; thank you. Dr. Schieber, I could not have asked for a better chair. You were always patient and willing to talk with me at any time. Thank you for all your efforts.

TABLE OF CONTENTS

DEDICATION iv

ACKNOWLEDGMENTS v

LIST OF TABLES x

LIST OF FIGURES xi

ABSTRACT xii

CHAPTER 1: INTRODUCTION TO THE STUDY 1

Study Background/Foundation 2

Current State of the Field in which the Problem Exists 4

Historical Background 6

Deficiencies in the Evidence 7

Problem Statement 8

Specific Leadership Problem 8

Audience 9

Purpose of the Study 10

Methodology and Research Design Overview 10

Research Questions and Hypotheses 11

Study Limitations 13

Study Delimitations 14

Definitions of Key Terms 14

Summary 16

CHAPTER 2: LITERATURE REVIEW 17

Accelerated Programs 18

High School Programs 19

Elementary Programs 21

Identification Processes for Accelerated Programs 22
Identification Methods 23
Multiple Measures 27
Disproportionate Representation in Accelerated Programs 28
Disproportionality: Roots and Effect 28
Bias in Individuals 30
Theoretical Foundation 31
Cultural Mobility Framework 32
Three-Stratum Theory 33
Educational Test Theory 34
Summary 36
CHAPTER 3: METHODOLOGY 38
Research Method 38
Research Design...40
Instruments...41
Naglieri Nonverbal Ability Test...41
Fountas & Pinnell Benchmark Assessment System...43
STAR Math 45
Independent Variables 46
Participants...47
Data Analysis Methods 48
Structural Equation Modeling 49
Research Questions 52
Limitations...57
Delimitations..58

Summary 59
CHAPTER 4: RESULTS 60
Presentation of Results 61
Demographic Statistics 62
Assessment Information 64
Research Question Statistics 65
Summary 85
CHAPTER 5: CONCLUSIONS AND DISCUSSION 87
Discussion of Results and Conclusions 88
Research Question 1 Conclusions 89
Research Question 2 Conclusions 91
Research Question 3 Conclusions 92
Application of Results and Conclusions to the Problem Statement 96
Application to Leadership 98
Recommendations for Action 100
Recommendations for Further Research 101
Concluding Statement 103
REFERENCES 104
APPENDIX Path Analysis Graphic Model 118

LIST OF TABLES

Table 1. Racial Demographics of Second Grade Cohort and Overall District Population in 2019....63

Table 2. Free and Reduced Lunch Count of Second Grade Cohort and Overall District Population in 2019....64

Table 3. Mean Assessment Score Data for Second Grade Cohort in 2018 and 2019....65

Table 4. BAS Results....66

Table 5. NNAT3 Results....68

Table 6. STAR Math Results....69

Table 7. Chi-Square Goodness of Fit....71

Table 8. Model Fit Summary....72

Table 9. Number of students of second grade cohort at or above the 90th percentile on BAS....75

Table 10. Chi-square test of BAS....77

Table 11. Number of students of second grade cohort in 90th percentile on NNAT3....79

Table 12. Chi-square test of NNAT3....81

Table 13. Number of students of second grade cohort in 90th percentile on STAR Math....83

Table 14. Chi-square test of STAR Math....85

LIST OF FIGURES

Figure 1. Multiple Factor Structural Equation Model for Research Question 1...............53

Figure 2. Multiple Factor Structural Equation Model for Research Question 2...............54

ABSTRACT

Low-income students and students of color are disproportionately identified at lower rates for accelerated programs in K-12 education compared to their peers in the United States. A lack of equitable identification processes that provide access to accelerated programs can be linked to issues of social inequity (Cao et al., 2017). An analysis of information from tools utilized in a Pacific northwest school district's identification process for accelerated programs was conducted. Cognitive assessment and academic measures used for universal screening processes at the second-grade level were examined to determine if they contributed to disproportionate outcomes. Information about accelerated programs in K-12 education, how students are identified for participation in these programs, and potential biases of the tools used for identification are provided in chapters one and two. Cultural mobility framework, three-stratum theory, and educational test theory were used to understand identification practices and tools that can lead to disproportionate outcomes in identification. Answering the research questions required testing the significance of between-group differences in assessment results for sub-groups of students, analyzing the relationships that exist between independent race and socioeconomic status variables and dependent assessment variables, and learning how identified between-group differences or bias could be counteracted to impact more equitable identification practices for accelerated programs. A quantitative methodology that applied a causal-comparative research design using structural equation modeling was able to identify the significance of between-group differences in assessments used for accelerated program identification in a K-12 school district. The population included close to the total population of second-grade students attending a Pacific northwest K-12 school district. The results of this study demonstrate that a statistical analysis can be used to identify unintentional bias in assessments and the identified differences can be used to mitigate the bias. School leaders may use these results to review identification policies and processes to more equitably identify low-income students and students of color for accelerated programs.

CHAPTER 1: INTRODUCTION TO THE STUDY

Disproportionate representation of student subgroups can be seen across K-12 education programs and outcomes, including discipline, graduation, state test results, special education, and accelerated programs. Typically, low-income students and students of color score lower on standardized tests and have lower levels of academic attainment than their peers (U.S. Department of Education, 2020). These students are excluded from school more frequently, over-identified for special education, and not identified for participation in accelerated programs. In the United States, compared to the larger student populations, low-income and students of color are frequently denied placement in K-12 programs that provide access to advanced or enriched learning (Hodges et al., 2018). The underrepresentation of low-income students and students of color in educational programs can be linked to issues of social inequity (Cao et al., 2017) and can have long lasting economic impacts due to reduced access to more advanced secondary and post-secondary learning opportunities, such as participation in college credit-bearing courses during or directly after high school (Crabtree et al., 2019).

Those who are identified for gifted services in K-12 education can be students who perform or show potential to academically perform at advanced levels. Low-income students and students of color have historically been identified for these services at lower rates compared to their peers. To identify more low-income students and students of color, school districts have implemented the use of locally normed measures into identification practices in place of using national norms and have implemented universal screening of students at one or more grade levels (Pfeiffer, 2015). For this study, an examination of the between-group differences for demographic sub-groups of students when using common assessment tools as variables was conducted to determine how the variance may impact identification of gifted student potential and access to accelerated programs.

Study Background/Foundation

Unequal participation rates in accelerated programs by student subgroups can be seen across the United States. In a study of gifted and talented identification practices, Hodges et al. (2018) noted that identification methods have failed to identify low-income students or students of color proportionally to other student demographics. The authors concluded that this underrepresentation is generally found across the United States. State laws related to gifted student identification vary greatly, which may contribute to the unequal participation by different sub-groups. Only 32 states have laws related to gifted education and not all of them have identification requirements. Gifted education laws in Washington state require elementary and secondary school districts to identify students for participation in their accelerated programs annually. The districts must use multiple criteria in the identification process (Washington State Legislature, RCW 28A.185.030, 2009).

Social inequities in a community can be seen and replicated in K-12 schools. Unequal opportunities for students can lead to unequal outcomes, regardless of the measure. School policies and practices can contribute to the continuation of inequities unless closely examined and corrected (Michelmore & Dynarski, 2017). To identify ways to enhance gifted education for under-represented groups, Ecker-Lyster and Niileksela (2017) posited that a paradigm shift toward embracing alternative identification methods could benefit the identification of gifted students. The authors suggested that researchers and districts could continue to explore equitable identification methods and measures including teacher training to identify and understand giftedness in students, and the use of local norms.

Research has been conducted to determine ways in which those involved in gifted education can identify more students from low-income families and students of color, including the impact of teacher training to identify gifted students. Steenbergen-Hu and Olszewski-Kubilius (2016) found that many states have shifted away from traditional identification methods, such as the use of a single IQ assessment, to identify gifted students. The authors also stated that teacher rating scales and teacher training in the identification of gifted student characteristics are two ways in which gifted education identification methods can increase the identification of historically disadvantaged students. As part of the shift from a single assessment, policies and practices have been implemented to support the use of multiple measures, including objective achievement measures as well as subjective perceptual measures, to identify students from historically underrepresented groups.

In an article about the roles of institutional and individual poverty in the identification of gifted students, Hamilton et al. (2018) found that school districts can demonstrate inequities in identification of gifted students between schools. The authors stated that low-income students are less likely to be identified for gifted programs but that school districts can interrupt these patterns through policy and practice and suggested two policy changes: using multiple measures to identify gifted students; and utilizing locally normed scores for gifted identification. By using locally normed cut scores on academic measures, school districts can identify the districts' gifted students in a way that more closely reflects the demographics within the district. Hamilton et al. (2018) found that by relying on national norms, such as national percentile rankings, school districts may limit their ability to identify gifted students in their district.

Current State of the Field in which the Problem Exists

There have been a number of studies examining gifted education in general; those researchers have established that identification practices are not equitable throughout the United States. McCluskey (2017) discussed the need to modify identification processes, observing that outdated identification practices are in place because stakeholders in school districts have not addressed the problem with creative solutions. To do so effectively, those who make policy and implement identification practices may benefit from alternative practices. These practices could provide more equitable ways of identifying groups of students who have been historically underrepresented in gifted education.

Studies about alternative identification practices exist but have not addressed specific practices or the tools that school districts use for identification. In an article about historically underrepresented sub-groups of students in gifted education, Lakin (2016) suggested that universal screening could address inequitable identification practices. Card and Giuliano (2016) also stated that universal screening could support more identification of low-income students and students of color. Universal screening ensures that all students are assessed at some point in their K-12 educational experience but does not ensure equitable identification of student sub-groups for gifted program participation.

Another recommended change in identification practices involves assisting teachers to better understand the characteristics of gifted students. Tirri (2017) suggested that teacher education could help teachers in identifying gifted students so they could refer them to gifted programs. The concept of teacher training in the identification of gifted students was also shared in an article by Hertzog (2017). Both authors suggested that teacher training could support the identification of students from historically underrepresented sub-groups of students and could benefit teachers in understanding how to best meet the needs of gifted students. Teacher professional learning on the topic of gifted education could be beneficial, but a design for implementation may need to be developed to meet the needs of school districts. Professional development opportunities would need to be ongoing as teacher turn-over could disrupt the proportion of those who received training.

The use of multiple measures or structures to identify students for gifted programs has also been recommended in studies. Cao et al. (2017) suggested that, historically, both cognitive assessments and academic measures have been used to identify students for gifted education. Ecker-Lyster and Niileksela (2017) noted that academic measures were a tool that could be used to identify students for gifted education. Multiple measures could be beneficial when schools are attempting to identify equitable numbers of students for gifted education, but school districts would need to establish the ways in which those measures are used for identification purposes.

Research in the area of gifted education identification has a common theme of inequitable identification of low-income students and students of color compared to their White and Asian peers. There are recommendations to limit identification inequities that include universal screening, teacher training, and the use of multiple measures for identification purposes. The implementation of these strategies has been successful to varying degrees. However, no clear evidence has been provided to show that these recommendations are working to increase the identification of low-income students and students of color for gifted programs.

Historical Background

Gifted and talented education has been active in the United States since the early 1900s and has gone through various iterations over the past century. During the 1920s gifted education was studied by psychologists, who determined that public schools were not meeting the needs of gifted students (National Association for Gifted Children, 2020). To meet these needs, education specifically for gifted children was started and has continued to evolve. Federal government legislative efforts expanded and supported gifted education through funding in the 1970s. In 1988, the Jacob K Javits Gifted and Talented Children and Youth Education Act (1988) was adopted. This act was the first federal legislation specific to gifted education funding in the United States.

Over the years, there have been numerous reports published about public education including A Nation at Risk (National Commission on Excellence in Education, 1983), National Excellence: A Case for Developing America's Talent (Ross & United States, 1993), and A Nation Deceived (Colangelo et al., 2004). The authors of these publications shared the idea that public education did not support students in the United States as effectively as it should. Through these reports, national standards for gifted education and the National Association for Gifted Children were established (National

Association for Gifted Children, 2020). However, in 2013, gifted and talented programs were defunded by the federal government due to a poor economy (Dougherty, 2014).

Identification of students who could access gifted education has traditionally relied on intelligence tests as students who were gifted were believed to be intellectually advanced, as compared to their peers. Sternberg (2017) suggested that the use of psychological intelligence testing is no longer appropriate to meet the needs of students who could demonstrate giftedness beyond their intelligence level. The author posited that students could be gifted in creativity, wisdom, and ethics; suggesting that identification methods for gifted education may need to change and evolve.

Deficiencies in the Evidence

While local school districts have explored multiple ways to identify underrepresented student populations, no specific or best practice has been identified (Hodges et al., 2018). Cao et al. (2017) suggested assessment data could be used in different ways to support gifted identification. One suggestion by the authors is to identify which assessments in use might limit disproportionate demographic representation of students in gifted programs. Another is to use assessment data to evaluate if gifted programs meet the needs of students.

Policy makers in the area of accelerated programs have recommended the use of multiple measures to identify and recommend students for program participation. However, there is limited information about what the measures should be or how they should be used to increase the participation of low-income and students of color. An exploration of potential bias in commonly used measures might be used to decrease disproportionality in the identification of low-income students and students of color for advanced programs and may benefit a Pacific northwest K-12 school district.

Problem Statement

The general problem is that the process through which students are identified for advanced education programs is not equitable. In the United States, low-income students and students of color are frequently denied placement in K-12 programs that provide access to advanced or enriched learning environments (e.g., Gifted and Talented, Advanced Placement, and International Baccalaureate) compared to the larger student populations (Hodges et al., 2018). The underrepresentation of low-income students and students of color in educational programs can be linked to issues of social inequity (Cao et al., 2017) and can have long lasting economic impacts due to reduced access to a range of secondary and post-secondary learning opportunities, such as participation in college credit-bearing courses in or directly after high school (Crabtree et al., 2019).

The measures used to identify students for accelerated programs rely on a narrow set of skills that, when used in isolation, fail to identify large numbers of students, specifically low-income and students of color, who could benefit from accelerated programs (Cao et al., 2017). Steenbergen-Hu and Olszewski-Kubilius (2016) stated that gifted education and accelerated programs can benefit students from all racial, ethnic, and socio-economic backgrounds. However, without equitable identification practices, students may be excluded from beneficial educational programs provided to their peers based on faulty assumptions or inaccurate evaluation tools including cognitive tests, academic assessments, prior academic performance, and teacher recommendations.

Specific Leadership Problem

The specific problem is that a K-12 school district in the state of Washington has failed to identify which commonly used measures that are considered fair and valid by district leaders may better inform accelerated program identification decisions and thus reduce disproportionate representation among low income students and students of color. As suggested by Hertzog (2017), the use of multiple measures of student abilities and skills, including academic assessments, can be analyzed to determine if multiple data points introduce unintended bias and if that bias can be compensated for to identify students for accelerated programs more equitably. An identification of how multiple measures may be used to counter-act implicit testing bias by analyzing the interaction between test results for low-income students and students of color can support school district leaders in adopting policies and procedures to support equitable identification of students for accelerated programs.

Audience

The information obtained from this study was utilized with the goal of testing the significance of between-group differences for historically underrepresented sub-groups of students compared to their peers when using locally administered assessment. School district leaders and policy makers in a K-12 school district may benefit from understanding the between-group differences when identifying student potential for accelerated programs. The data analyzed in this study may bring to light factors that contribute to bias implicit in assessment data used by school district leaders and policy makers to make program decisions. K-12 school district leaders and educators may be able to use the information from this study to inform decisions about the use of multiple measures for gifted identification for the assessments administered within their school district.

Purpose of the Study

The purpose of this quantitative casual-comparative study was to measure the between-group mean differences of assessment data for low-income students and students of color that may contribute to the disproportionate identification of second-grade students for accelerated programs in a K-12 school district in Washington state. Testing the significance of between-group differences for socioeconomic status and race and the assessment results used for student identification may help identify methods to address latent test bias. The independent variables used in this study for a structural equation model (SEM) are student race and socioeconomic status. Race was divided into the seven federal categories; determining applicable socioeconomic status factors focused on student eligibility for free or reduced-rate lunch. The dependent variables were cognitive and academic assessment data, including literacy and math, used by the school district to identify students enrolled in second grade for participation in accelerated programs.

Methodology and Research Design Overview

The study is an examination of how the use of cognitive and academic assessment tools in a K-12 school district may inequitably impact the identification of low-income students and students of color for accelerated program participation. In their meta-analysis of gifted education, Cao et al. (2017) found that a method for selecting which tools to use for gifted identification and how to use them together to utilize multiple methods has not been established. A quantitative methodology using assessments as dependent variables, as proposed in this study, may help reveal underlying factors associated with assessments that affect the results and could be used to create a formula or algorithm for the purpose of reducing disproportionality between subgroups of students in the selection of students for accelerated programs.

A quantitative research methodology with a large population size can support research questions focused on testing the significance of between-group differences of assessment results between low-income students and students of color compared to their peers (Watson, 2015). Specifically, a causal-comparative design was used to examine between-group differences and relationships in a K-12 school district through SEM. Causal-comparative designs can be used to test the significance of differences between groups and the relationships of the differences to the groups (Apuke, 2017). Lodico et al. (2010) stated that causal-comparative design may be beneficial in explaining differences between groups; the authors suggested that this design supports research in which the events affecting the variables have already occurred. A review of data from assessments that have already been administered can be considered causal-comparative or ex post facto. A causal-comparative design using SEM can measure the relationships between independent and dependent variables and identify the significance of the relationships between and amongst the variables (socioeconomic status, race, and assessment results) in a way that may alter identification outcomes in gifted education.

Research Questions and Hypotheses

The following research questions were used to address the problem of identifying assessment tools to inform accelerated program identification decisions that reduce disproportionate representation of low-income students and students of color.

1. What is the difference in assessment results for low-income students and students of color compared to their peers at a given grade level?

H_0 – There is no significant mean difference in assessment results for low-income students and students of color compared to their peers at a single grade level.

H_1 – There is a significant mean difference in assessment results for low-income students compared to their non-low-income peers at a single grade level.

H_2 – There is a significant mean difference in assessment results for students of color compared to their peers at a single grade level.

2. What is the significance of between-group differences when compounding dependent variables measured within and between independent variables?

H_0 – There is no significant mean difference in the relationships between dependent variables and independent variables when compounded.

H_1 – There is a significant mean difference in the relationships between dependent and independent variables when compounded.

3. Can assessment results used to identify students for accelerated programs be adjusted based on between-group differences to affect the identification of low-income students and students of color for accelerated programs?

H_0 – Adjustments based on between-group significant mean differences do not increase the identification of low-income students and students of color.

H_1 – Adjustments based on between-group significant mean differences increase the identification of low-income students for accelerated programs.

H_2 – Adjustments based on between-group significant mean differences increase the identification of students of color for accelerated programs.

There were five variables in this study: student low-income status (identified by Free and Reduced Lunch), student race (federal categories), standardized cognitive assessment (Naglieri Nonverbal Ability Test, 3rd Edition), teacher-administered reading comprehension test (Fountas & Pinnell Benchmark Achievement System), and a computer-adaptive mathematical skills test (Renaissance STAR Math).

Study Limitations

There are limitations to any study involving student demographic data and assessment scores. The entry of data into a student information system relies on an individual manually entering data or mass uploading data sets. Data entry errors which may impact the study can occur in these processes. Racial demographics are self-reported during student enrollment, so the race with which a family identifies could create a limitation. Families needing free and reduced-price lunch for their child or children are identified when a family member completes an application for federal funding. If the application is not completed and submitted, the number of those students may be under-reported, causing a potential limitation.

Data from a single grade level in the school district in a single year were reviewed in this study. Results may differ if different cohorts of students or data from different years were used in the data analysis.

Assessment data are also another limitation to the study. The school district providing data for the study utilizes standardized reading assessments at the second-grade level that rely on a degree of professional judgement to interpret a student's responses. Students participate in an online math assessment and a norm-referenced intelligence test, but the reading assessment is completed in a one-to-one setting with individual students and teachers, which allows for some subjectivity. While the scores can be normed, there are limitations in ensuring the assessment scores are equally reliable across the school district. The specific assessment tools used by the district and the results for grade two students may not apply to all school districts.

The limitation of the data accuracy can be mitigated by "cleaning" the data and comparing it to cohort data from years before second grade. A review of district trends over time can help in understanding if the data is accurate in terms of both demographics and academic performance.

Study Delimitations

The analysis of data used in this study was conducted using second grade student data in a K-12 school district for the 2019-2020 school year. The findings and results may not be applicable to other groups or grade levels because of the assessment data used, the composition of the student population, and the uniqueness of locally required assessments.

Definitions of Key Terms

The following terms and words provide clarification to allow for better understanding of the information in this study.

Accelerated programs – educational programs designed to create advanced academic opportunities for students. Accelerated programs vary depending on the age and grade of the student, and can include Advanced Placement (AP) courses, International Baccalaureate (IB) courses, or self-contained classes in which all students in the class are identified as gifted.

Bias – is a disproportionate weight inherent in a measure. It can be present in assessments that result in a student or group of students being unfairly penalized because of their personal characteristics. Teachers can also demonstrate conscious or unconscious bias when they treat students from different groups differently (Redding, 2019).

Computer adaptive assessment – is an assessment completed on a computer or digital device that adapts to an individual's level based on the correct or incorrect answers they provide to questions. Computer adaptive assessments can be used in literacy and math testing aimed at measuring a student's ability, as compared to similarly aged or grade-level peers.

Cultural capital – an individual's ability to utilize economic and social resources obtained from their family (Jæger, 2011). O'Shea (2016) suggested that cultural capital could be tangible goods or academic knowledge; students utilize that capital by taking advantage of their parents' direct knowledge or by passively accessing the education that parents with cultural capital know how to navigate.

Gifted education – students who perform at high levels in a given area such as, intellectually, creatively, artistically, or in a specific academic field (National Association for Gifted Children, 2020). These students may need services or activities to support their continued development.

Socioeconomic status – is the social class of an individual or group, often measured by education or income. Free- and reduced-price meal eligibility status in K-12 education was used to measure socioeconomic status, specifically to determine students who are considered low-income (Gajo et al., 2019). Students are either considered to be low-income or not low-income.

Standardized cognitive assessment – is an assessment administered under specified conditions that uses a common set of questions scored in a same way that measures an individual's cognitive ability compared to their peers. These assessments can provide information about a student's developmental strengths and areas of growth (Thompson et al., 2018).

Students of color – is a collective term used for non-White, non-Asian students. To receive federal funds, elementary and secondary schools must ask families to identify which of seven categories best fits their child's race: American Indian or Alaska Native, Asian, Black or African American, Hispanic, Native Hawaiian or Other Pacific Islander, White, and Two or More Races (U.S. Department of Education, 2008). The use of the phrase "students of color" i.e., identified by a category that is not White or Asian, and represents those who have historically been underrepresented in accelerated programs.

Teacher administered assessment – is an assessment given in a content area that may or may not be standardized and measures a student's skills in a subject. Teacher administered assessments can be beneficial in that they assist teachers in understanding the ways in which individual students perform on specific tasks with the goal of informing next steps in instruction (Bernstein et al, 2017).

Summary

This chapter provided background on gifted education and the inequitable identification of low-income students and students of color for gifted programs. The problem statement, purpose of the study, and research questions and hypothesis were also addressed.

CHAPTER 2: LITERATURE REVIEW

There have been decades of disparate outcomes for students from different races, ethnicities, and economic classes. Brought to light by the Coleman Report (Coleman, 1968) and continuing today, differences in educational outcomes for K-12 students across the United States persist and are influenced by the demographic composition of their schools (Reardon et al., 2019). Specific student sub-groups have disproportionate representation in areas like discipline, special education, and accelerated programs. Typically, more low-income students and students of color are excluded from school, over-identified for special education, and prevented from participating in accelerated programs, as compared to their peers. Locally made decisions about individual student behavior and academic potential can inhibit students' future academic outcomes and opportunities (Brown et al., 2020). The criteria used to identify students for accelerated programs in K-12 education are thought to be objective measures that demonstrate students' ability or potential to perform at advanced levels. However, the measures and processes used in accelerated program identification may contain unintended biases that result in low-income students and students of color being identified at lower rates compared to their peers. To counter these biases and identify more low-income and students of color, school districts have adopted policies and procedures to implement new practices that address the identification gap in accelerated programs (Lakin, 2016).

The purpose of this literature review is to provide information about accelerated programs and the means to identify students for these programs. A description of accelerated programs in K-12 school districts is provided, as well as the measures used in identifying students who can benefit from and should be allowed to access accelerated programs. The third section is a review of potential biases in the tools or processes used to identify students for accelerated programs and biases that lead to disproportionate representation. The final section of this literature review provides the theoretical foundation for the study.

Accelerated Programs

Accelerated programs are known by many names across the United States, including "academically advanced," "highly capable," and "gifted," with the latter being one of the more commonly used terms. Gifted education programs provide enrichment or classroom settings in which students can make continuous progress in school and support student strengths (National Association for Gifted Children, 2020). A survey conducted by the National Association for Gifted Children and the Council of State Directors of Programs for the Gifted (2015) found that only 32 of 50 states mandate identification and services for gifted students. The same survey found that 37 states define giftedness and only 27 states provide funding to support gifted students. Because there is no clear definition of giftedness, the range of what states and schools consider to be gifted can vary. The National Association for Gifted Children (2020) has said that gifted students are those who demonstrate outstanding levels of aptitude. To effectively meet the needs of those students, schools can and should implement programs and identify students who can benefit from these programs.

In the states that require accelerated programs, students can be identified at different grade levels and may be served in different ways over the course of their education. Generally, the goals of accelerated programs are to provide a selected group of students the opportunity to engage and be challenged to reach advanced levels of academic achievement (Hodges et al., 2018). The various processes and methods used to identify students for gifted or accelerated programs often include an adult's referral or a single assessment. Either or both methods can foster an intended or unintended bias that may exclude historically underrepresented students from accelerated programs.

In the state of Washington, school districts who serve students from kindergarten through twelfth grade are required to provide access to accelerated programs for students that have been identified as highly capable. School districts are responsible for the identification of students who can benefit from accelerated programs, as indicated by Washington State Legislature, RCW 28A.185.030 (2009). That mandate allows individual school districts to design their own programs and to identify students who access gifted education based on the goals of the program they create. The programs can also utilize a variety of strategies for providing instruction to these students, including ability grouping, acceleration of courses or grades, educational enrichment opportunities, and specialized schools or classes. Students can be identified early on and continue to receive access to accelerated programs throughout their K-12 experience. However, both identification processes and program offerings may inhibit enrollment of low-income students and students of color (Peters & Engerrand, 2016).

High School Programs

Accelerated programs in high school can challenge and support students in their learning. Depending on the age of the student, high schools can offer accelerated programs to meet the needs of students in the classroom or online. Frequently, high school programs center around providing students access to advanced coursework above the grade level standards to prepare them for post-secondary study. Students often take college-level courses such as Advanced Placement (AP) or International Baccalaureate (IB) in high school accelerated programs.

Advanced Placement courses provide students with an opportunity to be challenged, improve their college admission chances, be better prepared for college, and earn college credit (Study Point, 2020). IB courses offer students an opportunity to become culturally aware, engage with a rapidly changing world, and drive their own learning (International Baccalaureate, 2020). Accelerated programs for high school students may also include early entrance to college or universities, e.g., programs designed specifically for high school-aged students, such as those offered by the Robinson Center at the University of Washington and Bard College at Slippery Rock.

However, social inequities in the community can be replicated in high school accelerated programs, in which unequal opportunities for students can lead to unequal outcomes (Gordon & Bridglall, 2005). School policies and practices can contribute to the continuation of inequities unless further examined (Michelmore & Dynarski, 2017). As young people prepare to enter college, they can impact those who do as well as those who do not access accelerated programs by the end of their twelfth-grade education. In studies about accelerated program participation and demographic gaps, Cao et al. (2017) and Crabtree et al. (2019) found that issues of social inequality can have long-term effects. These researchers posited that students who do not have the opportunity to access accelerated programs may not have equal access to participation in college credit-bearing courses.

These limitations can contribute to social inequities for students and families beyond high school and have a negative economic impact on a student in college. A quantitative analysis of college completion for a cohort of high school graduates by Flores et al. (2017) suggested that there are differences between the race and income levels of students completing college. After a review of schools in the study, the authors noted that not all low-income students or students of color had access to the same accelerated programs to prepare them for college. Flores et al. (2017) highlighted potential impacts on students who do not access accelerated programs at some point during their educational experience, including lack of preparation for college courses and the high probability of living in poverty.

Elementary Programs

Elementary schools can offer students the opportunity to participate in more academically challenging programs through gifted and talented services. Students at that level can be identified as gifted and be provided services to support them through specially designed instruction (Steenbergen-Hu & Olszewski-Kubilius, 2016). Program design is determined by school districts and can include self-contained groups of students, a pull-out model, or small groupings of students. A school district can also choose for a program to have a specific focus, like science, technology, engineering, and math (STEM), language learning, art, or another core area. Through this flexibility, school districts can also design their identification process using measures they believe can identify students who may benefit from the specific program the district offers. Historically, school districts use intelligence tests for identification and acceptance into gifted programs, which has created the disproportionate representation of sub-groups of students in the programs.

The design of the accelerated program model at the elementary level can have impacts on which students participate and benefit from gifted education. For example, Steenbergen-Hu and Olszewski-Kubilius (2016) noted that a pull-out model that supports students less frequently may not meet all the needs of a single student due to inconsistent instruction and a failure to assist a student as they build their areas of strength. Another common program model is to have a self-contained classroom consisting only of students who have been identified as gifted. McCluskey (2017) indicates that a self-contained model can provide more consistent support to students who can benefit from gifted programs but is limited to students in the program. The author stated that students other than those identified could benefit if provided the opportunity. Different program models can meet the needs of gifted students in different ways. A school district may benefit from identifying their goals for gifted education to effectively meet the needs of students.

Identification Processes for Accelerated Programs

The exploration of gifted education started in American schools in the 1920s through research with the goal of better understanding individuals who appeared to learn quickly and, potentially, effortlessly (Subotnik et al., 2011). In their explanation of gifted education and the potential directions the program could go in, Subotnik et al. (2011) stated that as early gifted programs focused on students' intellectual advancement, intelligence testing became the primary resource for identifying students with potential giftedness. Through a focus on intelligence, gifted programs were meant to support the learning of individuals in areas associated with academic performance. Students are generally identified to participate in accelerated programs through a process determined by the school district. Because programs can vary from district to district, the process in each district can be different. Intelligence testing is one of the most frequently used methods for identifying students for gifted programs (Sternberg, 2017); other methods for identification can include academic testing (Hodges et al., 2018) and teacher referrals (Tirri, 2017). In many states, school districts are given broad flexibility in determining the tools and measures they want to use to identify students for gifted programs based on the goals of their program (National Association for Gifted Children & The Council of State Directors of Programs for the Gifted, 2015).

Identification Methods

Intelligence tests are one measure that can be used in the identification process; their use is considered the most traditional model and is seen as an objective measure (Ferrando et al., 2016). A number of intelligence tests have been utilized, including, amongst others, the Cognitive Abilities Test (CogAT), the Wechsler Intelligence Scale for Children (WISC), and the Naglieri Nonverbal Ability Test (NNAT), to identify those who perform at high cognitive levels. Some intelligence tests utilize non-verbal scores to limit the advantages for students who have developed strong language skills, but the results have demonstrated that there were still gaps based on race and income level (Peters & Engerrand, 2016). Some intelligence tests utilize several sub-tests to allow for a better understanding of different aspects of a student's intelligence, such as analytical skills, spatial recognition, and pattern awareness (Freeman & Chen, 2017).

Beyond intelligence testing, academic achievement testing may be considered as a measure to identify potential for success in accelerated or gifted programs. Academic measures are one way in which schools regularly collect information on student progress. Cao et al. (2017) stated that standardized academic tests can serve the gifted education identification process by helping to pinpoint areas in which students demonstrate strengths. The authors suggested that using standardized assessments can be justified in recognizing individual needs of students. These tests can be annual summative assessments such as state accountability tests, nationally normed tests (e.g., ITBS or SAT), or locally determined assessments that may or may not be nationally normed (e.g., i-Ready, STAR Math, or MAP). However, it should be kept in mind that achievement tests can also be susceptible to unintended bias when students have not had equal prior opportunities to learn (Peters & Engerrand, 2016).

Teacher or guardian referrals are also commonly used for identification practices (National Association for Gifted Children & The Council of State Directors of Programs for the Gifted, 2015). Teachers work with students daily and have knowledge about individual strengths, how students solve problems, and can potentially identify students who have gifted characteristics. Tirri (2017) suggested that the use of teacher input can be a strategy that is beneficial in the identification process, but teacher education would be needed in that area. The use of teacher input could support the identification of students and create a system that does not rely on a single test score. However, Redding (2019) found that teachers frequently gave higher academic ratings to students who shared the same race as the teacher.

Universal screening is a method to reduce or remove the impact of teacher or guardian referrals from the process. This consists of using the same identification methods and processes for all students in a school or at a single grade level. The strategy can be used by school districts to identify students who may not have been referred by a guardian or teacher for gifted program participation. Card and Giuliano (2016) suggested that universal screening is one way to address the underrepresentation of low-income students and students of color; its use can be a positive step towards ensuring that all students are assessed for gifted programs. However, this practice may involve students being assessed using traditional intelligence or academic tests. Hodges et al. (2018) found that identification methods, including universal screening, have not identified low-income students or students of color in proportion to other student demographics.

School districts can encounter challenges when reviewing test scores or teacher input in isolation or solely using universal screening. The use of a single measure to determine giftedness can disproportionately identify students as intellectually advanced by race. In a study to determine bias in intelligence testing, Cao et al. (2017) found that the intelligence tests they examined demonstrated bias amongst racial groups. Conclusions like this have also been made by McCluskey (2017) in a review of gifted identification practices and the disproportionate representation of students from low-income families. To counteract this bias, the author suggested lowering the threshold for acceptable achievement test scores for students who are identified as low-income compared to their non-low-income peers.

Similar to cognitive measures, academic measures used in isolation may create disproportionate representation. To better understand the underrepresentation of students of color in gifted programs, Grissom and Redding (2016) examined math and reading test scores as an identification method for gifted education in elementary schools. The authors stated that African American students who performed similarly to non-Hispanic White students were not identified for giftedness at the same rates. The authors suggested that teacher bias and referrals of students who could benefit from the services may follow racial lines. In their study of gifted identification and racial groups, Romanoff et al. (2009) found similar challenges of disproportionate identification using reading and math assessments. The authors of these studies did not use the same achievement tests demonstrating that even if identification measures change, disproportionate identification of low-income students and students of color may remain.

There are other challenges to address in this area. In 2009, Moule undertook a review of unconscious bias as grounded in critical race theory. This concept was supported via a qualitative study about African American students who take AP classes by Jeffries and Silvernail (2017), who noted that the students who participated did not believe white teachers supported them effectively or believed in their ability to succeed. Information provided by these authors dispute the concept that all teachers can effectively identify students of color that could benefit from gifted education.

There are several reasons historically underrepresented populations are under-identified for gifted programs. One factor identified by Olszewski-Kubilius and Corwith (2018) is the limitation of early opportunities to access gifted programs for low-income students and students of color. The authors cited identification practices such as intelligence or academic testing as barriers to accessing gifted programs; such methods require students from low-income homes to perform at levels equal to those of their peers who may have access to more educational learning opportunities outside of school. Because of the differences in identification of low-income students and students of color for gifted programs through intelligence and academic testing, recent changes have been recommended and attempted (Card & Giuliano, 2016). The performance gap between low-income students and students of color can start early in elementary school (García &

Weiss, 2015); therefore, identifying students early for gifted programs may be beneficial before performance gaps increase.

Multiple Measures

Historically, unitary measures of intelligence have been used for identification of students who could benefit from gifted programs (Romanoff et al., 2009). The use of multiple measures for gifted education identification is a shift in thinking from using a single test score. In a meta-analysis of identification practices, Hodges et al. (2018) noted that many states and school districts are starting to use multiple measures for identification. One challenge with this change and recommendation is that there is no guidance on what are considered appropriate measures that can be used together. Hodges et al. (2018) concluded that the use of multiple measures addresses the problem of a single score, but their study did not define what multiple measures would mitigate the problem or how to use them for identification. The use of a single measure to identify students for gifted education, regardless of the type of measure, can disproportionately identify students based on race and socio-economic status (Sablan, 2019).

The use of multiple measures as a structure for accelerated program identification processes was also studied by Cao et al. (2017). In their meta-analysis of the gifted education literature, the authors found that there were two issues of concern when recommending and using multiple measures. The first is that the literature does not identify which assessment tools or methods to use for the multiple measures, while the second is that the literature does not discuss or describe how the multiple measures should be used in combination. These issues highlight the challenge for K-12 educators attempting to use multiple measures to more equitably identify low-income students and students of color who have been historically underrepresented in gifted programs and accelerated classes.

One consideration in the use of multiple measures is to identify students based on the goals of the program. Hertzog (2017) stated that multiple measures could include teacher referrals for identification of gifted students. The author's recommendation was that a classroom teacher would have an understanding of the goals of the program to which the student would be referred and could therefore better identify students for participation. Based on the teacher's knowledge and understanding of the program, in conjunction with student performance, a referral would be a subjective measure that could be used in combination with intelligence or academic test scores to identify students.

Disproportionate Representation in Accelerated Programs

Disproportionate representation of low-income students and students of color is not a new challenge facing K-12 schools. There are numerous studies that have demonstrated that disproportionate populations exist in accelerated programs across the United States (Warne & Price, 2016). The challenge is also not limited to a specific sub-group of students. Bradbury and Corak's (2015) research discovered inequalities in opportunities and outcomes in both socioeconomic status and race that begin early and continue through students' K-12 education. The authors also stated that, per capita, these disparities are larger in the United States compared to other countries.

Disproportionality: Roots and Effect

Students from higher income homes have primarily comprised the gifted education population for many years (Steenbergen-Hu & Olszewski-Kubilius, 2016).

When identification processes for accelerated programs use a single measure, such as an intelligence test, to identify students, it may exacerbate disproportionate representation of specific sub-groups in those programs (McCluskey, 2017). Steenbergen-Hu and Olszewski-Kubilius (2016) also observed that school districts using these measures continue to limit participation in gifted services to a specific percent of students who qualify. The authors noted that the measure to identify students may change but limiting the number of students identified can negatively impact proportionate identification of students for gifted education.

School districts have attempted to more proportionately identify low-income and students of color by revising intelligence tests or measures. However, in a study to explain underrepresentation in gifted programs, Grissom and Redding (2016) determined that assessment changes did not increase the number of students identified as there were other challenges to low-income students and students of color performing at equal levels to their peers. The authors explained that teachers have an impact on student outcomes; specifically, teacher demographics and expectations have an impact on student performance. Students can demonstrate higher performance when they are in the class of a teacher with similar demographics, which could impact decisions about their giftedness made through the use of intelligence or academic measures (Redding, 2019).

Potential bias in the assessments used by school districts to identify students who can benefit for gifted education may impact proportionate identification; assessment bias may be present in the intelligence or academic measures available to school districts and can negatively impact equitable gifted identification for gifted programs if used unitarily. McCluskey (2017) posited that tests of intelligence may instead be designed to identify students who have had early access to books and other opportunities. In contrast, a student who could not utilize these tools may not perform at the same level. As identified in a study by Olszewski-Kubilius and Corwith (2018), this phenomenon can also be seen in academic assessments that are used for identification purposes. Through a study to better understand how quantitative methods can be used to impact disproportionality by race, Sablan (2019) learned that quantitative methods including descriptive and inferential statistics could be used to mitigate bias if the model went beyond traditional methods.

Bias in Individuals

Individual teacher bias, whether conscious or unconscious, can lead to more favorable academic outcomes for students that identify as the same race as their teachers (Redding, 2019). In a randomized control study, Copur-Gencturk et al. (2019) discovered that, when estimating students' mathematical abilities, teachers had favorable bias for non-Hispanic White students over students of color and boys over girls. Card and Guiliano (2016) found that when a large urban district changed from random referral of students in second grade to universal screening, more historically underrepresented students were identified for accelerated programs. The authors of these studies suggested that when teachers were taken out of the identification process, identification of students of color was more representative of the population.

Bias may not be explicit, but rather unconscious and unknown to an individual teacher. In an article written to increase the understanding of unconscious bias and unintentional racism, Moule (2009) suggested that all individuals have bias that they may not know is there; in addition, according to the author, when an individual is made aware of their bias, they can change their beliefs, but that process takes effort and time. Milner (2013) stated that bias can be created based on the location of a school. Through the lens of critical race theory, the author indicated that an educator may assess a student's abilities based on where they live or the school they attend, and impact decisions made about who should have access to gifted programs.

Due to both conscious or unconscious bias in individuals, decisions about participation in gifted programs using teacher recommendations or rating scales can introduce that bias into the process. Via a student-teacher race study, Redding (2019) found that teachers can make decisions in favor of overrepresented sub-groups based on race. Milner (2013) found that the neighborhood in which students lived could impact the decision-making of adults. Relying on teacher perceptions and input as identifiers for students best suited to gifted education may perpetuate the underrepresentation of low-income students and students of color in gifted programs, suggesting that the common measures and methods used in the process need to be carefully examined by K-12 school districts.

Theoretical Foundation

The following sections introduce a framework and theory to support the use of multiple measures to identify students for accelerated program participation. As discussed in previous sections, students are disproportionately identified for accelerated programs by income level and race in K-12 school districts. The intentional or unintentional bias in program identification processes can put groups of students on different educational tracks in primary, secondary and post-secondary education. Those who ascribe to the cultural mobility theory support the need to increase access to accelerated programs for low-income students and students of color with the aim of improving their social standing. Using a test score can be one way of measuring student knowledge and skills in each subject or context (Farley-Ripple et al., 2019). Changing processes to include multiple measures for identification of student strengths is supported by the three-stratum theory in intelligence.

Cultural Mobility Framework

The cultural mobility framework suggests an individual's social standing can be improved through education. DiMaggio (1982) argued that if resources were in place to support students who did not have access to early learning or accelerated coursework, they would better utilize these opportunities compared to their peers. In an article about cultural capital and educational inequality, Jæger and Karlson (2018) observed that cultural mobility suggested that the cultural capital return would be higher for students who had a low socio-economic standing compared to higher socioeconomic students.

Cultural capital can be explained as the ability to utilize economic and social assets to benefit one's family (Jæger, 2011). Jæger (2011) suggested that students utilize cultural capital obtained from their parents by taking advantage of their parents' direct knowledge or by passively accessing the education that parents with cultural capital know how to navigate. One way in which parents may do this is by understanding how to access accelerated programs. This knowledge can present an advantage to students and create inequitable access to accelerated programs for students who come from families that do not have this level of cultural capital.

Beyond K-12 education, cultural capital can be beneficial for older individuals from less advantaged backgrounds. In a study about adult education, Cincinnato et al.

(2016) posited that there is a link between social background and participation in adult education. However, the authors suggested that those connections could be broadened by providing opportunities for individuals to engage in learning opportunities earlier in their lives. Policies that ensure early and more equitable access to accelerated education programs provided to students who come from disadvantaged backgrounds can limit the negative impact of cultural capital (Cincinnato et al., 2016).

Cultural mobility theory may support the concept that students identified as low-income would benefit more from accelerated program identification compared to their non-low-income peers. The increase of historically underrepresented students identified for accelerated programs may result in improved cultural capital in the community and positive outcomes for students in the future.

Three-Stratum Theory

The processes in K-12 education for the identification of students qualified for accelerated programs typically use individual cognitive ability assessments. Cognitive intelligence testing has been conducted for decades and has utilized a number of models. Some researchers have suggested that only measuring for cognitive ability is a limited way to view intelligence. The three-stratum theory, developed by John Carroll in 1993, suggested that individuals have different abilities of intelligence that are used to perform different tasks (as cited in Benson et al., 2018). This theory supports the idea that individuals can demonstrate their intelligence or specific cognitive traits in different ways.

The three-stratum theory has been advanced by other researchers. Beaujean (2015) discussed Carroll's model of cognitive testing in an article comparing different cognitive assessments. The author suggested that Carroll's three-stratum theory is widely accepted as a common way to measure cognitive abilities. Carroll, in 1993, utilized a factor analysis to demonstrate that cognitive ability is made up of multiple factors that are independent of one another. Carroll also suggested that the factors of intelligence differed among individuals; combined, those factors together comprised an individual's cognitive ability (as cited in Beaujean, 2015).

The three-stratum theory supports the use of multiple measures to identify student strengths, particularly for processes that use intelligence tests as a measure. An understanding of different cognitive areas can guide the selection of measures to use in the identification of students who could benefit from accelerated programs. Not all students perform well on a single measure and no single measure can capture all the cognitive dimensions. Instead, students may perform well on different assessments to demonstrate they can benefit from accelerated courses. Researchers have suggested using multiple measures of academic potential to improve identification processes for accelerated programs (Hertzog, 2017).

Educational Test Theory

Educational test theory is a method of understanding a student's knowledge and skill using statistical tools (Mislevy 1996). The theory is based on the idea that a test can measure cognitive or academic performance on a given topic at both a point in time and over time. The information obtained from assessments can be used in multiple ways, including understanding individual progress, grouping students, adapting instruction, and planning next steps for learners (Farley-Ripple et al., 2019). Educational assessments can be used as part of a teaching and learning cycle in which teachers use assessment data to identify steps that can benefit learners, as well as to identify student potential or areas of strength that could be supported through accelerated programs.

There are different ways to test student knowledge and skills. In an article about cognitive and educational testing, Sternberg (2017) noted that not all tests are beneficial for all students, depending on their culture and social backgrounds. The author suggested that students who benefit the most from educational testing are those who come from environments that place a high value on testing and achievement. Sternberg (2018) also mentioned that identifying ways in which students can demonstrate their knowledge and potential could benefit school systems in understanding how they can better support students from various backgrounds.

Both educational test theory and three-stratum theory provide a foundational framework for the exploration of how multiple measures can be used for accelerated program identification, how statistical variances may be identified on assessment outcomes for different subgroups and how the variances may be accounted for to identify students who have been historically overlooked. Educational test theory can be used to provide a foundation to assess understanding in a variety of content areas, while three-stratum theory lends support to the use of multiple measures when attempting to understand more than one aspect of student strengths. Together, these theories can be used to support understanding of how assessments are impacted by cultural and social backgrounds as identified in cultural capital theory.

The theories identified in this section help ground a study about the processes and results that can lead to or disrupt disproportionate identification of low-income students and students of color for accelerated programs. If a school district adopts policies that support the use of multiple measures of academic potential to select students for participation in accelerated courses, it may increase the number of these historically underrepresented students. The use of these theories in combination support the concept that school district policies and practices for accelerated program identification can reduce disadvantages for individuals and improve future life outcomes by adopting methods to increase participation of low-income students and students of color in accelerated courses.

Summary

The purpose of this literature review was to review processes, methods, and measures that have been used by school districts to identify students for accelerated programs and to better understand why low-income students and students of color are identified at disproportionately lower rates as compared to their peers. Information in the literature review examined four themes to provide (a) background about accelerated programs in K-12 education; (b) processes that school districts use to identify students for accelerated programs; (c) how bias can impact disproportionate numbers of sub-groups in accelerated programs; and (d) theories that can be used to support a need to disrupt the historical underrepresentation of low-income students and students of color in accelerated programs.

Past practice and recent recommendations for gifted identification have suggested the use of multiple measures that may include an assortment of cognitive tests, academic measures, and teacher input (Hamilton et al., 2018). Bias in both intelligence and academic measures can negatively impact the reliability and validity of these data for identification purposes (Olszewski-Kubilius & Corwith, 2018), and unconscious bias can impact teacher perceptions about students (Moule, 2009). There is evidence that past practice has negatively impacted historically underrepresented students from accessing accelerated programs, and a deliberative use of multiple measures could change this pattern.

This review of the literature provides information to support the idea that the assessments and measures used to identify students for accelerated programs may be a cause of the disproportionate representation of low-income students and students of color. The utilization of several different tests and assessments with the goal of mitigating bias for historically underrepresented students has been unsuccessful. Researchers have recommended the use of multiple measures to lessen bias, but no suggestions have been made about which assessments or measures to use or how to use the measures together. Without a clear method for using multiple measures for identification of students for accelerated programs, the underrepresentation of low-income students and students of color may continue.

CHAPTER 3: METHODOLOGY

The goal of this study was to test the significance and examine the between-group differences for student socioeconomic status and race based on the results used for the identification of students for accelerated programs. To reach the goal, a statistical analysis of the results of three assessments administered to second grade students in a K-12 school district for accelerated program identification purposes was conducted. To analyze the between-group differences using assessment data, three research questions guided the study:

1. What is the difference in assessment results for low-income students and students of color compared to their peers at a given grade level?
2. What is the significance of between-group differences when compounding dependent variables measured within and between independent variables?
3. Can assessment results used to identify students for accelerated programs be adjusted based on between-group differences to affect the identification of low-income students and students of color for accelerated programs?

The information in this chapter details the research method and design, data analysis methods, the instruments, and participants.

Research Method

A quantitative methodology can be used in a study undertaken to analyze the relationships between variables and test hypotheses. In an explanation of the differences between quantitative and qualitative methods, Park and Park (2016) stated that one purpose of quantitative research is to identify the underlying reasons a phenomenon may exist. The authors suggested that a quantitative study can help demonstrate statistical interactions and increase the understanding of phenomena. In a study that analyzes assessment data used in the process to identify students for accelerated programs, a quantitative method can be used to examine statistical differences and relationships for students who identify as low-income or non-White or Asian.

Neither qualitative methods nor mixed methods were selected for use in this study. Qualitative research methods can be focused on the collection of data that help in understanding lived experiences, behaviors, and emotions in which the findings can evolve (Levitt et al., 2018). Mixed methods research utilizes both qualitative and quantitative research to integrate findings that are insightful and result based (Levitt et al., 2018). The purpose of this study was to test the significance of between-group differences for student demographics and student assessment data when used as the variables. Quantitative research can be used to measure and analyze relationships between variables. In an article describing quantitative research, Watson (2015) stated that there are two types of variables: independent and dependent. The author suggested that independent variables can influence dependent variables. A quantitative method was used in this study to determine if the independent variables of low-income and racial demographics influence cognitive and academic assessment outcomes for students. Measuring the potential impact of independent variables on dependent variables make this method an appropriate choice.

Cognitive and academic assessments in a K-12 school district were examined to understand if they are impacted by student demographic factors which may then create inequities for the identification of low-income students and students of color for accelerated program participation. In their meta-analysis, Cao et al. (2017) found that a method for identifying the appropriate tools to use for gifted identification and how to use them together has not been identified. A quantitative methodology using assessments as dependent variables, as proposed in this study, may help reveal underlying relationships or biases associated with assessments. If factors are found that indicate bias selection, then they may be accounted for in the identification process for the purpose of reducing disproportionality between subgroups of students in the selection of students for accelerated programs.

Research Design

A quantitative research methodology with a large population size can support research questions focused on analyzing the between-group differences for assessment results (Rahi, 2017). A causal-comparative design can be used to examine the between-group differences and the relationships of those differences to the groups (Apuke, 2017). In separate examples, Lodico et al. (2010) and Salkind (2010) stated that causal-comparative design may be beneficial in explaining measurement differences between groups in which the events have already occurred or cannot be manipulated, including socioeconomic status, race, and gender. A review of data from assessments that have already been administered can be causal-comparative or ex post facto.

Descriptive, correlational, and experimental research designs were not selected for use in this study. A descriptive research design can support the understanding of a setting or situation through the collection of data, but it does not measure between-group differences (Solheim et al., 2017). Salkind (2010) stated that correlational research designs can measure the effect of variables, but that only one group of individuals are used in a study. The purpose of this study was to test the significance of between-group mean differences for more than one group: low-income students and students of color. Experimental research designs can be used when independent variables can be manipulated and random sampling can occur (Borgianni & Maccioni, 2020). In this study, events supporting the data have already taken place; thus, that data cannot be manipulated or changed. This makes a causal-comparative design an appropriate choice for the study of between-group differences for student assessments based on low-income status and race.

Instruments

Assessment data used in a study to test the significance of between-group differences for sub-groups of students were obtained from the K-12 school district student information system with the approval of a research project proposal. The student information system, Skyward, is a software database that securely houses all student information collected and reported for the normal operations and legal requirements of the district, including demographics, schedules, grades, and longitudinal data.

Three assessment instruments were used for this study, all of which are administered to all second-grade students in the K-12 school district for the purpose of accelerated program identification. The results of those assessments are recorded and stored in the student information system. The assessment results from these three instruments are the dependent variables for this study and include: a standardized cognitive assessment (Naglieri Nonverbal Ability Test, 3rd), teacher-administered reading comprehension test (Fountas & Pinnell Benchmark Assessment System), and a computer-adaptive mathematical skills test (Renaissance STAR Math).

Naglieri Nonverbal Ability Test

The Naglieri Nonverbal Ability Test, 3rd Edition (NNAT3), is a web-based, nonverbal cognitive assessment that can be used to screen students for identification into gifted education programs; it is one of many available cognitive assessments from which school districts can choose. The district in this study selected the NNAT3 test for easier standardized administration, computerized scoring and to reduce potential bias for English learners of a language-dependent assessment. Giessman et al. (2013) found the use of a nonverbal assessment can benefit students from non-English speaking homes or culturally diverse backgrounds, as well as individuals with limited early learning opportunities. In a comparison of student performance using the NNAT3 and Cognitive Abilities Test, Form 6 (CogAT6), the authors noted that Latino, multiracial, and ELL students performed less disparately on the NNAT3 compared to their White and native English-speaking peers.

A full cognitive test, such as the CogAT6, often analyzes areas such as verbal and quantitative areas of intelligence, in addition to nonverbal abilities. In an article about concerns and disagreements with the NNAT3, Naglieri and Ford (2015) addressed the use of a nonverbal assessment to measure cognitive abilities suggesting that all assessments have limitations and a nonverbal test can assess a students' general ability level. The authors stated that they believed, along with academic measures, verbal and quantitative tests could be used for gifted education testing; however, such an assessment could be biased against students of color and those who did not speak English as their first language. The NNAT3 is a nonverbal cognitive assessment that attempts to limit bias that can be identified through verbal and quantitative tests and limit the under identification of low-income students and students of color for accelerated programs by focusing on the identification and completion of increasingly complex patterns rather than vocabulary or computation skills. In the NNAT3 Manual, Pearson (2018) discussed the test reliability using the Rasch item response model to compare the NNAT2 with the NNAT3, utilizing both online and paper tests. Pearson found that the NNAT3 was reliable and the scores were comparable across the samples. To test the validity of the test, Pearson (2018) used the NNAT2 and Otis-Lennon School Ability Test (8th edition). The scores from all three tests were correlated and no significant effect was identified, which indicated the NNAT3 was valid and reliable as a cognitive ability measure.

To assess a student's nonverbal cognitive ability, all second-grade students in the Pacific northwest school district are administered the NNAT3 as part of the universal screening process. Both the national and local norms are reviewed, and the more favorable score of the two is used for identification purposes. The assessment is administered in the first trimester in each student's classroom throughout all fifteen elementary schools. Classroom teachers are provided training on test administration and administer the assessment as their schedule allows during the school day. Scores are collected through the web based NNAT3 application and uploaded into Skyward by the district for recording and reporting purposes.

Fountas & Pinnell Benchmark Assessment System

Leveled reading comprehension assessments are administered in the Pacific northwest K-12 school district using the Fountas & Pinnell Benchmark Assessment System (BAS). Three times a year, in the fall, winter, and spring, the reading comprehension test, BAS, is administered by teachers one-on-one with students in kindergarten through fifth grade. The assessment utilizes a series of leveled fiction and non-fiction books that increase progressively in complexity to measure a student's reading comprehension skills (Fountas & Pinnell, 2017). Individuals can take between 10-30 minutes to complete the assessment with a teacher. Teachers assess a student on oral reading fluency and comprehension skills to identify an independent and instructional reading level for each student.

The use of assessments that are administered individually can create challenges in training and time. Klingbeil et al. (2015), in a study that compared validity of screening tools, suggested that time and resources would need to be allocated to administer individual assessments. In addition, the authors mentioned that subjective assessments like the BAS could necessitate that teachers be continuously trained on their use and implementation to validate the results. The reliability of the BAS test has been evaluated by Heinemann (2012) by comparing student scores from both fiction and non-fiction texts. The use of the test-retest model demonstrated that BAS was a reliable assessment. The validity of the reading test was determined by comparing student scores with scores from other reading tests (Heinemann, 2012). This comparison was done using the Reading Recovery and Slosson Word Test. The results of the comparisons demonstrated that BAS was a valid assessment of student reading comprehension levels.

In the K-12 school district used for this study, teachers are trained on the administration of the assessment annually by school-based instructional coaches. During the training, teachers calibrate their scoring through videos and practice using the BAS. Student independent reading scores are recorded in Skyward by classroom teachers during all three assessment windows. The school district locally norms the fall literacy scores for use in identification of students for accelerated programs in the district.

STAR Math

Students' mathematical skills are measured through STAR Math, an online assessment from Renaissance Learning (2020). STAR Math is an online computer adaptive test that is norm-referenced and provides information to educators about individual student understanding and computational skills in mathematics (Topping, 2020). Computer adaptive assessments use student responses from past questions to adjust the difficulty of questions (Johnson et al., 2020). In a study aimed at improving understanding of the growth model of STAR Math, Johnson et al. (2020) posited that the online assessment can accurately measure student progress in relation to grade-level expectations for student math abilities. STAR Math can be taken multiple times during a school year to measure student progress in that subject over time.

STAR Math assesses math skills in counting, ratios, operations, number systems, geometry, measurement, and equations and reports comparisons to expected student grade level progress during the year, national norms, and recommended instructional levels (Meador, 2018). This information can be used to identify areas of student strength within math and potential supports that can be provided to students to continue their mathematical growth. Reliability and validity studies involving STAR Math were provided by Renaissance Learning (2020). Reliability was tested by having the same students take the test during multiple administrations while, over time, validity was determined by utilizing construct validity to ensure the test measured math skills and the scores reported accurate information (Renaissance Learning, 2020). STAR Math was determined to be both reliable and valid in a study by Renaissance Learning.

The STAR Math assessment is administered to all first through fifth grade students in the school district used in this study. Students only participate in the assessment in the fall, allowing teachers to understand how they are performing at the beginning of the year. The data from the assessment are collected from the STAR Math online application and uploaded into Skyward for district recording and reporting. For the accelerated program identification process, the data are locally normed to determine if the national norm or local norm for individual students is more favorable. The higher norm is then used for identification of students for accelerated programs.

Independent Variables

Two independent variables were used in this study including socioeconomic status and race. Socioeconomic status is the social class of an individual or group, measured by family income level, educational attainment or other social or economic measures. Free- and reduced-price meal eligibility status in K-12 education were used to measure socioeconomic status, specifically to determine students who are considered low-income (Gajo et al., 2019). Students are either considered to be low-income or not low-income.

Students of color is a collective racial term used for non-White, non-Asian students. To receive federal funds, elementary and secondary schools must ask families to identify which of seven categories best fits their child's race: American Indian or Alaska Native, Asian, Black or African American, Hispanic, Native Hawaiian or Other Pacific Islander, White, and Two or More Races (U.S. Department of Education, 2008). The use of the phrase, students of color in this study, i.e., identified by a racial category that is not White or Asian, represents those who have historically been underrepresented in accelerated programs. All seven racial categories were analyzed to test the significance of the relationships between each category and the dependent assessment variables. Both low-income students and students of color have been historically underrepresented in accelerated programs which is why both sub-groups have been included as independent variables.

Participants

With more than 15,000 students in kindergarten through twelfth grade, the school district used for this study is considered a large-sized school district in Washington state. The district is comprised of an urban community serving five municipalities that includes a diverse student population. After seeing overall growth in student enrollment numbers over the previous four years, overall enrollment has remained steady for the past three. An equal number of Asian, Hispanic, and White students make up seventy-five percent of the student demographics, followed by Black, Pacific Islander, and Native American students. The proportion of bi-racial students is growing while the overall number of students has remained steady. From 2017 to 2019, the percentage of students receiving free and reduced lunch has gone down slightly from 53.7 to 52.4 percent, while, for the past five years, the proportion of English language learners has fluctuated between sixteen to eighteen percent reflecting the district's continuous influx of non-English speaking immigrants from East Africa, Asia, and Eastern Europe.

This study was conducted using data from approximately 1,200 second grade students in a K-12 school district in the Puget Sound region. Data from all second-grade students enrolled during the 2019-20 school year that participated in the district's universal screening process for accelerated program identification were included in the study population. The school district utilizes second grade as the universally screened grade, in which an entire grade level is screened for gifted services eligibility as mandated by state guidance.

Second grade students in the 2019-20 school year were largely representative of the overall elementary student population in the school district but not for the district as a whole. There were 1,207 total students in the second grade, and the racial demographics of the students were: 27% White, 23% Asian, 19% Hispanic, 15% African American, 14% Multi Racial, and 2% Native American/Pacific Islander. The percent of low-income second grade students was 45%, as measured by free and reduced lunch eligibility, which is lower than the district as a whole. The second-grade cohort also had 31% of the student population who were identified for English language services, which was nearly twice the proportion of the district as a whole. The study was close to a total population study because the screening assessment is given to all second-grade students, except for students from families electing to opt-out of the process.

Data Analysis Methods

SEM can be used to test the significance of between-group mean differences of socioeconomic status and race on assessment data used in the identification of students for accelerated programs (Tarka, 2018). SEM, also known as confirmatory factor analysis, can be used to analyze relationships between all variables, observed and unobserved, and measure the significance of the relationships between variables when analyzing sets of data (Vogt, 2007). Crossman (2020) stated that SEM allows for multiple regression analyses of relationships between independent variables and dependent variables to be examined.

There are multiple methods that could be used to analyze the between-group relationships of assessment data used in this study. SEM is a better choice to analyze the relationship between independent and dependent variables compared to a common factor analysis or principal components analysis as it allows for correlations between and among independent and dependent variables (Crossman, 2020). An analysis of variance (ANOVA) and a t-test can compare mean scores of groups, but the output for these tests only provide significance and not coefficients (Emerson, 2017). The use of SEM may allow for a model to be created, analyzed, and developed further in the future with the inclusion of more data if needed.

In a study to identify the significance between student demographics and assessment outcomes, SEM was utilized to look for significant coefficients and relationships that may influence disproportionate accelerated program identification for low-income students and students of color compared to their peers (Hall & Malmberg, 2020). The outcomes from SEM can be used to understand the research questions and test the hypotheses of the study, which are centered around understanding how the assessments used for identifying students for accelerated programs can create the disproportionate outcomes.

Structural Equation Modeling

SEM is a combination of statistical techniques that can support the proposed analysis of data by identifying coefficients and relationships between independent and dependent variables and within measured variables that have not been observed from other means. This data analysis method allows for flexibility in creating the model, and combines factor analysis and regressions analysis (Crossman, 2020). In a description of

SEM, Breitsohl (2019) indicated that this method can be used to test large sample sizes and multiple variable relationships at one time in social and behavioral sciences. Alessandri et al. (2017) stated that SEM could be used to measure the variance and covariance of test scores and that it provides more information compared to other methods for analyzing data such as ANOVA.

In an explanation of SEM, Qureshi and Compeau (2009) suggested that SEM can measure between group differences for separate variables or in combination to measure latent constructs. The authors stated that SEM can test for variance or linear relationships of variables, as well as identify the significance level between the independent variables of low-income and racial demographics, and dependent variables of assessment data. The significance level may be used to identify correlations and relationships to help understand the disproportionate identification of historically underrepresented sub-groups for accelerated programs.

A statistics software package, SPSS Amos, was used to conduct the SEM. To begin this process, all the data were entered from a single spreadsheet obtained from the school district's Skyward database that included all second grade students' race, free and reduced lunch status, grade level, NNAT3 cognitive assessment score, BAS reading level, and STAR math score. Crossman (2020) stated that a graphic diagram of the relationships, called a path diagram, is fundamental to SEM as it provides a visual of the graphic model and relationships between variables that were measured. The author wrote that path diagrams provide SEM with hypothetical models and sets of relationships that can be used for the analysis.

A chi-square test was used to run the goodness of fit test for each path diagram created to identify which diagram can provide data for the SEM analysis (Meyer, 2020). Adjustments can be made to the path diagram by removing variables or creating alternative paths to find one that has statistical significance and sufficiently low error. This produced an estimated matrix to ensure that the model was good and adequate to conduct the SEM. Once goodness of fit is determined from the chi-square test, a measurement model can be specified (Crockett, 2012). In an explanation of how to conduct SEM, Crockett (2012) stated that the model will identify the direct relationship between independent and dependent variables, in addition to any indirect relationships in the model.

A graphic model that includes the variables must be created and identified for SEM to work effectively. To increase the likelihood of an effective structural model, Crockett (2012) suggested setting the causal path from each independent variable to the corresponding dependent variable to zero. This fixed number can help support more reliable scores when the SEM is run. To successfully run SEM, Crockett (2012) mentioned that the output produces unidirectional relationships, meaning that the number of elements produced in the covariance matrix were less than what were identified in the measurement model.

There are assumptions that must be met to ensure SEM is as accurate as possible; these can include sample size, no missing data, and model specification (Sarstedt & Ringle, 2017). Sample size can create data that are unstable or that have high variability. In an article about sample size requirements for SEM, Wolf et al. (2013) stated that sample size could be adjusted for when evaluating the model for fit. The authors also suggested that missing data can impact the results more than the size of the sample. The data analyzed for this study were close to total population with all data for each student, so the assumptions of sample size were addressed and students with missing data were removed from the sample. Model specification errors can occur when relevant variables are left out of the model (Vogt, 2007). Model errors can vary in impact and can be addressed by testing the model for consistency and fit (Bollen & Noble, 2011).

Prior to selecting the final path diagram, a chi-square test was used to run the goodness of fit which assesses the overall discrepancies between the model and covariance matrices. If goodness of fit is less than or equal to 0.05, the parameter estimates would be considered significant (Crockett, 2012). The structural model can then be analyzed to determine the extent to which the model fits. This was done by analyzing the critical value of each parameter.

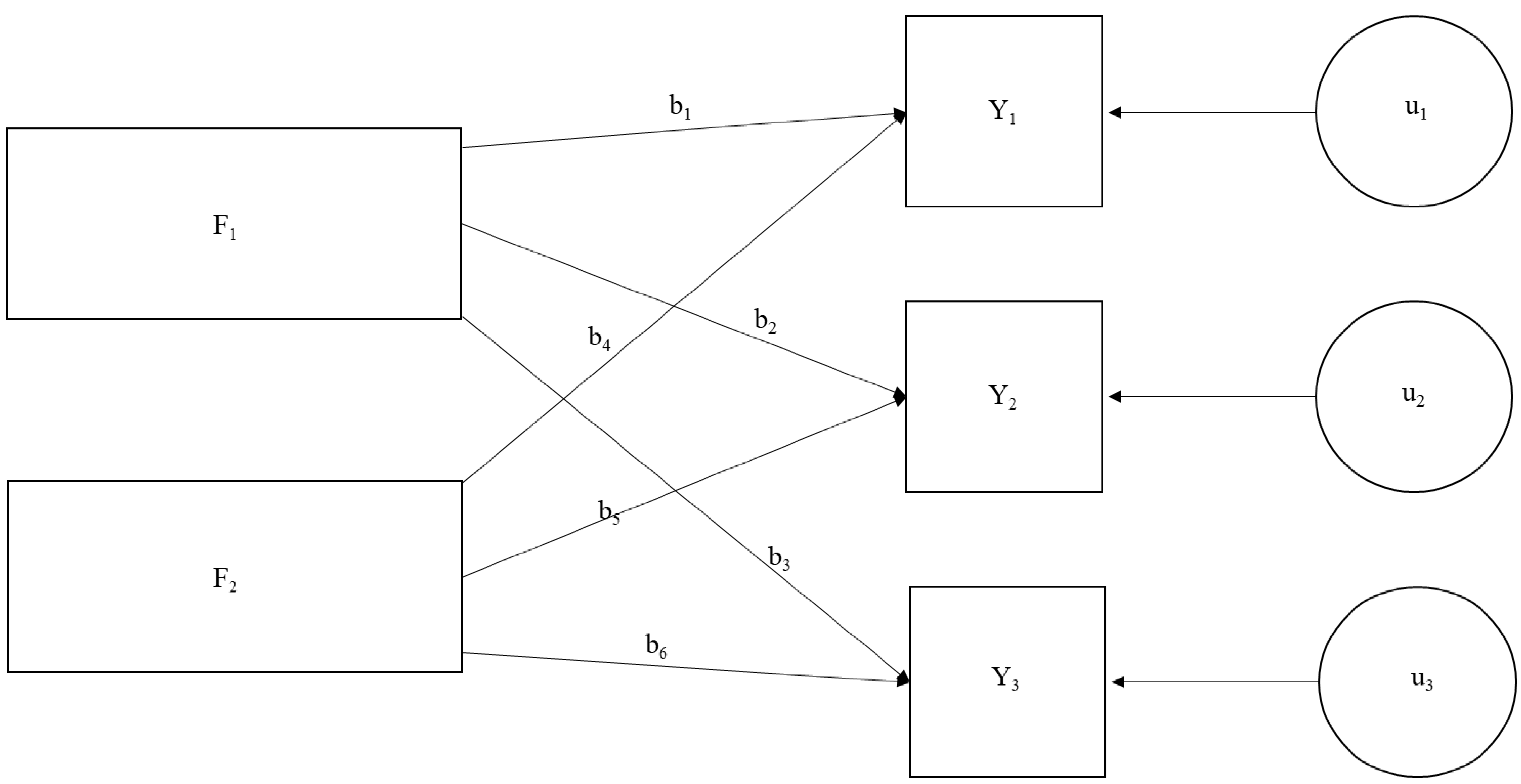
Research Questions

Once the constructs from SEM were evaluated, coefficient scores were calculated to determine the significance of the relationships between the independent variables and the dependent variables in the path diagram. Significance was measured by a *p*-value less than or equal to 0.05. SEM can help in analyzing the score variance in assessment results and identify significant coefficients. The resulting model can help clarify how low-income status and race impact the outcomes on multiple identification measures and be used to answer the research questions used for this study.

The first research question was: What is the difference in assessment results for low-income students and students of color compared to their peers at a given grade level? Based on the results of a linear regression run for the SEM analysis, the difference between low-income students compared to non-low-income students can be measured for the nonverbal cognitive assessment and the academic measures used for accelerated program identification. Figure 1 represents a multiple factor SEM model in which the two independent variables may impact multiple dependent variables, as examined in the first research question. Regression coefficients for all racial groups and socioeconomic status may be identified for the same three assessments. The between-group differences in the dependent variables as identified by the coefficients may help in understanding the degree to which race and low-income demographics of historically underrepresented sub-groups of students play a role in different rates of identification for accelerated programs.

Figure 1

Multiple Factor Structural Equation Model for Research Question 1

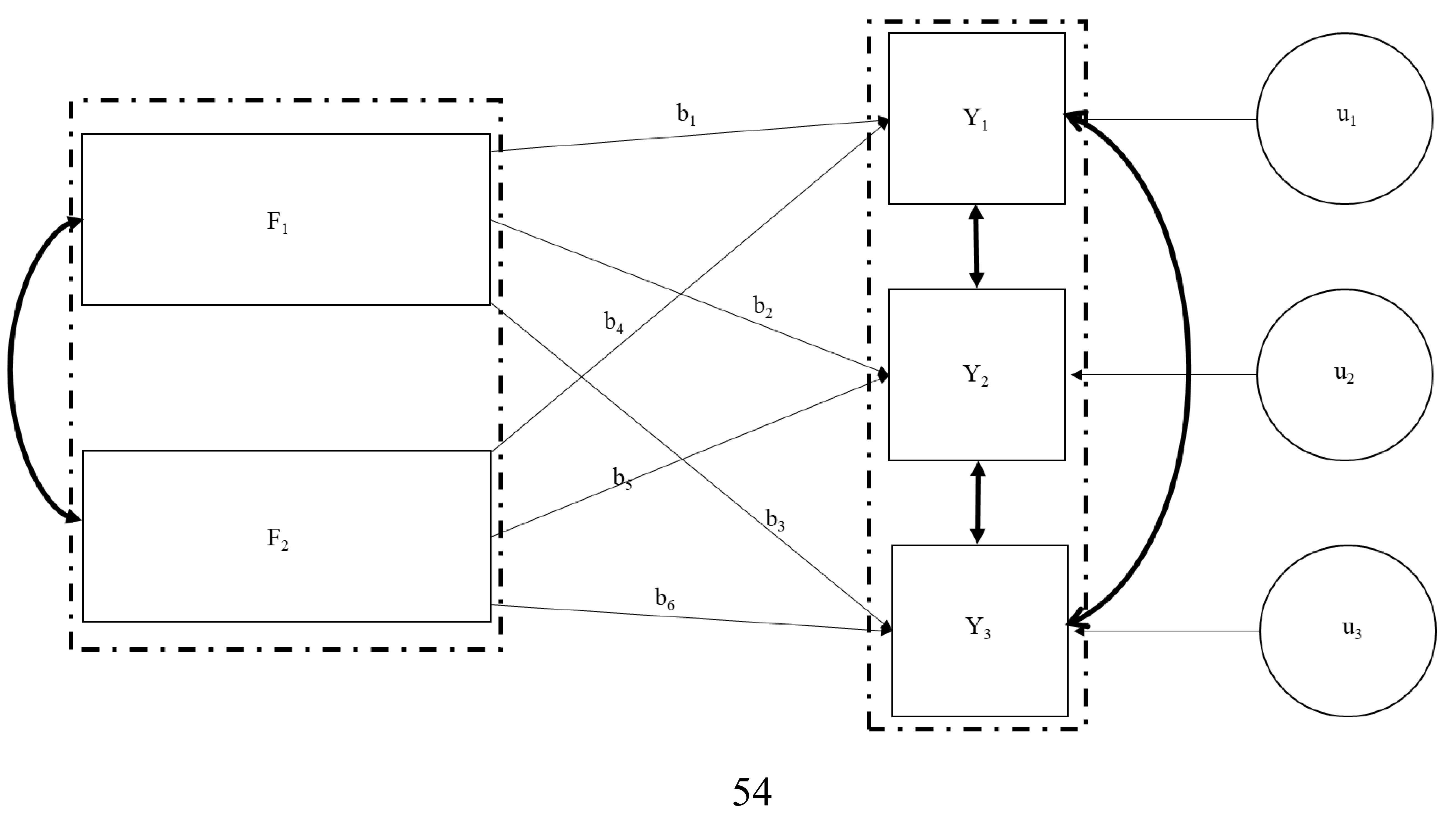


Note. F represents the independent variables that can have a direct impact on the measured dependent variables, which are designated as Y. The arrows between F and Y are the regression coefficients represented by b (Breitsohl, 2019). Indicator intercepts are represented by u, which is the error in measurement associated with each dependent variable.

The second research question sought to test the significance of between-group differences for dependent assessment variables when compounding independent demographic variables. A multiple variable model identifying differences between and within compounded variables is seen in Figure 2. This model shows how SEM was used to answer research question two. SEM can help when analyzing between-group differences when low-income and race are used as compounding independent variables. These compounded variables may identify if the significance is greater or smaller when a student identifies as both low-income and non-White, non-Asian compared to their peers. A significant coefficient may also be found between combinations of assessments in relation to low-income and race, which can help in further understanding how the assessment data impacts identification of low-income students and students of color for accelerated programs.

Figure 2

Multiple Factor Structural Equation Model for Research Question 2



Note. F represents the independent variables that can have a direct impact on the measured dependent variables, which are designated as Y. The arrows between F and Y are the regression coefficients represented by b (Breitsohl, 2019). Indicator intercepts are represented by u, which is the error in measurement associated with each dependent variable. The thicker arrows represent variable combinations when they are compounded, and the dash-dotted lines represent the groupings of variables.

The final research question was: Can assessment results used to identify students for accelerated programs be adjusted based on between-group differences to affect the identification of low-income students and students of color for accelerated programs? Once the relationships between all of the observed and unobserved variables are understood for low-income students and students of color, significant coefficients from the SEM process that may impact identification outcomes can be accounted for and adjusted for one or more of the individual student assessment results based on the variances found in research questions 1 or 2. These adjustments can be made by applying the coefficients to assessment scores by socioeconomic status or race. The adjusted results can be used to order students by percentiles and compare the results to the order of students in percentiles without adjustments.

Comparative descriptive statistics can be used to analyze the rank order of students in second grade with and without the coefficients found to be statistically significant at $p < 0.05$ applied for the dependent variables. Analysis of the student assessment scores before and after the inclusion of the coefficients may help determine if more students identified as low-income and students of color are rank ordered above a certain threshold, such as the 90^{th} percentile of the grade. The demographic differences between the two sets of data, in the various areas assessed, may be used to answer research question three. Based on the adjustments made to the ranking of students after accounting for the significant factors in assessment outcomes arising from SEM, more low-income students and students of color may be identified for accelerated programs.

To test the significance of between-group differences between the assessment results without and with the coefficient applied, a chi-square test of homogeneity can be used. In an article about Pearson's chi-square tests, Nihan (2020) stated that a chi-square test could be used to interpret findings from two or more groups. The author suggested that a chi-square test of homogeneity could be used to test the significance between two groups and to determine if the distribution fits the expected distribution of observed data. A chi-square test of homogeneity can be used to compare the racial and economic distribution of second grade students with and without the coefficients within the second-grade cohort.

A chi-square test was used to determine if disproportionality exists between students at or above the 90^{th} percentile compared to the distribution of all students in the cohort by race and socioeconomic status. The threshold of at or above the 90^{th} percentile was selected because students selected for accelerated programs are generally identified based on higher scores on standardized tests. Some districts may use different thresholds but at or above the 90^{th} percentile was assumed to be a reasonably high threshold. The chi-square test was significant at $p < 0.05$. First, a chi-square test determined if the distribution of students at or above the 90^{th} percentile before the coefficient is applied is the same as the expected distribution for the total cohort population by race and socioeconomic status. Next, another chi-square test was used to determine if the distribution between groups at or above the 90th percentile with and without the coefficient applied are similar. A third chi-square test determined if the distribution of students at or above the 90th percentile after the coefficient is applied is the same as the expected distribution for the total cohort population by race and socioeconomic status.

Limitations

There are limitations to a study involving student demographic data and assessment scores. Data entry into a student information system relies on an individual manually entering data or mass uploading data sets. Guardians of students' complete documents online or on paper to identify with one of seven broad, federal race categories, but there is no verification process to ensure the racial category identified is accurate. Spicker (2018) suggested that different individuals from different regions can interpret categories differently, creating potential limitations to variables in a quantitative study.

Individuals at all fifteen school locations throughout the school district enter the data into the student information system, Skyward. Data entry errors which may impact the study can occur during these processes. Data from a single grade level in the school district in a single year were reviewed. Results may be different if different cohorts of students or data from different years were used in the data analysis, which can be a limitation.

Assessment data is also another limitation to the study. The school district providing data for the study utilizes BAS as the standardized reading assessment at the second-grade level, an assessment which relies on some professional judgement to interpret a student's responses. Teachers participate in annual training to administer the assessment, but there are times when individual teachers may not have been trained in a given year. The BAS scores are not systematically calibrated across the district and, based as they are on teacher skill and judgement, may skew the data's reliability. This can be a limitation when second-grade data are locally normed.

Students participate in an online math assessment and nonverbal cognitive test, but as mentioned, the reading assessment is performed during a one-on-one meeting between individual students and teachers, which allows some subjectivity. While the scores can be normed, there are limitations in ensuring the assessment scores are equally reliable across the school district.

The limitation of the data accuracy can be mitigated by "cleaning" the data and comparing it to cohort data from second grade students in prior years. Cleaning the data can be completed by checking for duplicate information, verifying all data is complete, and deleting inaccurate or irrelevant information.

Delimitations

There are delimitations in a quantitative study utilizing student assessment data. Data were used from second grade students in a K-12 school district for the 2019-2020 school year. The school district in this study identifies locally administered assessments that are used for instructional purposes at the second-grade level. Other school districts may implement different assessments to support instruction.

Another delimitation is the use of second grade students for universal screening. The state of Washington requires that universal screening be implemented, but school districts have autonomy in making decisions about the grade level in which universal screening takes place. Both the grade level used for universal screening and the assessments used may be different for different districts.

Finally, the findings and results may not be applicable to other groups or grade levels because of the data used, make-up of the students, and uniqueness of locally required assessments. Every cohort of students may be of a different size and demographic make-up, including race, socioeconomic status, and of English language learners, and experiences or access to early learning or other educational supports.

Summary

A quantitative method, SEM, was used to measure the between-group differences, identify significant coefficients, and analyze the relationships within and between independent and dependent variables. The purpose of this causal-comparative study was to test the significance of between-group differences for student socioeconomic status and race and the results of assessments used to identify second grade students for accelerated programs. An attempt to determine if statistically significant observed or unobserved bias can be used to reduce disproportionate identification of low income and students of color in a Pacific northwest K-12 school district was conducted through this study. Second grade students were selected as they are universally screened in the school district for identification of accelerated programs.

CHAPTER 4: RESULTS

The purpose of this quantitative casual-comparative study was to measure the between-group mean differences in assessment data for low-income students and students of color that may contribute to the disproportionate identification of second grade students for accelerated programs. A finding of significant between-group differences in assessment results for socioeconomic status and race can help identify potential test bias. SEM can be used in a quantitative causal-comparative design to measure the significance of the relationships between and amongst independent and dependent variables for a large population of second grade students in a diverse K-12 school district. Data from assessments universally administered to the entire population of second grade students in the district were available for the study. The following research questions and hypotheses were used to measure the significance of between-group differences of assessment tools used for accelerated program identification.

1. What is the difference in assessment results for low-income students and students of color compared to their peers at a given grade level?

H_0 – There is no significant mean difference in assessment results for low-income students and students of color compared to their peers at a single grade level.

H_1 – There is a significant mean difference in assessment results for low-income students compared to their non-low-income peers at a single grade level.

H_2 – There is a significant mean difference in assessment results for students of color compared to their peers at a single grade level.

2. What is the significance of between-group differences when compounding dependent variables measured within and between independent variables?

H_0 – There is no significant mean difference in the relationships between dependent variables and independent variables when compounded.

H_1 – There is a significant mean difference in the relationships between dependent and independent variables when compounded.

3. Can assessment results used to identify students for accelerated programs be adjusted based on between-group differences to affect the identification of low-income students and students of color for accelerated programs?

H_0 – Adjustments based on between-group significant mean differences do not increase the identification of low-income students and students of color.

H_1 – Adjustments based on between-group significant mean differences increase the identification of low-income students for accelerated programs.

H_2 – Adjustments based on between-group significant mean differences increase the identification of students of color for accelerated programs.

This chapter provides the racial demographic information and free and reduced lunch breakdown of the second-grade cohort compared to the overall district population. Next, the second-grade students' data are analyzed to answer each of the three research questions. In the analysis for question two, results from a chi-square tests to determine goodness of fit for the path analysis model are shared.

Presentation of Results

The data used to answer the research questions and test the hypotheses are presented in summaries along with details of the information in tables. Demographic descriptive statistical information about the second-grade cohort data is compared to the total district population to determine how representative the demographics are when compared. Data from the cohort used in the study is compared to cohort data from second grade students in other years. Tables and narrative summaries are shared to present the results from the SEM to answer the research questions.

Demographic Statistics

Second-grade data collected for the SEM analysis included 1209 student records. After the data were cleaned, 1082 student records remained for students enrolled in October 2019. 110 students with incomplete data, such as a missing assessment score, were removed as were all data for 17 students that identified as American Indian or Pacific Islander. The total number of students for these two racial categories was less than ten respectively and were removed to protect the identity of the students in these categories. Each remaining record included a dummy ID, one of seven racial categories taken from the district student information system, and individual student scores for the assessments used for accelerated program identification: BAS, NNAT3, and STAR Math.

The racial distribution of the 1082 second-grade students was 23% Asian (n = 253), 14% Black (n = 150), 23% Hispanic (n = 251), 26% White (n = 278), and 12% Two or more races (n = 134). Together, this distribution comprises 98% of all second-grade students enrolled in the district in 2019. The demographic distribution of both the second-grade records included in the study and the district as a whole can be seen in Table 1, along with counts and percentages of the population. The second-grade cohort in the study had proportionally fewer Asian students (- 2%) and Hispanic students (- 3%) than the district in 2019. The cohort has proportionally more White students (+ 4%) and students of two or more races (+ 2%) than the district.

Table 1

Racial Demographics of Second Grade Cohort and Overall District Population in 2019

Race	Second Grade		District		Difference
	n	%	*n*	%	%
American Indian	0*	0*	62	0%	0**
Asian	253	23%	3678	25%	- 2%
Black	150	14%	2191	15%	- 1%
Hispanic	251	23%	3854	26%	- 3%
Pacific Islander	0*	0*	167	1%	0**
White	278	26%	3319	22%	4%
Two or More Races	134	12%	1501	10%	2%
Total	1082	98%	14772	100%	

Note. * Reflects the number of students in that category was less than 10 and was removed to protect personally identifiable information. ** Reflects that no comparison between second grade and overall district demographics could be made due to the removal of the second-grade student data in this category.

There were no incomplete free and reduced lunch data for the 1082 records of second grade students. The free and reduced lunch distribution of the second-grade students was 47.9% (*n* = 518) eligible and 52.1% (*n* = 564) not eligible. The distribution of both the second-grade cohort and district counts, and the percentages of free and reduced lunch are represented in Table 2. The proportion of students eligible for free and reduced lunch was comparable (- 2%) to the proportion of eligible students in the entire district in 2019.

Table 2

Free and Reduced Lunch Count of Second Grade Cohort and Overall District Population in 2019

Free and Reduced Lunch	Second Grade		District		Difference
	n	%	n	%	%
Eligible	518	47.9%	7369	49.9%	- 2%
Not eligible	564	52.1%	7403	50.1%	2%

Assessment Information

Three assessments were administered to all 1082 second grade students during the fall of 2019. The scale scores were used in SEM to calculate a regression from each race category and free and reduced lunch status to each assessment. The model used in SEM also measured the significance of the between-group mean differences in the regression.

To ensure the scores are valid and reliable for the second-grade students in the study, assessment results taken by second-grade students in the fall of 2018 were compared to results on the same assessments by second grade students in 2019. Descriptive statistics provide the mean and standard error for the assessment data for two cohorts of second grade students. The standard error is the approximate error from the mean of the whole sample. An individual student score, low to high, for BAS could range between 0 - 28, the score range for the NNAT3 is 1452 - 1689, and the score range for STAR Math is 23 - 731. The BAS scale score means in 2019 was 10.61 with a standard error of 3.39. The NNAT3 scale score means in 2019 was 1591.24 with a standard error of 35.26. The STAR Math scale score means in 2019 was 426.98 with a standard error of 97.21. The scores for all three tests for 2019 and two assessments given to students 2018 are presented in Table 3 for comparison purposes. There were no results for the NNAT3 test for 2018 because the school district first administered the test in 2019.

Table 3

Mean Assessment Score Data for Second Grade Cohort in 2018 and 2019

Test	Mean	Standard Error
	2019	
BAS Scale Sore	10.61	3.39
NNAT3 Scale Score	1591.24	35.26
STAR Math Scale Score	426.98	97.21
	2018	
BAS Scale Sore	10.53	3.47
STAR Math Scale Score	425.71	90.48

Both years have BAS and STAR Math scale score information, and the scores are comparable. The BAS assessment has a mean difference of a 0.08 increase from 2018 to 2019 and, with a range of 28, this difference is reasonable. STAR Math had a mean increase of 1.27 from 2018 to 2019 and, with a range of 708, this difference is also reasonable. Based on a comparison of the two years, the assessment scores for the 2019 cohort in the study are considered to be both valid and reliable.

Research Question Statistics

The following sections present the results of the data analysis for each research question. The use of SEM enabled calculation of the regression coefficients between race and low-income status and all the assessments for questions one and three, as well as the path analysis model created for question two.

Research Question 1

The first research question concerned the significance of between-group differences for low-income students and students of color compared to their peers on the dependent variables. For this analysis, the characteristic values were assigned 0 or 1; for example, Asian students (1) compared to non-Asian students (0). A negative coefficient means students in each category (1) scored lower when compared to other students without that characteristic (0). The inverse of the coefficients that were statistically significant were used to counteract the potential bias as a treatment.

The information obtained from SEM about the significance of these between-group differences in the BAS test is represented in Table 4. The scores were significant at $p < 0.05$. Students who are low-income had the largest coefficient value ($\beta_{FRL} = - 2.15$, $p < .001$) overall. Compared to the other race categories, students identified as Hispanic had the most significant coefficient ($\beta_h = - 2.11$, $p < .001$). The coefficient values for students of color, non-White and non-Asian, were larger than those of their peers. The variance from the mean in scale scores of one standard deviation is represented by the *z*-score. Both Hispanic and low-income students had a larger standard deviation for BAS scores compared to all other groups. All *p*-values were less than 0.05.

Table 4

BAS Results

Race and FRL	Coefficient	Standard Error	*z*	*p*
Asian	-0.62	0.22	-3.80	.005
Black	-0.69	0.27	-2.55	.011
Hispanic	-2.11	0.22	-9.51	< .001

Race and FRL	Coefficient	Standard Error	*z*	*p*
White	-0.55	0.22	-2.58	.010
Two or More Races	-0.70	0.29	-2.45	.014
FRL	-2.15	0.19	-11.43	< .001

Note. FRL is free and reduced lunch which is used as the indicator for low-income. The race categories of American Indian and Pacific Islander were removed so these data were not recorded. Scores were significant at $p < 0.05$.

The results of the regression performed in SEM to measure between-group differences for BAS reject the null hypothesis by race. There are large and significant BAS score mean differences for all racial groups when a single racial group is compared to all other students. Specifically, Hispanic student scores had a large coefficient difference compared to other races, and these scores also showed a wider standard deviation. Students identified as free and reduced lunch also show a significant difference compared to their peers as measured by SEM with $p < 0.05$, and a wide standard deviation. The null hypothesis was rejected for significant mean differences as measured by the BAS for students identified as free and reduced lunch.

The information obtained from SEM about the significance of the between-group differences in the NNAT3 test is represented in Table 5. The scores were significant at $p < 0.05$. Students who are low-income had the largest coefficient (β_{FRL} = - 16.09, $p < .001$) overall. Within the race categories, students identified as Black had the largest coefficient (β_b = - 14.03, $p < .001$) compared to the other race categories. The coefficients for students of color were larger than their peers. The coefficients and standard deviations for Asian and White race categories were smaller than those of their peers and did not exhibit significant *p*-values (i.e., less than 0.05). Both Black and low-income students had a higher standard deviation for NNAT3 scores as compared to all other groups.

Table 5

NNAT3 Results

Race and FRL	Coefficient	Standard Error	*z*	*p*
Asian	3.39	2.38	1.43	.154
Black	-14.03	2.91	-4.82	< .001
Hispanic	-11.13	2.39	-4.67	< .001
White	-1.48	2.30	-0.64	.520
Two or More Races	-11.11	3.06	-3.64	< .001
FRL	-16.09	2.02	-7.99	< .001

Note. FRL is free and reduced lunch which is used as the indicator for low-income. The race categories of American Indian and Pacific Islander were removed so these data were not recorded. Scores were significant at $p < 0.05$.

The results of the regression performed in SEM to measure between group differences for the NNAT3 rejected the null hypothesis by race. Students of color (i.e., non-White and non-Asian) had significant and large mean between-group differences with $p < 0.05$. Students identified as free and reduced lunch also showed a significant and large difference as measured by SEM with $p < 0.05$. The null hypothesis was rejected for significant mean differences as measured by the NNAT3 for students identified as free and reduced lunch.

The information obtained from SEM about the significance of between-group differences in the STAR Math test is presented in Table 6. The results of the analysis were similar to the results for the NNAT3. Students who are low-income had the largest significant coefficient (β_{FRL} = - 52.37, p < .001). Within the race categories, students identified as Hispanic had the largest coefficient (β_h = - 50.36, p < .001) compared to the other race categories followed by students identified as Black and then students identified as Two or more races. The coefficients for White and Asian students were both smaller than their peers and did not have p-values less than 0.05. Both Hispanic and low-income students had a higher standard deviation for STAR Math scores compared to all other groups, and the z-score for White students was the lowest.

Table 6

STAR Math Results

Race and RFL	Coefficient	Standard Error	*z*	*p*
Asian	8.62	6.34	1.36	.174
Black	-34.41	6.36	-4.43	< .001
Hispanic	-50.36	6.36	-7.92	< .001
White	-4.02	6.14	-0.65	.513
Two or More Races	-17.02	8.15	-2.09	.037
FRL	-52.37	5.37	-9.75	< .001

Note. FRL is free and reduced lunch which is used as the indicator for low-income. The race categories of American Indian and Pacific Islander were removed so these data were not recorded. Scores were significant at $p < 0.05$.

The results of the regression performed in SEM to measure between group differences for STAR Math rejected the null hypothesis by race. All three groups of students of color had significant between-group mean differences with $p < 0.05$. Students identified as free and reduced lunch also showed a significant difference as measured by SEM with $p < 0.05$. The null hypothesis was rejected for significant mean differences as measured by the STAR Math for students identified as free and reduced lunch.

Research Question 2

The second research question attempted to identify the significance of between-group differences when compounding the dependent assessment variables, measured within and between, and the independent demographic variables of race and low-income. A SEM path analysis model was created for this analysis but did not allow achievement of goodness of fit through the use of chi-square tests, comparative fit index (CFI), and the root mean square error of approximation (RMSEA).

To determine if a path analysis model created is acceptable, goodness of fit can be measured using statistical tests. Chi-square tests are the most frequently utilized tests in SEM to measure goodness of fit (Jiang et al., 2017). A chi-squared test compares groups in a SEM model to determine population covariance and is generally considered acceptable at < 0.05 significance level. In an article about equivalence testing, Jiang et al. (2017) stated that the rejection of a model (> 0.05) using chi-square tests does not indicate that the model cannot be used. The authors suggested that chi-square tests cannot control type I errors which occur when a false positive result is recorded, and the null hypothesis is incorrectly rejected. Other tests can be used to identify goodness of fit when a type I error occurs. Yuan and Chan (2016) stated that chi-square tests in SEM can be impacted by the number of variables and directions of comparisons, which can limit the chi-square test output and approximations.

Alternative tests that can be conducted to determine goodness of fit can include comparative fit index (CFI) and root mean square error of approximation (RMSEA). CFI assesses the incremental fit of a model and can be used to determine fit of uncorrelated variables (Garrido et al., 2016). Yuan and Chan (2016) suggested that RMSEA can be used in combination with adjusted cutoff values to determine good, fair, or poor fit. The authors stated that a model with poor fit can still be used for approximation, but the model may need to be adjusted for further analysis. If the model has limitations like a limited number of variables or unidirectional or multi-directional relationships, a goodness of fit adjustment for the model may not be possible.

The chi-square goodness of fit test is presented in Table 7. The test included all 1082 records and returned a value of 4558.37 with 18 degrees of freedom. The chi-square test was run through SPSS Amos. Since the returned value was so high, there was likely a type I error, or the variables were too restrictive, and the level of significance is inaccurate. This result can occur when there are a limited number of variables or the relational directions are too restrictive as posited by Yuan and Chan (2016).

Table 7

Chi-Square Goodness of Fit

	Value	df	*p*
Chi-Square	5016.81	18	.000

Because the chi-square test did not provide goodness of fit, the CFI and RMSEA tests were evaluated. CFI has an index range of 0 to 1, where higher values indicate stronger fit (Garrido et al., 2016). RMSEA is a measure of absolute fit and the significance cutoff can be set at any level. Yuan and Chan (2016) suggested multiple cutoffs to identify "good fit" at .01, "fair fit" at .05, "close fit" at .08, and "poor fit" at .10 and above. Table 8 presents the goodness of fit as measured by CFI and RMSEA. Both methods demonstrated poor fit for the path model analyzed. The default model has scores based on the model as it is designed (see Appendix). The independent model scores test the model with the assumption that all the variables are correlated with other variables.

Table 8

Model Fit Summary

	RMSEA	CFI
Default Model	.483	.105
Independence Model	.361	.000

The model produced covariance information between all race categories, each race category and free and reduced lunch, and between each assessment. Failure to reach goodness of fit suggests the data available cannot provide a reliable path model. The data obtained from SEM failed to reject the null hypothesis for research question two.

There are several potential reasons for failure to achieve goodness of fit. Tarka (2018) stated that a limited number of variables could cause errors in the calculations to achieve goodness of fit. The author also suggested that unidirectional and multi-directional relationships may create constraints that limit the model in calculating goodness of fit. The model used for this study had only two independent variables with multiple categories. The limited number of demographic variables available from the district may have impacted the goodness of fit in this model. The direction of the relationships between the low-income and race variables may also have had an impact. No results are presented for research question two as goodness of fit was not achieved, and no results from a path diagram could be shared with confidence to provide a clear answer to research question two.

Research Question 3

The goal of the third research question was to determine if significant between-group mean differences could be used to adjust the outcomes of the measures used for accelerated program identification and increase the proportion of low-income students and students of color that may be identified for accelerated program participation based on these measures. To test the hypotheses associated with this research question the inclusion of regression coefficients from question one that were significant were factored into student scores. The inclusion of coefficients was completed by adding the inverse of the coefficient to each individual original scale score for those groups that had significant mean group differences, indicated by their *p*-values on a given test.

Students' assessment scale scores for all 1802 records of second grade students' assessment scores were ordered, high to low, without the addition of the coefficients and then re-ordered after the addition of the coefficients. A threshold of the 90th percentile was applied to each rank ordering and those at or above the 90th percentile was compared by race and low-income. Students with identical scale scores, before or after coefficients were added, landed in the same percentile rank.

Coefficients for significant mean group differences ($p < 0.05$) for each test were applied to the scores for all students in that group based on race and free and reduced lunch status. The BAS reading test was the only assessment for which every group had coefficients added. The coefficients used with BAS by race were Asian, $\beta_a = - 0.62$, Black, $\beta_b = - 0.69$, Hispanic, $\beta_h = - 2.11$, White, $\beta_w = - 0.55$, and Two or more races, β_{tmr}

= - 0.70, and the coefficient used for low-income was β_{frl} = - 2.15. NNAT3 used the coefficients Black, β_b = - 14.03, Hispanic, β_h = - 11.13, and Two or more races β_{tmr} = - 11.11 for race, and low-income was β_{frl} = - 16.09. The coefficients for STAR Math by race were Black, β_b = - 34.41, Hispanic, β_h = - 50.36, and Two or more races, β_{tmr} = - 17.02, and β_h = - 52.37 was used for low-income. These coefficients are also shown in Table 4, Table 5, and Table 6 respectively.

The number of students who scored at or above the 90th percentile for BAS before and after the coefficients were included and are presented in Table 9 differentiated by race and low-income. The total number of students above the threshold after the coefficient was applied increased. The largest change was for students identified as low-income with 17 additional students above the threshold. Students identified as Black had the largest increase within the race categories with 10 additional students above the threshold. There were also five more Asian and Two or more races students at or above the 90th percentile after the coefficient was applied. The number of White and Hispanic students above the threshold remained the same after the coefficient was applied. There are disproportionate numbers of students before and after the application of coefficients by race compared to the total population. When calculating the difference in what would be proportionate representation for students of color, Black students are closer to being proportionately represented while Hispanic students are not.

Table 9

Number of students of second grade cohort at or above the 90th percentile on BAS

Race and FRL	Total Students in Cohort	Percent of Total Students in Cohort	Student n Before Treatment (A)	Percent Student n Before Treatment (B)	Student n After Treatment (C)	Percent Student n After Treatment (D)	Difference Before and After Treatment (C-A)	Proportionate n of Students for A (E)	Disproportionality of Student n Before Treatment (A-E)	Proportionate n of Students for C (G)	Disproportionality of Student n After Treatment H (C-G)
Asian	264	23%	31	31%	36	31%	5	22	9	26	10
Black	167	14%	5	5%	15	13%	10	14	-9	17	-2
Hispanic	270	23%	7	7%	7	6%	0	23	-16	27	-20
White	295	25%	41	41%	41	35%	0	25	16	29	12
Two or More Races	151	13%	14	14%	16	14%	2	13	1	15	1
Other Races	17	1%	1	1%	1	1%	0	1	0	2	-1
FRL	578	41%	17	21%	34	31%	17	33	-16	45	-11

Note. Column A represents the number of students at or above 90th percentile before treatment. Column B represents the percent of students at or above 90th percentile before treatment. Column C represents the number of students at or above 90th percentile after treatment. Column D represents the percent of students at or above 90th percentile after treatment. Column E represents the total number of students at or above 90th percentile before treatment when proportionate to cohort. Column F represents the difference of number of students at or above 90th percentile before treatment to proportionate number. Column G represents the total number of students at or above 90th percentile after treatment when proportionate to cohort. Column H represents the difference of number of students at or above 90th percentile after treatment to proportionate number.

A Pearson's chi-square test was used to determine if the race and low-income distribution of students in the entire 2019 second grade cohort was the same in terms of the distribution of students at or above the 90th percentile by race and income. A comparison of total cohort population by race and income with those at or above the 90th percentile can help determine if the student scores with and without the coefficient treatment are similar. A p-value of < 0.05 meant that the difference in the ranking of students in relation to their race or socioeconomic status was statistically significant. A p-value > 0.05 meant that the distribution was proportional to the overall population.

The results of the chi-square test for the BAS assessment are presented in Table 10, including the degrees of freedom and p-values. The distribution of students by race and socioeconomic status who scored above the threshold without the application of the coefficient was compared to the distribution of the total number of students who took the test and the p-values were < 0.05. The comparison of students above the threshold with and without the coefficient also returned *p*-values < 0.05 for race and low-income. Lastly, the number of students who scored above the threshold was compared to the total number of students by race and socioeconomic status when the coefficients were applied; *p*-values were also < 0.05. This means the distribution of the students by race and income at or above the 90th percentile for their BAS scores was not representative of the population as a whole, with or without the addition of the coefficients. The distribution of students above the threshold was also not representative of the total population after the coefficient was applied because the *p*-values were < 0.05. The null hypothesis is rejected because the proportion of students are not similarly distributed by race or low-income for the BAS test when ranked either with or without the coefficient added for groups that had significant mean differences.

Table 10

Chi-square test of BAS

	χ2s	Df	*P*
Comparing 90th percentile and population without coefficient			
Race	28.60	5	< .001
Income	38.64	1	< .001
Comparing 90th percentile groups only with and without coefficient			
Race	448.88	5	< .001
Income	449.44	1	< .001
Comparing 90th percentile and population with coefficient			
Race	27.74	5	< .001
Income	22.62	1	< .001

Table 11 represents the results for NNAT3 before and after the coefficient was applied to student scale scores. The NNAT3 had the largest increase in the number of low-income students that met the threshold of at or above the 90th percentile with an increase of 43 students. The largest increase by race was for Hispanics with an increase of 20 students above the threshold, followed closely by Black students with an increase of 18. Students identified as Two or more races above the threshold increased by 8. Both the number of Asian and White students above the threshold decreased after the coefficient treatment was applied by 14 and 11 students, respectively. There are disproportionate numbers of students before and after the application of coefficients by race compared to the total population. When calculating the difference in what would be proportionate representation for students of color, Black and Hispanic students are slightly overrepresented. White and Asian students are slightly underrepresented.

Table 11

Number of students of second grade cohort in 90th percentile on NNAT3

Race and FRL	Total Students in Cohort	Percent of Total Students in Cohort	Student n Before Treatment (A)	Percent Student n Before Treatment (B)	Student n After Treatment (C)	Percent Student n After Treatment (D)	Difference Before and After Treatment (C-A)	Proportionate n of Students for A (E)	Disproportionality of Student n Before Treatment (A-E)	Proportionate n of Students for C (G)	Disproportionality of Student n After Treatment H (C-G)
Asian	261	23%	34	36%	20	17%	-14	21	13	26	-6
Black	165	14%	2	2%	20	17%	18	14	-12	17	3
Hispanic	270	24%	14	15%	34	29%	20	22	-8	27	7
White	294	26%	35	37%	24	21%	-11	24	11	30	-6
Two or More Races	141	12%	8	9%	16	14%	8	12	-4	14	2
Other Races	16	1%	1	1%	2	2%	1	1	0	2	0
FRL	568	40%	17	21%	60	44%	43	33	-16	55	5

Note. Column A represents the number of students at or above 90th percentile before treatment. Column B represents the percent of students at or above 90th percentile before treatment. Column C represents the number of students at or above 90th percentile after treatment. Column D represents the percent of students at or above 90th percentile after treatment. Column E represents the total number of students at or above 90th percentile before treatment when proportionate to cohort. Column F represents the difference of number of students at or above 90th percentile before treatment to proportionate number. Column G represents the total number of students at or above 90th percentile after treatment when proportionate to cohort. Column H represents the difference of number of students at or above 90th percentile after treatment to proportionate number.

The results of the chi-square test for the NNAT3, with degrees of freedom and *p*-values, are presented in Table 12. Students who scored above the threshold without the application of the coefficient treatment were compared by race and socioeconomic status to the total number of students who took the test; *p*-values were < 0.05. The comparison of students above the threshold with and without the coefficient treatment applied also returned *p*-values < 0.05 by race and low-income. Lastly, the students who scored at or above the 90th percentile were compared to the total number of students by race and socioeconomic status with the application of the coefficient; *p*-values were also < 0.05. Ultimately, the null hypothesis was rejected because the proportion of students at or above the 90th percentile is not similarly distributed by race and income for the NNAT3 test to the population as a whole when ranked either with or without the coefficient applied.

Table 12

Chi-square test of NNAT3

	χ2s	df	*P*
	Comparing 90th percentile and population without coefficient		
Race	26.20	5	< .001
Income	33.60	1	< .001
	Comparing 90th percentile only with and without coefficient		
Race	524.00	5	< .001
Income	510.76	1	< .001
	Comparing 90th percentile and population with coefficient		
Race	11.99	5	.035
Income	4.80	1	.028

As presented in Table 13, the STAR Math results for students in the 90th percentile before and after the coefficient were applied by race and low-income. The largest increase in the number of students that met the threshold of at or above the 90th percentile was for low-income students with an increase of 29 students. Based on race, Hispanic students had the largest increase with 15 additional students, followed by Black students with an increase of 11. After the application of the coefficient treatment for STAR Math, the number of White and Asian students who met the threshold decreased by 14 and 13 students, respectively. Students from the Two or more races category had the smallest change with one additional student at or above the 90th percentile. There are disproportionate numbers of students before and after the application of coefficients by race compared to the total population. When calculating the difference in what would be proportionate representation for students, Black and Asian students are slightly overrepresented. White and Hispanic students are slightly underrepresented.

Table 13

Number of students of second grade cohort in 90th percentile on STAR Math

Race and FRL	Total Students in Cohort	Percent of Total Students in Cohort	Student n Before Treatment (A)	Percent Student n Before Treatment (B)	Student n After Treatment (C)	Percent Student n After Treatment (D)	Difference Before and After Treatment (C-A)	Proportionate n of Students for A (E)	Disproportionality of Student n Before Treatment (A-E)	Proportionate n of Students for C (G)	Disproportionality of Student n After Treatment H (C-G)
Asian	255	23%	43	40%	30	28%	-13	25	18	25	5
Black	151	14%	6	6%	17	16%	11	15	-9	15	2
Hispanic	258	24%	9	8%	24	22%	15	25	-16	25	-1
White	279	25%	37	35%	23	21%	-14	27	10	27	-4
Two or More Races	137	12%	12	11%	13	12%	1	13	-1	13	0
Other Races	17	2%	0	0%	0	0%	0	2	-2	2	-2
FRL	527	38%	21	25%	50	39%	29	33	-12	49	1

Note. Column A represents the number of students at or above 90th percentile before treatment. Column B represents the percent of students at or above 90th percentile before treatment. Column C represents the number of students at or above 90th percentile after treatment. Column D represents the percent of students at or above 90th percentile after treatment. Column E represents the total number of students at or above 90th percentile before treatment when proportionate to cohort. Column F represents the difference of number of students at or above 90th percentile before treatment to proportionate number. Column G represents the total number of students at or above 90th percentile after treatment when proportionate to cohort. Column H represents the difference of number of students at or above 90th percentile after treatment to proportionate number.

The results of the chi-square test for the STAR Math test are presented in Table 14 with degrees of freedom and *p*-values. When the demographic breakdown of students who scored at or above the 90^{th} percentile without the application of the coefficient treatment was compared to the total number of students who took the test by race and socioeconomic status, *p*-values were < 0.05. A comparison of students above the threshold with and without the coefficient also returned *p*-values < 0.05. The distribution of students at or above the 90^{th} percentile with the coefficient treatment applied compared to the total population returned a *p*-value of .851 for race and .651 for low-income. Both these *p*-values were > 0.05. This meant the distribution of the students when the coefficient treatment was applied for STAR Math was representative of the distribution for the overall population. The null hypothesis cannot be rejected meaning the proportion of the students are similarly distributed for the STAR Math test when ranked with the coefficient treatment applied. However, the distribution of students by race and income are different when the coefficient is not applied for students above the threshold compared to the entire population and for students at or above the 90th percentile before and after the coefficient was applied.

Table 14

Chi-square test of STAR Math

	χ2s	df	*p*
Comparing 90th percentile and population without coefficient			
Race	29.48	5	< .001
Income	32.52	1	< .001
Comparing 90th percentile only with and without coefficient			
Race	497.57	5	< .001
Income	495.06	1	< .001
Comparing 90th percentile and population with coefficient			
Race	1.99	5	.851
Income	.20	1	.651

Summary

The purpose of this quantitative casual-comparative study was to measure the between-group mean differences of assessment data for low-income students and students of color. SEM was used to measure the between-group mean differences for low-income students and students of color on three assessments: BAS, NNAT3, and STAR Math. Significant mean differences were identified on all three assessments for low-income students and students of color. SEM was unable to achieve goodness of fit for a path analysis model to measure the covariance between and amongst variables, so the researcher was unable to test the hypothesis for research question two. The statistically significant coefficients obtained from SEM in research question one was used to determine if their addition could impact the demographic distribution of students at or above the 90^{th} percentile for all three tests. From the results of the chi-square tests the null hypothesis is rejected for BAS and the NNAT3. The students are not similarly distributed before and after the coefficients are applied to the scale scores of student groups with significant mean differences for these two tests. However, from the results of the chi-square test for STAR Math the null hypothesis cannot be rejected in one of the three tests; the distribution of the demographic make-up of students at or above the 90^{th} percentile was similar to the distribution of the entire population after the coefficients were applied to the scale scores of student groups with significant mean differences.

CHAPTER 5: CONCLUSIONS AND DISCUSSION

Historically underrepresented subgroups of students are not identified for accelerated program participation in the same proportions as their peers. As a result, low-income students and students of color do not have the same access to accelerated programs in K-12 education. The exclusion of these students can lead to social inequities and have economic impacts beyond K-12 education (Cao et al., 2017). School districts who serve students in K-12 education often identify students for participation in accelerated programs through universal screeners, the use of multiple measures, and the scores from seemingly objective assessments. The use of these methods can have negative impacts due to bias, both implicit and explicit, and continue the disproportionate identification of low-income students and students of color compared to their peers.

The purpose of this quantitative casual-comparative study was to measure the between-group mean differences of assessment data for low-income students and students of color that may contribute to the disproportionate identification of second grade students for accelerated programs in a K-12 school district in Washington state. Using assessment tools and practices already in place in a school district, between-group mean differences could be identified and evaluated for bias. The use of SEM for data analysis provided an opportunity to examine the between-group differences and relationships between and within dependent and independent variables. The results of the analysis and actions can be utilized by leaders as they attempt to address the latent bias discussed in this chapter.

First, a discussion and application of the results along with conclusions to the problem statement are shared. Next, application of the results from this study and recommendations for action are discussed. The chapter concludes with recommendations for further research and a concluding statement.

Discussion of Results and Conclusions

The results of this study included between-group mean differences for student assessment scores on three academic and cognitive tests used by the school district for accelerated program identification. The self-identified racial category recorded by the school district, and low-income status as identified by free and reduced lunch, were used as the independent variables to compare between-group differences. The following research questions and hypotheses were used to analyze between-group differences of assessment results used for accelerated program identification.

1. What is the difference in assessment results for low-income students and students of color compared to their peers at a given grade level?

H_0 – There is no significant mean difference in assessment results for low-income students and students of color compared to their peers at a single grade level.

H_1 – There is a significant mean difference in assessment results for low-income students compared to their non-low-income peers at a single grade level.

H_2 – There is a significant mean difference in assessment results for students of color compared to their peers at a single grade level.

2. What is the significance of between-group differences when compounding dependent variables measured within and between independent variables?

H_0 – There is no significant mean difference in the relationships between dependent variables and independent variables when compounded.

H_1 – There is a significant mean difference in the relationships between dependent and independent variables when compounded.

3. Can assessment results used to identify students for accelerated programs be adjusted based on between-group differences to affect the identification of low-income students and students of color for accelerated programs?

H_0 – Adjustments based on between-group significant mean differences do not increase the identification of low-income students and students of color.

H_1 – Adjustments based on between-group significant mean differences increase the identification of low-income students for accelerated programs.

H_2 – Adjustments based on between-group significant mean differences increase the identification of students of color for accelerated programs.

The analysis found statistically significant between-group mean differences for assessment results by race and low-income status in answering research questions one and three. Latent bias in the assessments used by the district can impact the number and proportion of students identified for accelerated programs. The data in the study were insufficient to create a path model to use in SEM to answer research question two. The student demographic data available from the district could not create a valid path analysis model to fully test the hypothesis of between-group differences for dependent and independent variables for research question two.

Research Question 1 Conclusions

To understand the differences in assessment results for low-income students and students of color compared to their peers, SEM was used to perform a regression for between-group differences for each race category and low-income. The regression provided a coefficient and *p*-value for each group of students. The *p*-value was used to determine the significance of between-group mean scores with $p < 0.05$ considered significant. All three assessments provided evidence of statistically significant differences for students of color and low-income students.

BAS results

For the BAS assessment, the null hypothesis was rejected for both low-income students and students of color. All mean score differences were significant at $p < 0.05$. There were significant between-group differences for students by income and race when teachers administered the BAS. Some latent bias in the assessment or its administration resulted in lower average scale scores for low-income students and students of color for this assessment which is used for accelerated program identification in the K-12 school district.

NNAT3 results

The NNAT3 assesses student cognitive levels, and the null hypothesis was rejected for both low-income students and students of color. When compared to students not in the same category, mean score differences for students identified as low-income, Black, Hispanic, and Two or more races all returned *p*-values < 0.05; their NNAT3 scores were significantly lower than other groups. The mean differences for the scores of students identified as Asian and White were shown not to be significant. A latent bias in the NNAT3 may impact student identification for accelerated programs for low-income students and students of color.

STAR Math results

Th STAR Math results were the basis of rejecting the null hypothesis for both low-income students and students of color. *P*-values were < 0.05 for low-income, Black, Hispanic, and Two or more races. The between-group mean score differences were significant, and the coefficients were all negative. The mean score differences of Asian and White students were not considered significant. As with the other assessments used by the K-12 district, latent bias in STAR Math scores can impact the proportion of historically underrepresented students when used for accelerated program identification.

Research Question 2 Conclusions

To answer research question two, a path analysis model was created to better understand the size and relationship of between-group differences for the dependent and independent variables. Through the use of SEM, a path analysis can help in understanding how dependent and independent variables interact with one another. The goal of research question two was to understand if there were significant between-group differences between and amongst variables and what the differences were when variables were combined.

A valid path analysis model could not be achieved to effectively measure and evaluate the hypothesis for research question two. The values returned by the chi-square goodness of fit test conducted by SEM returned a value that was too high for the model to be considered. A review of RMSEA and CFI scores also returned values that confirmed the model did not meet the goodness of fit criteria. The inability to achieve goodness of fit may be due to a limited number of variables in the model which caused errors to occur in the calculations. While there were five race categories within the race variable measured in the model, there were only two total independent variables, which limited the possibility of removing or modifying variable relationships to achieve goodness of fit. There were no other data related to students' race or socioeconomic status available from the district to add to the model.

Without goodness of fit, the hypothesis was unable to be tested. Between-group differences were not evaluated when measuring the relationships between and within variables individually and in combination. This does not mean that significant differences did not exist, rather that more variables were needed to create a valid path analysis model for the relationship between race, socioeconomic status and the assessment results.

Research Question 3 Conclusions

The statistically significant scale score coefficients obtained in research question one was used to answer research question three, which focused on the application of the coefficients to individual student scores for each assessment. Research question three attempted to identify differences in the rank ordering of students after scale score adjustments were made using the significant coefficient values for low-income students and students of color in each assessment. A comparison of the demographic breakdown by race and low-income status of students scoring at or above a threshold of the 90^{th} percentile of their cohort was conducted for each test to see the direct impacts of the coefficient treatment. A second step was then completed using a chi-square test of homogeneity to measure whether any differences in the demographic make-up for those at or above the 90^{th} percentile were significant when compared to the second-grade cohort as a whole, before and after assessment scale scores were adjusted. The chi-square test was also used to measure the distribution of race and low-income students above the threshold for each assessment with and without the coefficient.

BAS results

The application of the scale score coefficient for the BAS reading test increased the number and proportion of low-income and students of color. It doubled the number of low-income students at or above the 90th percentile, from 17 to 34. The chi-square test showed that the demographic distribution of students above the threshold did not match the overall student population with or without the coefficient adjustments. Adjusting low-income student's scores showed positive increases in the number of low-income students at or above the 90th percentile.

The outcomes of the BAS coefficient treatment for students by race also had positive increases for non-White and non-Asian students. Ten more Black students scored above the threshold, and the Two or more races group increased by two students. Using statistically significant scale score differences to adjust student's scores by race increased the number of students of color at or above the 90th percentile and did not reduce the number of students in other race categories.

NNAT3 results

The NNAT3 test also evidenced changes in the demographic data by income when the coefficient was applied to students' scale scores. The number of students eligible for free and reduced lunch scoring at or above the 90th percentile increased by 43 students. The chi-square test that compared the income breakdown of the top scoring students to the total cohort population after the coefficient was added returned a *p*-value of .028, which is still statistically significant at < 0.05 but not as low as *p* value found for the difference before the coefficient was added.

Changes in the number and proportion of students by race for the NNAT3 test also occurred when the coefficient was added. The number of Black students above the threshold increased by 18 students, and the number of Hispanic students increased by 20. The number and proportion of White and Asian students above the threshold dropped. After using the coefficient, there was an equal number of Asian and Black students in the top scoring group and there were more Hispanic students than Asian or White students. The inclusion of the coefficient increased the total number of non-White and non-Asian students and made a statistically significant difference in the distributions of students by racial demographics above the threshold as well as between students at or above the 90th percentile to the full cohort.

STAR Math

The application of the coefficient treatment to the STAR Math test produced a large increase in the number of low-income students at or above the 90th percentile. The number of low-income students more than doubled from 21 to 50. The chi-square test comparing the distribution of low-income students in the top scoring group to the total population after the coefficient was applied returned a p-value of .651 failing to reject the null hypothesis. The distribution of low-income students was similar for the second-grade students after the coefficient was added, but not before.

Students of color also increased in the top scoring student group with the coefficient treatment application to the STAR Math test. With an increase of 15, the number of students above the threshold that identified as Hispanic more than doubled, and the number of students identified as Black increased by 11. The chi-square test after the coefficient was applied returned a p-value of .851 when comparing the racial make-up of students at or above the 90th percentile and the overall cohort. Similar to the results of the distribution of low-income students for the STAR Math scores, the difference in the distribution of racial categories was statistically significant between students at or above the 90th percentile before and after the coefficients are applied, but the difference between the top scoring group and the total student population after the coefficients were applied was not statistically different.

In all cases, using scale score coefficients to counteract latent bias in assessments used for accelerated program identification in the K-12 district reduced the disproportionality of students scoring at or above the 90th percentile by race and by low-income status. In most cases this change increased the proportion of students of color and low-income students in the top scoring group than would have otherwise been ranked or distributed while also increasing the total number of students in the top scoring group. In most cases, the number of White and Asian students scoring in the top group fell but the distribution of races became more similar to the racial distribution of all students in the second-grade cohort by assessment scale scores. The treatment had a net positive effect on low-income students and students of color except for BAS. Before the application of the coefficient treatment, White and Asian students were overrepresented in all three tests while low-income, Black, and Hispanic students were underrepresented. This was no longer the case after the treatment.

The results from this study show that there are significant mean between-group differences for low-income students and students of color for academic and cognitive assessments used for accelerated program identification. These differences could impact the identification of historically underrepresented students for accelerated programs. The results for research question three demonstrate that an application of coefficients to individual test scores that have significant mean difference can increase the number of low-income students and students of color included in the top scoring groups, both academically and cognitively. The analysis of data has revealed that there is latent bias in assessments used for accelerated program identification that affect the proportion of low-income students and students of color compared to the total population. Statistically validated adjustments can be made to reduce the disproportionality in assessment outcomes for historically underrepresented groups. However, there were not enough data found in the study to identify how differences in race and socioeconomic status interact to create these differences.

Application of Results and Conclusions to the Problem Statement

The goal of this study was to identify potential bias in measures commonly used in a K-12 school district, thus reducing the disproportionate identification of low-income students and students of color for accelerated programs. Hertzog (2017) suggested that the use of multiple measures can compensate for unintended bias in academic assessments but did not provide direct evidence supporting that claim. Cao et al. (2017) stated that multiple measures could be used, but no researchers have provided a potential method for using those measures in combination to reduce the disparities in identification. The results from this study suggest that there is bias in the assessments used by a K-12 school district, and that the application of a statistical treatment may decrease disproportionate identification of students for accelerated programs.

The finding of significant between-group differences for the three assessments used in the district demonstrates that there is latent bias in the measures used for accelerated program identification. This means that a K-12 school district should review the assessments they use for accelerated program identification; the district can then counteract the bias by identifying statistically significant between group differences and adjusting how they calculate student scores for identification accordingly. Results from this study showed that when the significant mean scale score difference coefficients were used to interrupt the unintended bias, more low-income students and students of color achieved scores at or above the 90th percentile in all three assessments, one cognitive and two academic. Students in the top scoring groups, such as those at or above the 90th percentile, are often considered for accelerated program participation because their academic and cognitive scores stand out above their peers.

Scoring in the 90th percentile does not mean students will be identified to participate in accelerated programs, but these students can be considered for participation because they may benefit from the program or services available. The inclusion of more low-income students and students of color could improve their social standing, as suggested by the cultural mobility theory (DiMaggio, 1982). The improvement of social standing through inclusion in gifted academic programs could increase a student's cultural capital in the future, as suggested by Jæger (2011). School districts' policies and practices in the use of assessment data can disrupt disproportionate identification of accelerated programs and may improve cultural capital for low-income students and students of color if they are identified for participation in accelerated programs.

While significant between-group differences were not found for all groups on all tests used by the district, when statistically significant coefficients were applied, there was an increase in the number of low-income students and students of color at or above the 90th percentile of the second grade cohort. This increase in underrepresented students at the top of the ranking suggests that there are statistically valid ways to rectify the disproportionate numbers of students in accelerated programs.

The information obtained from this study contributes to prior research on universal screening for accelerated program identification and the use of multiple measures. Hodges et al. (2018) found that universal screening alone did not identify more low-income students or students of color. Through universal screening, a school district could use multiple measures as Herzog (2017) recommended. The assessments analyzed for this study were conducted as part of the K-12 school districts' universal screening process. The district used multiple measures for accelerated program identification. All students in grade two in the district used in this study were administered all three assessments as part of the accelerated program identification process, one cognitive and two academic. Other districts that use multiple measures in a universal screening process could analyze the results to identify any between-group differences for each assessment. A statistical treatment could be applied to the assessment scores that had significant mean group score differences. The addition of significant coefficients may identify more low-income students and students of color for consideration in accelerated program participation reducing disproportionate identification of students by race and socioeconomic status.

Application to Leadership

Leaders can use the information and results from this research study to create or improve school district policies and practices. Michelmore and Dynarski (2017) posited that school district policies and practices can contribute to the continuation of school and social inequities. To positively impact these inequities, there are three actions that school leaders can take: review accelerated program identification policy with a goal of reducing disparities between racial and socioeconomic groups, change accelerated program identification practices to counteract biased evaluation of student ability, and continuously monitor outcomes for low-income students and students of color in accelerated programs.

Policy at both the state and local levels can change to positively impact accelerated program identification practices. State and district leaders can limit unintended bias through the adoption of policies that explicitly require implementing methods such as the use of multiple measures and universal screening for identification. Leaders can analyze their methods and measures to identify potential sources of bias and take steps to mitigate all identified biases. The information from this study suggests that a statistical method can be used to identify between-group differences in commonly used assessments, and the identified differences can be used to adjust how students are ranked from high to low based on assessment scores in identification processes.

One practice that can be implemented is to continuously monitor the demographic distribution of students participating in accelerated programs. This practice can inform leaders about who is being identified for accelerated programs and help in decision making as to when a review of identification steps is warranted. Once students are identified for accelerated programs, leaders can continuously monitor outcomes for students participating in the programs. This information can be beneficial in understanding who is benefiting from services. Data from student performance in accelerated programs may be monitored to determine who benefits the most from services by reviewing pre and post data outcomes from local and state assessments. Information about student pathways beyond high school can also be monitored to determine if the cultural capital of historically underserved groups is improving. The increased identification of low-income students and students of color in accelerated programs may be beneficial in increasing cultural mobility, but continued monitoring can help inform how to improve upon the services provided or identification methods used in the future once students are identified.

Recommendations for Action

There are actions that K-12 school districts can take based on the results from this research study. School district leaders may want to understand any unintended bias that potentially impact the results of the measures used in accelerated program identification. Based on that knowledge, adjustments can be made to mitigate the bias so that low-income students and students of color are equitably considered for accelerated program identification and participation.

School districts can implement universal screening to ensure all students are considered for accelerated program participation. Identification processes that screen a group of students can utilize multiple measures to identify students who can benefit from accelerated program goals. Carroll's three-stratum theory suggested that students can perform well on a single measure and no single measure can capture how students will perform in all measurable areas of interest (as cited by Beaujean, 2015). The K-12 school district's data reviewed for this study identified three assessments that aligned to their accelerated program goals and measured multiple dimensions of student performance.

If a school district implements universal screening and identifies multiple measures to use for accelerated program identification, an analysis of the data to measure for bias can be conducted. That analysis could be completed using various statistical tools, not just through SEM. The use of a one-way ANOVA could identify the significance of bias in assessment. A linear regression could support the identification of statistically significant coefficients that may be applied to student scores similar to the way they were used in this study. The use of statistical significance could be applied to identify more low-income students and students of color that may represent a similar population distribution to the total population used during universal screening.

Finally, school district leaders can continue to monitor student demographics and accelerated program participation. Continuous monitoring can support equitable identification practices and determine how students are performing once they are in the program. The information obtained from student performance after participation in accelerated programs can help inform district staff on the benefits of identifying not only high-performing students, but also those who can benefit from services while in the program. Continuous monitoring of identification demographic distributions, student growth while in the program, and the bias of the assessments used for identification could help district leaders in continued improvement in equitable identification practices for accelerated programs.

Recommendations for Further Research

Future research could be conducted that would extend the findings from this study. The path analysis model developed for research question two could be further developed to include more variables, and a more in-depth investigation of the relationship of low-income status and race on test scores could be conducted. Future research could be undertaken in a district that has larger American Indian and Pacific Islander student populations. Other grade levels could be evaluated to determine if significant mean differences are present with older or younger students, or if latent bias increases or decreases at different grade levels. Additionally, other cognitive or academic assessments used for accelerated program identification could be analyzed in a similar manner. Based on the results of this study, these topics could benefit from future research.

The path analysis graphic model created for this study in SEM was limited by the number of variables. Goodness of fit could not be obtained because of the limited variables and the inability to modify the model to fully analyze the covariance of variables. There are other variables that could be included for further analysis, such as English language learner status, gender, teacher experience, early learning opportunities, and teacher perceptions about accelerated programs. There are more data variables that could be considered including student grades, attendance, and other forms of assessment that can measure student performance. The inclusion of these variables may allow for model fit to be acceptable and the significance of between-group differences to be measured.

Throughout K-12 education, there are assessments given to students that could limit a students' ability to increase their cultural capital. The between-group differences were significant for low-income students and students of color in almost all the statistical tests in this study. Further research on assessment tools could be conducted to determine if there are significant between-group differences for other assessment tools and the impacts of these differences on disproportionate representation of historically underrepresented students.

Concluding Statement

A quantitative research model was able to identify the significance of between-group differences in assessments used for accelerated program identification in a K-12 school district. The results of this study demonstrate that a statistical analysis can be used to identify unintentional bias in assessments and the identified differences can be used to mitigate the bias. For many years, historically underrepresented subgroups of students have been identified at lower rates for accelerated programs compared to their peers. The implementation of statistical methods can address this problem, and the application of these methods may benefit low-income students and students of color beyond their K-12 educational experience.

Accessing accelerated programs can provide opportunities for low-income students and students of color to increase their cultural capital, as well as offering instruction and programs in which they can accelerate their learning and experience new learning opportunities. The implementation of universal screening and the use of multiple measures for program identification can ensure that all students are equitably evaluated for program participation. With the inclusion of a statistical method to limit the bias from the multiple measures, more low-income students and students of color may access and benefit from a K-12 school district's accelerated program goals.

REFERENCES

Alessandri, G., Zuffianò, A., & Perinelli, E. (2017). Evaluating intervention programs with a pretest-posttest design: A structural equation modeling approach. *Frontiers in Psychology, 8,* 1-12. https://doi.org/10.3389/fpsyg.2017.00223

Apuke, O. D. (2017). Quantitative research methods: A synopsis approach. *Kuwait Chapter of the Arabian Journal of Business and Management Review, 6*(11), 40-47. https://doi.org/10.12816/0040336

Beaujean, A. (2015). John Carroll's views on intelligence: Bi-factor vs. higher-order models. *Journal of Intelligence, 3*(4), 121-136. https://doi.org/10.3390/jintelligence3040121

Benson, N., Beaujean, A., McGill, R., & Dombrowski, S. (2018). Revisiting Carroll's survey of factor-analytic studies: Implications for the clinical assessment of intelligence. *Psychological Assessment, 30*(8), 1028-1038. https://doi.org.10.1037/pas0000556

Bernstein, J., Cheng, J., Balogh, J., & Rosenfeld, E. (2017). Studies of a self-administered oral reading assessment. *7th ISCA Workshop on Speech and Language Technology in Education*, 172-176. https://doi.org/10.21437/SLaTE.2017-30

Bollen, K. A., & Noble, M. D. (2011). Structural equation models and the quantification of behavior. *Proceedings of the National Academy of Sciences, 108*(Supplement_3), 15639-15646. https://doi.org/10.1073/pnas.1010661108

Borgianni, Y., & Maccioni, L. (2020). Review of the use of neurophysiological and biometric measures in experimental design research. *Artificial Intelligence for*

Engineering Design, Analysis and Manufacturing, 34(2), 248-285. https://doi.org/10.1017/S0890060420000062

Bradbury, B., & Corak, M. (2015). *Too many children left behind: The U.S. achievement gap in comparative perspective*. Russell Sage Foundation.

Breitsohl, H. (2019). Beyond anova: An introduction to structural equation models for experimental designs. *Organizational Research Methods, 22*(3), 649-677. https://doi.org/10.1177/1094428118754988

Brown, R. S., Gallagher-Mackay, K., & Parekh, G. (2020). Redefining risk: Human rights and elementary school factors predicting post-secondary access. *Education Policy Analysis Archives, 28*(21), 1-21. doi:10.14507/epaa.28.4200

Cao, T. H., Jung, J. Y., & Lee, J. (2017). Assessment in gifted education: A review of the literature from 2005 to 2016. *Journal of Advanced Academics, 28*(3), 163-203. https://doi:10.1177/1932202X17714572

Card, D., & Giuliano, L. (2016). Universal screening increases the representation of low-income and minority students in gifted education. *Proceedings of the National Academy of Sciences, 113*(48), 13678-13683. https://doi.org/10.1073/pnas.1605043113

Cincinnato, S., De Wever, B., Van Keer, H., & Valcke, M. (2016). The influence of social background on participation in adult education: Applying the cultural capital framework. *Adult Education Quarterly, 66*(2), 143-168. https://doi.org/10.1177/0741713615626714

Colangelo, N., Assouline, S. G., & Gross, M. U. M. (2004). A nation deceived: How schools hold back America's brightest students. *Gifted Child Today Magazine., 28*(1). http://www.accelerationinstitute.org/nation_deceived/nd_v1.pdf

Coleman, J. (1968). The concept of equality of educational opportunity. *Harvard Educational Review, 38*(1), 7-22. https://doi.org/10.17763/haer.38.1.m3770776577415m2

Copur-Gencturk, Y., Cimpian, J. R., Lubienski, S. T., & Thacker, I. (2019). Teachers' bias against the mathematical ability of female, black, and Hispanic students. *Educational Researcher, 49*(1), 1-47. https://doi:10.3102/0013189X19890577

Crabtree, L. M., Richardson, S. C., & Lewis, C. W. (2019). The gifted gap, stem education, and economic immobility. *Journal of Advanced Academics, 30*(2), 203-231. https://doi:10.1177/1932202X19829749

Crockett, S. A. (2012). A five-step guide to conducting sem analysis in counseling research. *Counseling Outcome Research and Evaluation, 3*(1), 30-47. https://doi.org/10.1177/2150137811434142

Crossman, A. (2020). *Structural Equation Modeling*. Thought Co. https://www.thoughtco.com/structural-equation-modeling-3026709

DiMaggio, P. (1982). Cultural capital and school success: The impact of status culture participation on the grades of U.S. high school students. *American Sociological Review, 47*(2), 189-202. https://doi.org/10.2307/2094962

Dougherty, J. (2014). *From '88 to the new millennium: The rise and decline of gifted and talented programs*. Trinity College.

http://commons.trincoll.edu/edreform/2014/05/from-88-to-the-new-millennium-the-of-gifted-and-talented-programs/

Ecker-Lyster, M., & Niileksela, C. (2017). Enhancing gifted education for underrepresented students: Promising recruitment and programming strategies. *Journal for the Education of the Gifted, 40*(1), 79-95. https://doi.org/10.1177/0162353216686216

Emerson, R. W. (2017). Anova and t-tests. *Journal of Visual Impairment & Blindness, 111*(2), 193-196. https://doi.org/10.1177/0145482X1711100214

Farley-Ripple, E. N., Jennings, A. S., & Buttram, J. (2019). Toward a framework for classifying teachers' use of assessment data. *AERA Open, 5*(4), 4-18. https://doi.org/10.1177/2332858419883571

Ferrando, M., Ferrándiz, C., Llor, L., & Sainz, M. (2016). Successful intelligence and giftedness: An empirical study. *Anales de Psicología, 32*(3), 672-682. https://doi.org/10.6018/analesps.32.3.259431

Flores, S. M., Park, T. J., & Baker, D. J. (2017). The racial college completion gap: Evidence from Texas. *The Journal of Higher Education, 88*(6), 894-921. https://doi:10.1080/00221546.2017.1291259

Fountas, I. C. Pinnell, G. S. (2012). *The Fountas & Pinnell benchmark assessment system* (2nd ed.). Heinemann.

Fountas, I. C., Pinnell, G. S., & Pearson Australia. (2017). *The Fountas & Pinnell benchmark assessment system 1*. Heinemann

Freeman, A. J., & Chen, Y.-L. (2019). Interpreting pediatric intelligence tests: A framework from evidence-based medicine. *Handbook of Psychological*

Assessment (pp. 65-101). Elsevier. https://doi.org/10.1016/B978-0-12-802203-0.00003-1

Gajo, E., Oberwetter, J., Mathew, M., Dam, M., Sanborn, T., & Chehab, L. G. (2019). Correlation of sugar-sweetened beverage consumption and school free and reduced lunch eligibility as a measure of socioeconomic status. *Journal of Community Health, 44*(2), 307-312. https://doi.org/10.1007/s10900-018-0588-8

García, E., & Weiss, E. (2015). *Early education gaps by social class and race start U.S. children out on unequal footing: A summary of the major findings in "inequalities at the starting gate."* Economic Policy Institute. https://www.epi.org/research/education/

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2016). Are fit indices really fit to estimate the number of factors with categorical variables? Some cautionary findings via Monte Carlo simulation. *Psychological Methods, 21*(1), 93-111. https://doi.org/10.1037/met0000064

Giessman, J. A., Gambrell, J. L., & Stebbins, M. S. (2013). Minority performance on the Naglieri nonverbal ability test, second edition, versus the cognitive abilities test, form 6: One gifted program’s experience. *The Gifted Child Quarterly, 57*(2), 101-109. https://doi.org/10.1177/0016986213477190

Gordon, E., & Bridglall, B. (2005). *Nurturing talent in gifted students of color*. (pp. 120-146) Cambridge University Press. https://doi:10.1017/CBO9780511610455.009

Grissom, J. A., & Redding, C. (2016). Discretion and disproportionality: Explaining the underrepresentation of high-achieving students of color in gifted programs. *AERA Open, 2*(1), 1-25. https://doi.org/10.1177/2332858415622175

Hall, J., & Malmberg, L. (2020). The contribution of multilevel structural equation modelling to contemporary trends in educational research. *International Journal of Research & Method in Education, 43*(4), 339-347. https://doi.org/10.1080/1743727X.2020.1796066

Hamilton, R., McCoach, D. B., Tutwiler, M. S., Siegle, D., Gubbins, E. J., Callahan, C. M., Brodersen, A. V., & Mun, R. U. (2018). Disentangling the roles of institutional and individual poverty in the identification of gifted students. *Gifted Child Quarterly, 62*(1), 6-24. https://doi:10.1177/0016986217738053

Hertzog, N. B. (2017). Designing the learning context in school for talent development. *Gifted Child Quarterly, 61*(3), 219-228. https://doi:10.1177/0016986217705712

Hodges, J., Tay, J., Maeda, Y., & Gentry, M. (2018). A meta-analysis of gifted and talented identification practices. *Gifted Child Quarterly, 62*(2), 147-174. https://doi:10.1177/0016986217752107

International Baccalaureate. (2020). *Benefits for students*. https://www.ibo.org/benefits/benefits-for-students/

Jæger, M. M. (2011). Does cultural capital really affect academic achievement? New evidence from combined sibling and panel data. *Sociology of Education, 84*(4), 281-298. https://doi.org/10.1177/0038040711417010

Jæger, M., & Karlson, K. (2018). Cultural capital and educational inequality: A counterfactual analysis. *Sociological Science, 5*, 775-795. https://doi.org/10.15195/v5.a33

Jeffries, R., & Silvernail, L. (2017). Barriers to black student enrollment in honors and advanced placement courses. *Negro Educational Review, 68*(1-4), 56-79,157. https://www.questia.com/library/p142508/negro-educational-review

Jiang, G., Mai, Y., & Yuan, K.-H. (2017). Advances in measurement invariance and mean comparison of latent variables: Equivalence testing and a projection-based approach. *Frontiers in Psychology, 8*, 1823. https://doi.org/10.3389/fpsyg.2017.01823

Johnson, K. N., Thompson, K. L., & Farmer, R. L. (2020). Determining growth sensitivity of star math with a latent growth curve model. *Canadian Journal of School Psychology, 35*(3), 197–209. https://doi.org/10.1177/0829573520922678

Klingbeil, D. A., McComas, J. J., Burns, M. K., & Helman, L. (2015). Comparison of predictive validity and diagnostic accuracy of screening measures of reading skills: Predictive validity and diagnostic accuracy of screening measures. *Psychology in the Schools, 52*(5), 500-514. https://doi.org/10.1002/pits.21839

Lakin, J. M. (2016). Universal screening and the representation of historically underrepresented minority students in gifted education: Minding the gaps in Card and Giuliano's research. *Journal of Advanced Academics, 27*(2), 139-149. https://doi.org/10.1177/1932202X16630348

Levitt, H. M., Bamberg, M., Creswell, J. W., Frost, D. M., Josselson, R., & Suárez-Orozco, C. (2018). Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA publications and communications board task force report. *American Psychologist, 73*(1), 26-46. https://doi.org/10.1037/amp0000151

Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2010). *Methods in educational research: From theory to practice* (2nd ed). Jossey-Bass.

McCluskey, K. W. (2017). Identification of the gifted redefined with ethics and equity in mind. *Roeper Review, 39*(3), 195-198. https://doi:10.1080/02783193.2017.1318999

Meador, D. (2018). Comprehensive review of the STAR math online assessment. https://www.thoughtco.com/review-star-math-online-assessment-program-3194775

Meyer, J. (2020). *First steps in structural equation modeling: Confirmatory factor analysis.* The Analysis Factor. https://www.theanalysisfactor.com/structural-equation-modeling-first-step-confirmatory-factor-analysis-2/

Michelmore, K., & Dynarski, S. (2017). The gap within the gap: Using longitudinal data to understand income differences in educational outcomes. *AERA Open, 3*(1), 1-18. https://doi:10.1177/2332858417692958

Milner, H. R. (2013). Analyzing poverty, learning, and teaching through a critical race theory lens. *Review of Research in Education, 37*(1), 1-53. https://doi.org/10.3102/0091732X12459720

Mislevy, R. J. (1996). Test theory reconceived. *Journal of Educational Measurement, 33*(4), 379-416. https://doi.org/10.1111/j.1745-3984.1996.tb00498.x

Moule, J. (2009). Understanding unconscious bias and unintentional racism. *Phi Delta Kappan, 90*(5), 320-326. https://doi:10.1177/003172170909000504

Naglieri, J. A., & Ford, D. Y. (2015). Misconceptions about the Naglieri nonverbal ability test: A commentary of concerns and disagreements. *Roeper Review, 37*(4), 234–240. https://doi.org/10.1080/02783193.2015.1077497

National Association for Gifted Children. (1988). *Jacob K. Javits Gifted and Talented Children and Youth Education Act, S. 303, 100th Cong*. https://www.nagc.org/resources-publications/resources-university-professionals/jacob-javits-gifted-talented-students

National Association for Gifted Children. (2020). *A brief history of gifted and talented education.* http://dev.nagc.org/resources-publications/resources/gifted-education-us/brief-history-gifted-and-talented-education

National Association for Gifted Children. (2020). *Why are gifted program needed?* https://www.nagc.org/resources-publications/gifted-education-practices/why-are-gifted-programs-needed

National Association for Gifted Children & The Council of State Directors of Programs for the Gifted. (2015). *2014-2105 State of the States in Gifted Education: Policy and Practice Data. November.* https://www.nagc.org/resources-publications/gifted-state/2014-2015-state-states-gifted-education

National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. *The Elementary School Journal, 84*(2), 113-130. https://doi.org/10.1086/461348

Nihan, S. T. (2020). Karl Pearson's chi-square tests. *Educational Research and Reviews, 15*(9), 575-580. https://doi.org/10.5897/ERR2019.3817

Olszewski-Kubilius, P., & Corwith, S. (2018). Poverty, academic achievement, and giftedness: A literature review. *Gifted Child Quarterly, 62*(1), 37-55. https://doi:10.1177/0016986217738015

O'Shea, S. (2016). Avoiding the manufacture of 'sameness': First-in-family students, cultural capital and the higher education environment. *Higher Education, 72*(1), 59-78. https://doi.org/10.1007/s10734-015-9938-y

Park, J., & Park, M. (2016). Qualitative versus quantitative research methods: Discovery or justification? *Journal of Marketing Thought, 3*(1), 1-7. https://doi.org/10.15577/jmt.2016.03.01.1

Pearson. (2018). *NNAT3 manual: Levels A-D*. NCS Pearson, Inc.

Peters, S. J., & Engerrand, K. G. (2016). Equity and excellence: Proactive efforts in the identification of underrepresented students for gifted and talented services. *Gifted Child Quarterly, 60*(3), 159-171. https://doi:10.1177/0016986216643165

Pfeiffer, S. I. (2015). *Essentials of gifted assessment* (First edition). Wiley.

Qureshi, I., & Compeau, D. (2009). Assessing between-group differences in information systems research: A comparison of covariance- and component-based sem. *MIS Quarterly, 33*(1), 197. https://doi.org/10.2307/20650285

Rahi, S. (2017). Research design and methods: A systematic review of research paradigms, sampling issues and instruments development. *International Journal of Economics & Management Sciences, 06*(02), 1-5. https://doi.org/10.4172/2162-6359.1000403

Reardon, S. F., Kalogrides, D., & Shores, K. (2019). The geography of racial/ethnic test score gaps. *American Journal of Sociology, 124*(4), 1164-1221. https://doi.org/10.1086/700678

Redding, C. (2019). A teacher like me: A review of the effect of student–teacher racial/ethnic matching on teacher perceptions of students and student academic and behavioral outcomes. *Review of Educational Research, 89*(4), 499-535. https://doi.org/10.3102/0034654319853545

Renaissance Learning. (2020). *Research foundation for star adaptive assessment: Science of star.* Renaissance Learning, Inc.

Romanoff, B. S., Algozzine, B., & Nielson, A. B. (2009). Achievement of African American and Caucasian students referred and placed or not placed in gifted programs. *Journal for the Education of the Gifted, 33*(2), 156-175. https://doi.org/10.1177/016235320903300202

Ross, P. O. & United States (Eds.). (1993). *National excellence: A case for developing America's talent.* Office of Educational Research and Improvement, U.S. Dept. of Education.

Sablan, J. R. (2019). Can you really measure that? Combining critical race theory and quantitative methods. *American Educational Research Journal, 56*(1), 178-203. https://doi.org/10.3102/0002831218798325

Salkind, N. J. (Ed.). (2010). *Encyclopedia of research design*. SAGE Publications.

Sarstedt, M., Ringle, C. M., & Hair, J. F. (2017). Partial least squares structural equation modeling. *Handbook of Market Research, 26*, 1-40. https://doi:10.1007/978-3-319-05542_15-1

Solheim, E., Plathe, H. S., & Eide, H. (2017). Nursing students' evaluation of a new feedback and reflection tool for use in high-fidelity simulation – formative assessment of clinical skills. A descriptive quantitative research design. *Nurse Education in Practice*, *27*, 114-120. https://doi.org/10.1016/j.nepr.2017.08.021

Spicker, P. (2018). The real dependent variable problem: The limitations of quantitative analysis in comparative policy studies. *Social Policy & Administration, 52*(1), 216-228. https://doi.org/10.1111/spol.12308

Steenbergen-Hu, S., & Olszewski-Kubilius, P. (2016). Gifted identification and the role of gifted education: A commentary on "evaluating the gifted program of an urban school district using a modified regression discontinuity design." *Journal of Advanced Academics, 27*(2), 99-108. https://doi:10.1177/1932202X16643836

Sternberg, R. J. (2017). Accel: A new model for identifying the gifted. *Roeper Review, 39*(3), 152-169. https://doi.org/10.1080/02783193.2017.1318658

Sternberg, R. J. (2018). Context-sensitive cognitive and educational testing. *Educational Psychology Review, 30*(3), 857-884. https://doi.org/10.1007/s10648-017-9428-0

Study Point. (2020). *AP classes: To take or not to take?* https://www.studypoint.com/ed/ap-classes

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest, 12*(1), 3-54. https://doi:10.1177/1529100611418056

Tarka, P. (2018). An overview of structural equation modeling: Its beginnings, historical development, usefulness and controversies in the social sciences. *Quality & Quantity, 52*(1), 313-354. https://doi.org/10.1007/s11135-017-0469-8

Thompson, T., Coleman, J. M., Riley, K., Snider, L. A., Howard, L. J., Sansone, S. M., & Hessl, D. (2018). Standardized assessment accommodations for individuals with intellectual disability. *Contemporary School Psychology, 22*(4), 443-457. https://doi.org/10.1007/s40688-018-0171-4

Tirri, K. (2017). Teacher education is the key to changing the identification and teaching of the gifted. *Roeper Review, 39*(3), 210-212. https://doi.org/10.1080/02783193.2017.1318996

Topping, K. J. (2020). Implementation fidelity and attainment in computerized practice of mathematics. *Research Papers in Education, 35*(5), 529-547. https://doi.org/10.1080/02671522.2019.1601759

U.S. Department of Education. (2008). *New race and ethnicity guidance for the collection of Federal education data.* https://www2.ed.gov/policy/rschstat/guid/raceethnicity/index.html

U.S. Department of Education. (2020). *Reading and mathematics score trends.* Institute of Education Sciences, National Center for Education Statistics. https://nces.ed.gov/programs/coe/indicator_cnj.asp

Vogt, W. P. (2007). *Quantitative research methods for professionals.* Pearson/Allyn and Bacon.

Warne, R. T., & Price, C. J. (2016). A single case study of the impact of policy changes on identification for gifted programs. *Journal for the Education of the Gifted, 39*(1), 4961. https://doi.org/10.1177/0162353215624159

Washington State Legislature. (2009). RCW 28A.185.030 https://apps.leg.wa.gov/rcw/default.aspx?cite=28A.185.030

Watson, R. (2015). Quantitative research. *Nursing Standard, 29*(31), 44-48. https://doi.org/10.7748/ns.29.31.44.e8681

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 73*(6), 913-934. https://doi.org/10.1177/0013164413495237

Yuan, K.-H., & Chan, W. (2016). Measurement invariance via multigroup SEM: Issues and solutions with chi-square-difference tests. *Psychological Methods, 21*(3), 405-426. https://doi.org/10.1037/met0000080

APPENDIX

Path Analysis Graphic Model