EVALUATING PROGRAM DIVERSITY AND THE PROBABILITY OF GIFTED IDENTIFICATION USING THE TORRANCE TEST OF CREATIVE THINKING

Lindsay Ellis Lee, B.A.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

August 2021

APPROVED:

Anne Rinn-McCann, Committee Chair and
Chair of the Department of Educational Psychology
Rachel Mun, Committee Member
Selcuk Acar, Committee Member
Brenda Barrio, Committee Member
Karen Rambo-Hernandez, Committee Member
Randy Bomer, Dean of the College of Education
Victor Prybutok, Dean of the Toulouse Graduate School

Lee, Lindsay Ellis. *Evaluating Program Diversity and the Probability of Gifted Identification Using the Torrance Test of Creative Thinking*. Doctor of Philosophy (Educational Psychology), August 2021, 215 pp., 25 tables, 35 figures, chapter references.

Multiple criteria systems are recommended as best practice to identify culturally, linguistically, economically diverse students for gifted services, in which schools often incorporate measures of creativity. However, the role of creativity in identification systems and its recruitment of diverse student populations is unclear. The Torrance Test of Creative Thinking (TTCT) is the most widely used norm-referenced creativity test in gifted identification. Although commonly used for identifying talent, little is known on the variability in composite scores on the TTCT-Figural and student demographics (i.e., race/ethnicity, sex, socioeconomic status, English language learning status). This study evaluated student demographic subgroup differences that exist after the initial phase of an identification process (i.e., universal screening, referrals) and examined the relationship among student demographics (i.e., race/ethnicity, free/reduced lunch status, English language learning status, sex), cognitive ability, academic achievement, and creativity, as measured by the TTCT-Figural Form A or B, to the probability of being identified for gifted programs. In a midsized school district in the state of Texas, findings indicate several demographic differences for students who were referred or universally screened across the measures of cognitive ability, academic achievement, and creativity. However, there were lower differences when using the TTCT-Figural. Results of a hierarchical generalized linear regression indicate underrepresented groups showed no difference in the probability of being identified after controlling for measures of cognitive ability, academic achievement, and creativity. Though, cognitive ability and academic achievement tests were more predictive of identification compared to the TTCT-Figural. Implications and recommendations for future research are discussed.

Copyright 2021

by

Lindsay Ellis Lee

ACKNOWLEDGMENTS

The arduous process of writing this dissertation was supported by a wide network of people, both personally and professionally. I would like to thank my husband, Marty, and our son, Teddy, whose unwavering, multifaceted support at home provided me a solid foundation to pursue my goals. Thank you to my entire family, along with our close family friends, the Neymans, the Muths, the Coopers, and the Butters, who also reminded me to practice self-care. Also, a very special thank you to my fellow classmates, Dr. Melanie Meyer, Jessica Ottwein, Kacey Crutchfield, Leah Murphy, and Janessa Bower, whose mutual encouragement of one another continues to inspire me.

Thank you to Dr. Anne Rinn for her endless support and encouragement throughout this entire process as both by chair and advisor. I could not have accomplished this without her. Further, I am forever grateful for the opportunity to have been a graduate student in the educational psychology department focused on gifted and talented education at the University of North Texas. I would also like to especially thank my committee, Dr. Karen Rambo-Hernandez, Dr. Rachel Mun, Dr. Selcuk Acar, and Dr. Brenda Barrio, whose strengths helped me consider various perspectives, learn new methodological techniques, and improve my dissertation.

Lastly, I would like to acknowledge and thank Dr. Scott Peters, Dr. Matt McBee, Dr. Betsy McCoach, Dr. Tamra Stambaugh, and Dr. Kiana Johnson. I am eternally indebted to this team of people who supported and provided me valuable feedback. Specifically, I appreciate the time, moral support, and guidance afforded by Dr. Scott Peters and Dr. Matt McBee, whose work has been a huge inspiration throughout my doctoral studies and continues to influence my research.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

EVALUATING PROGRAM DIVERSITY AND THE PROBABILITY OF GIFTED IDENTIFICATION USING THE TORRANCE TEST OF CREATIVE THINKING

1.1 Introduction

For decades, equitable access to gifted programs for culturally, linguistically, ethnically, and economically diverse (CLED) students has been a persistent problem within gifted education (Baldwin, 2002; Ford et al., 2020; Lohman, 2009; Mun et al., 2020; Peters et al., 2019b; Plucker & Peters, 2016; Plucker et al., 2017). The United States Department of Education (USDOE) Office of Civil Rights (OCR) estimated that Black/African American and Hispanic/Latinx students comprise 42.3% of enrollment in schools, however only 25.7% are enrolled in gifted and talented programs. Additionally, English learners comprise 10.4% of the student population in schools nationwide, and only 2.6% of them are enrolled in gifted and talented programs (U.S. Department of Education, 2020). Beyond racial and linguistic disparities, students who are economically disadvantaged are further underserved in advanced classrooms (Borland, 2004; Mun et al., 2016; Siegle et al., 2016). As demographics within the United States continually become more diverse, educational systems need to prioritize how they will provide more equitable services to their student populations; this includes access to gifted education (Mun et al., 2020). Furthermore, there needs to be evaluation of existing systems of identification to understand which students are more likely to be identified for gifted services (Peters et al., 2019a; Peters & Engerrand, 2016), as well as an evaluation of the differential predictions of demographic subgroups with existing instruments in identification, inclusive of measures of creativity.

The entanglement of intelligence and creativity has been well-documented for decades

(Jauk et al., 2013; Karwowski et al., 2016; Kaufman & Plucker, 2011; Runco, 1993), and creativity is considered a facet of giftedness in numerous theories (Gagné, 2017; Renzulli, 1978; Tannenbaum, 2003). Thus, it is not surprising that both cognitive ability and creativity tests are often used within processes of gifted identification to mirror the complementary nature of their relationship to the construct of giftedness. Not only has cognitive ability and/or intelligence been discussed in relation to creativity (Guilford, 1967), but academic achievement has shown to relate with measures of creativity. Gajda et al. (2017) found in their meta-analysis a small positive relationship (r = .23 out of 100 studies, between tests of creativity and academic achievement. However, missing from their discussion was how this relationship connects with gifted identification and the relationship with demographic differences in gifted programs. Likewise, Desmet et al. (2021) evaluated a Dutch identification protocol and found no significant relationships among measures of intelligence and academic achievement with the Test of Creative Thinking-Drawing Production (TCT-DP; Urban & Jellen, 2010), but nevertheless concluded that the TCT-DP still provides useful information for gifted and talented programs despite not providing information on the likelihood of identification.

Torrance (2004) was a major proponent of including creativity measures in the gifted identification process, saying "creativity should almost always be one of the criteria, though not the sole criterion. In general, when creativity indicators are used, students who might otherwise be missed, should be included rather than exclude anyone" (p. 85). However, the creation or search for the 'perfect' creativity test to elicit a single score to identify creative giftedness or creative talents remains (see creativity quotient fallacy; Sternberg, 2018; Treffinger, 2004). Despite decades of research on creativity tests, there are several existing creativity instruments that should be evaluated within identification systems. There are numerous creativity tests used for gifted identification that need further evaluation for their performance and how correlated they are with other tests (Lee & Peters, 2021; McBee et al., 2014). Some of the most common assessments of creativity (e.g., divergent thinking, problem-solving, creative personality or behavioral characteristics) in gifted identification include the Gifted Rating Scales (GRS; Pfeiffer & Jarosewich, 2003), Scales for Identifying Gifted Students (SIGS; Ryser & McConnell, 2004), the Scales for Rating the Behavioral Characteristics of Superior Students (SRBCSS; Renzulli & Hartman, 1971), the Discovering Intellectual Strengths and Capabilities through Observation while allowing for Varied Ethnic Responses (DISCOVER; Maker, 2005), and the Torrance Test of Creative Thinking (TTCT; Torrance, 2017). More broadly, assessments that incorporate creativity have been suggested as more equitable in identifying diverse students and should be included as a criterion considered within the process (Kaufman, 2010; Luria et al., 2016). Torrance (2004) suggested that the use of creativity tests, like the TTCT, could help identify more students who demonstrate creative potential. Although scholars have enhanced their focus on cognitive ability (i.e., verbal and non-verbal) and academic achievement tests, there is scarce research on how creativity tests perform within a specific gifted identification system in a school district and the relationship to a range of student demographics (i.e., sex, ethnicity/race, socioeconomic status, English learner status) to the probability of being identified for gifted services.

Previous research on identification for gifted services has focused on combination rules (Lakin, 2018; McBee & Makel, 2019; McBee et al., 2014), nomination practices (McBee, 2006; McBee et al., 2016), deficit perspectives (Ford, 2014; Ford & Grantham, 2003), local norms (Peters et al., 2019b), universal screening (Card & Guiliano, 2016), behavioral rating scales (e.g., Peters & Pereira, 2017), and the usage of non-verbal cognitive ability measures (e.g., Carman et al., 2020; Lohman et al., 2008; Naglieri & Ford, 2003; Peters & Engerrand, 2016). Despite greater attention to providing more equitable identification practices (Plucker & Callahan, 2014), there has been limited recent empirical research on student demographic subgroup differences related to race/ethnicity, English learner status, and socioeconomic status when including creativity tests in gifted identification, particularly the figural form of the Torrance Test of Creative Thinking (Acar et al., 2021; Lee & Rinn, 2021; Torrance, 1971). The current best practice in identification of students for gifted services is to lessen barriers to access and incorporate multiple sources of evidence (Acar et al., 2016; Lakin, 2018; McBee et al., 2014). A multiple criteria system can be defined as a system that uses more than one determinant (e.g., standardized tests, portfolios, rating scales) for each student that informs consideration for admittance into the gifted program. This process commonly uses cognitive ability measures, achievement tests, and other alternative assessments (e.g., creativity measures, checklists, student work samples). Other decisions contribute to gifted identification, such as designated cut-scores, different combination rules, and other decisions for program development (e.g., funding, hiring personnel, program availability). As multiple criteria systems vary across the United States, the evaluation of these systems is necessary to determine how equitable they are for CLED students, especially in terms of how tests are used and the specific variability of scores across demographics in school districts. Thus, there is practical significance for school administrators to understand how creativity tests compare to measures of cognitive ability and academic achievement in identification for gifted services.

Even though some scholars have propagated creativity assessments (i.e., divergent thinking, creative behavior/personality, problem-solving) as one solution to help diversify gifted programs and allow more access to gifted services (Luria et al., 2016; Kaufman, 2010; Torrance, 2004), there is limited recent empirical research to support the notion that creativity tests (or divergent thinking tests) equate to greater representation of CLED student populations within gifted programs. Specifically, the TTCT (Torrance, 2017) is the most popular norm-referenced divergent thinking test used for gifted identification (Kaufman et al., 2008; Kaufman et al., 2012). However, more critical research is needed to determine whether the TTCT-Figural provides increased access for diverse populations, as well as how the TTCT-Figural relates to other tests (e.g., cognitive ability, achievement) used for gifted identification.

The purpose of this study is to examine demographic subgroup differences that exist beyond the first phase of a two-phase identification system (i.e., universal screening, nominations by teachers, parents, self, or community members) in relation to performance on the TTCT-Figural (Torrance, 2017) compared to cognitive ability and academic achievement measures. This includes comparing how hypothetical combination rules are used in gifted identification with different measures (i.e., cognitive ability, achievement, creativity) and the relation to gifted program diversity. Further, this study will investigate the relationship among student demographics, cognitive ability, academic achievement, creativity (i.e., TTCT-Figural), and the probability of being identified for gifted services, as well as the potential interactions among student demographics and scores on the TTCT-Figural in a midsized urban school district. Hence, the following review of literature will examine underrepresentation in gifted programs, best practices in gifted identification (e.g., the usage of multiple criteria, combination rules, alternative assessments), and the role of creativity tests in gifted identification prior to discussing the current study.

1.1.1 Underrepresentation in Gifted Program

More than half of United States public school students who are entering kindergarten through 12th grade are from culturally diverse backgrounds (i.e., Black, Hispanic, Asian, multiracial) and several underrepresented groups in gifted programs are expected to increase by 2028 (National Center for Education Statistics, 2019). Even so, school demographics within gifted education often do not reflect this broadened diversity and instead are further reflective of systemic inequities that have permeated throughout United States history (i.e., economic inequality, racial discrimination; Hamilton et al., 2018; Peters, 2021). For instance, although residential segregation laws ended during the Civil Rights Era (Fair Housing Act, 1968), along with desegregation in schools (Brown v. Board of Education, 1954), there are neighborhoods throughout the United States that still reflect the social inequality that was supposedly eradicated. Despite the earlier mid-century efforts, systemic inequities persist and impact school systems across the nation (e.g., Title I designations).

Borland (2003, 2004), Ford et al. (2020), and Ford et al. (2008), among others, have discussed criticisms of gifted education historically serving students from more advantageous backgrounds and providing them coveted educational resources (i.e., more individualized instruction, quality teachers, rigorous and engaging curriculum) and have urged more inclusivity of CLED students in advanced programs. Further, Plucker and Peters (2016) have found that the United States has "the largest income-based achievement gaps in the industrialized world" (p. 57) and few low-income students (i.e., eligible for free/reduced lunch) show advanced levels of performance on national tests. Olszewski-Kubilius and Corwith (2018) posit that underrepresentation of low-income students in gifted programs is due to the usage of cognitive ability and academic achievement tests in the gifted identification process, specifically with mandated high cut scores. The disparities in performance between students from high and low-socioeconomic backgrounds are likely due to having disparate access for opportunities to learn

(Olszewski-Kubilius & Corwith, 2018; Peters & Engerrand, 2016). To exacerbate the problem further, schools that lack funding (Hodges et al., 2021) or parents who do not have access to childcare to increase school readiness (Ricciardi et al., 2020) may not be able to maximize learning opportunities for students, thus further contributing to advanced student needs' not being adequately met (Dixson et al., 2020). Moreover, there are various environmental ramifications from continued inequity that impact students before they even are tested for gifted services (e.g., adverse childhood experiences, lead exposure, access to childcare or prenatal care; Peters, 2021). Thus, entangled with racial/ethnic inequities of who gets served, social and economic disparities at a micro level (student familial background) and macro level (school systems) continue to play a large role in perpetuating students being disproportionately identified for gifted services (Peters et al., 2019a; Plucker & Peters, 2018; Yoon & Gentry, 2009).

Access has been a persistent issue for scholars in the field of gifted education and can be attributed to similar causes of the achievement gaps in performance, as well as inconsistent definitions of giftedness and practices in gifted programs (Erwin & Worrell, 2012; Worrell & Dixson, 2018). Scholars in gifted education continue to research and advocate for underserved students to eliminate barriers to access (Gentry et al., 2020; Siegle et al., 2016) and provide more inclusive programming (e.g., culturally responsive pedagogy; Ford et al., 2000; Mun et al., 2021). Although there has been an increase in attention on identification practices, there still remains serious issues of underrepresentation across the United States (Hodges et al., 2018). Gentry et al. (2020) found that students who are Black/African American, Native American, Hispanic/Latinx, or Native Hawaiian/Pacific Islander remain underrepresented compared to White and Asian students. Likewise, Peters et al. (2019a) found similar results across racial/ethnic demographics and increases in representation for African/American and

Hispanic/Latinx students when identification is mandated. To extend the work of Yoon and Gentry (2009), Peters et al. (2019a) included students with limited English proficiency (LEP) found LEP students to be severely underrepresented within gifted programs, especially within states that mandated identification with no clear indication of why.

Mun et al. (2016) noted that while there has been progress made to incorporate inclusive definitions and policies for gifted identification for English learners (also for students as a whole), these practices have been inconsistently implemented. Their review of the literature suggest that cognitive ability and academic achievement tests were some of the greatest obstacles to identification for gifted programs and encouraged usage of non-traditional tests that are more dynamic and performance-based (e.g., creativity) for English learners. Similarly, scholars have criticized systems that emphasize IQ-based conceptions of giftedness and/or high thresholds of cognitive ability and achievement tests as significant barriers to identification of underrepresented groups (Ford & Grantham, 2003; Mun et al., 2020; Worrell & Dixson, 2018). Furthermore, Ford et al. (2008) posit that recruitment and retention barriers for CLED students are associated with identification practices that allow deficit perspectives to inform decisions (i.e., negative, stereotypical, and prejudicial beliefs that associate reasons for failure to internal deficiencies).

Notable barriers of entry for underserved populations have been addressed in gifted education research, including the impact of nomination practices (McBee et al., 2016), combination rules (e.g., mean and rules; Lakin, 2018; Lohman, 2005a, 2005b; McBee et al., 2014), outdated policies (Plucker et al., 2017; Siegle et al., 2016), and using insensitive or culturally biased tests (Naglieri & Ford, 2003, 2005). It is important to note that although culturally biased tests have been widely popularized as a barrier, there are not clear empirical studies that show standardized test bias is the cause of underrepresentation in gifted programs (Worrell & Dixson, 2020). More so, the discrepancies found indicate more evidence of consequential validity (Messick, 1989; Popham, 1997; Warne et al., 2013). Thus, meaning giving a test to a sample of students can show the systemic consequences, both positive and negative, of using a particular test in an identification system (e.g., different scores on a cognitive ability measure being due to environmental or social factors and influencing identification decisions). Moreover, there are other notable issues that complexly contribute to inequitable access to gifted programs beyond using different tests (e.g., high cut scores, misaligned tests to services provided, not providing frontloading opportunities; Lee et al., 2020; McBee et al., 2016; Peters, 2021; Peters & Engerrand, 2016; Plucker et al., 2017). For instance, Hodges et al. (2018) found within their meta-analysis of gifted and talented identification processes that despite use of non-traditional tests for identification, inequitable access for underrepresented groups still persists. Thus, evaluation of existing school systems is required to further understand the intricacies of various multiple criteria systems of identification and relation to demographic representation.

1.1.1.1 Multiple Criteria Systems

Multiple criteria systems have long been the gold standard to identify individuals for gifted programs (Frasier, 1997; Jolly & Robins, 2018). Like Torrance (2004) advocated for more than creativity to be included in decisions for gifted programs, Erwin and Worrell (2012) propose that standardized tests of intelligence (i.e., cognitive ability) should not be the only criteria used for gifted identification. Instead, multiple sources of evidence should be used to make decisions for placement in gifted programs. In the 2014-2015 *State of the States in Gifted Education* report (National Association of Gifted Children and Council of State Program Directors for the Gifted, 2015), respondents reported that 19 states utilize multiple criteria identification systems, with all 19 using two or more data points. This indicates that more than a third of states consider more than one data point and have to consider how they will combine these criteria to reach a decision on entry into a gifted program. Many states use multiple criteria as a way to improve the chances of students being identified through recognizing particular domains of strength (e.g., performance/artistic ability, creativity). However well-intended, the usage of multiple criteria is a muddled process that is often inconsistent with various measures used, some of which are possibly uncorrelated (McBee et al., 2016), and can be misaligned with program outcomes (Gubbins et al., 2021; Peters et al., 2020b). As there are countless combinations of criteria to choose from, testing commonly occurs in multiple phases to initially screen or refer students before they receive all tests in the formal evaluation phase, also known as a two-phase identification system.

1.1.1.1.1 Two-Phase Identification Systems

A two-phase identification system, as the name implies, consists of two-phases. In Phase 1, students can be either referred by teachers, parents, community members, or self-select for formal evaluation (Phase 2). This commonly consists of the person (e.g., teacher, parent, self, or other community member) referring the individual by filling out an observation behavioral checklist (e.g., Scales for Identifying Gifted Students [SIGS], Ryser & McConnell, 2004; Gifted Rating Scales [GRS], Pfeiffer & Jarosewich, 2003) to meet a specific threshold to be formally evaluated. Alternatively, students could be universally screened in Phase 1 with a specified test and meet a designated cut-score to be formally evaluated in Phase 2 (Lee & Peters, 2021). For example, students could be universally screened with a measure of verbal or non-verbal ability and score in the 90th percentile to be formally evaluated with other measures. Phase 2 consists of students receiving the full-battery of tests (e.g., multiple criteria designated by the district) to be considered for a gifted program. Once students are tested in Phase 2, a selection committee evaluates the student scores on multiple criteria and makes the determination if students are identified for gifted services.

1.1.1.1.2 Alternative Assessments

The formal evaluation criteria (within Phase 2) usually consist of cognitive ability measures (e.g., intelligence tests) and achievement tests. However, alternative assessments are also used to identify students for gifted services in some states and districts (Rinn et al., 2020). Alternative assessments are non-traditional tests that are combined with traditional tests in either the initial phase or the formal evaluation phase (Lakin, 2018; Lee et al., 2020; Lohman, 2009). Alternative assessments are inclusive of non-verbal cognitive ability tests, rating scales/checklists, performance-based assessments (e.g., portfolio), and creativity tests (VanTassel-Baska, 2008). VanTassel-Baska (2008) expressed that "perhaps the most popular of these non-traditional approaches is the use of nonverbal tests that purport to find more equal representation of minority and English language learners than more traditional measures" (p. 8). Non-verbal cognitive ability tests have been popularized to possibly mitigate barriers in access (e.g., non-verbal Cognitive Ability Test [CogAT], Lohman, 2012; Raven Progressive Matrices, Raven, 1998; Naglieri Nonverbal Ability Test [NNAT-2], Naglieri, 2008) through presenting test takers with visual stimuli (e.g., concrete objects, patterns, lines) that require non-verbal responses (e.g., completing a puzzle, filling in missing components). Naglieri and Ford (2003) suggested that the usage of non-verbal tests is necessary to lessen underrepresentation through avoidance of culturally biased or unfair test practices. However, more research has revealed that using nonverbal tests (e.g., NNAT-2) does not equate to equitable access or more diverse programs (e.g., Carman et al., 2020; Carman & Taylor, 2010; Lohman, 2005a, 2005b; Lohman &

Gambrell, 2012; Peters & Engerrand, 2016) and there is indication of potential publication bias based on authorship (Lee et al., 2021).

Thus, if non-verbal tests (i.e., non-verbal CogAT and NNAT-2) show similar discrepancies in performance with verbal tests (Peters & Engerrand, 2016), then that questions if using figural creativity tests (e.g., TTCT-Figural; Torrance, 2017) is helping to identify diverse students for gifted programs as has been purported by creativity test advocates (Cramond & Kim, 2008; Torrance, 2004). As states vary in their usage of specific criteria, it is imperative that there is more evaluation of multiple criteria systems in terms of alternative assessments (e.g., creativity tests) and combination rules used in gifted identification, and how that relates to identification for gifted programs.

If a goal within gifted education is to provide more equitable demographic representation, then we need to further understand how assessments used in the identification process are related to student demographics. Earlier research by McBee (2010) found that the probability of being identified for gifted services in Georgia elementary schools was related to student race and socioeconomic status and varied across schools; however, that research does not account for how specific creativity tests related to the probability of being identified. Likewise, Lakin (2018) addresses how combination rules would generalize to creativity tests used in gifted identification, but empirical research on combination rules with specific creativity tests used in an specific school district context is non-existent. Thus, further research is warranted to provide a contextual understanding of the probability of being identified with a creativity test (e.g., TTCT; Torrance, 2017), in addition to cognitive ability and achievement tests, within a school district that uses these specific criteria.

1.1.1.1.3 Combination Rules

Lohman (2009) suggested that “combining scores from different tests is thus almost a better policy than using a single score” (p. 986). Combination rules can be defined as rules that districts use to integrate multiple data points (e.g., test scores) to designate the gifted label and determine entry into gifted programs. This integration can take the form of rules that use the average of scores (i.e., average of Test 1 and Test 2; McBee et al.’s [2014] compensatory model), use either score from multiple scores collected where only one score might be high (i.e., use Test 1 OR Test 2; McBee et al.’s [2014] disjunctive model), or require multiple data points at a certain cut off (i.e., must meet criteria on Test 1 AND Test 2; McBee et al.’s [2014] conjunctive model). Lee et al. (2020) described how the gifted education policy for identification in Georgia uses an OR rule, in that they offer alternative pathways to identification. That said, students must score highly in three of four categories to be identified for gifted services (using a combination of AND and OR rules). Lee et al. also discuss how the use of an AND rule makes the group of students identified “more homogeneous” (p. 74), whereby an OR rule can make the group more heterogeneous (i.e., more diverse). McBee et al. (2014) found that the conjunctive model (i.e., AND) was the most restrictive in terms of sheer number of students identified for gifted services, followed by the compensatory model (i.e., average), and last the disjunctive model (i.e., OR).

Likewise, Lohman (2009), Lakin (2018), and McBee et al. (2014) all found that the OR rule was the least restrictive rule, because it allows a large pathway for more students to be identified for gifted services based on different domains of strength. However, a downside of implementing an OR rule is that the students who are identified are so vastly different in terms of ability and strength, that one gifted program may not meet the needs of all students identified.

The AVERAGE rule could be seen as a Goldilock's principle (e.g.,Capps, 2020), in that the rule seeks to balance two extremes and find the answer somewhere in the middle; it can both increase who is identified and create a more homogeneous group of students to be provided gifted services (McBee et al., 2014). McBee et al. (2016) and McBee and Makel (2019) cautioned if there are several measures used and are weakly correlated, they could result in fewer students being identified. For example, McBee et al. examined the correlation of the Gifted Rating Scales (GRS; Pfeiffer & Jarosewich, 2003) with the TTCT-Figural (Torrance, 2017) and found weak relationships (r = .19 with the creativity subscale of the GRS). Thus, if an identification system were to try to combine the GRS (Pfeiffer & Jarosewich, 2003) and TTCT-Figural with an AND rule, there would be much less sensitivity (i.e., not many students identified).

Past research has suggested that using AVERAGE and OR rules could increase program diversity (Lohman, 2009). Lakin (2018) found evidence of increased access for students who differed by race/ethnicity, students who qualify for free/reduced lunch, and English learners when OR and AVERAGE rules were applied. However, when program size was held constant, all three combination rules yielded similar diversity of students. Lakin (2018) posits that no matter the combination rules utilized when using multiple measures, there will be similar diversity of students identified for gifted programs. Additionally, Lakin suggests that these results would generalize to districts that use teacher rating scales, creativity tests, measures of cognitive ability, and achievement tests. Even if this is true, the current research on the combination of tests with divergent thinking and/or creativity tests in relation to student demographics and combination rules is limited (e.g., McBee & Makel, 2019; McBee et al., 2014). Considering Lakin's (2018) results that only used the CogAT7 (Non-Verbal, Quantitative, Verbal; Lohman, 2012) to test combination rules, would the results differ based on adding different tests with the full battery of the CogAT7 (Lohman, 2012), particularly achievement and a creativity test? Most school districts use more than a single test within the identification process, so it is crucial to know how different tests perform with each combination rule. More research is needed to understand usage of combination rules with a diverse battery of tests (inclusive of creativity tests), that is commonly seen within identification systems, and how combination rules relate to student demographics considered for gifted services.

1.1.2 Creativity and Gifted Identification

Creativity has played a central and integral role within conceptions of giftedness and intelligence since the mid-20th century (e.g., Guilford, 1967; Renzulli, 1978; Tannenbaum, 2003). Furthermore, creativity has been said to play a vital role in talent development because it helps transform giftedness to eminence (Cramond & Kim, 2008; Subotnik et al., 2011). With wider conceptions of giftedness and intelligence that incorporate views of creativity (e.g., three-ring theory of giftedness, Renzulli, 1978; triarchic theory of intelligence, Sternberg, 1983) and outcomes associated with creative productivity (e.g., schoolhouse giftedness versus creative productive giftedness, Renzulli, 2016), it seems logical to incorporate measures of creativity in identification for services in gifted education. More recently, Subotnik et al. (2011) proposed the movement of the field of gifted education towards identification processes that relate to domain-specific talent trajectories and should be inclusive of identifying creative potential that can be further developed into competence and later into expertise (Ericsson et al., 2007; Feldhusen, 2005). This is supported by a longitudinal study that noted creative behaviors found with the use of the TTCT-Verbal or Figural were predictive of longer-term adult creative performance (Clapham et al., 2005; Cramond et al., 2005; Harrington et al., 1983).

In the *2018-2019 State of the States in Gifted Education* report (Rinn et al., 2020), respondents report that 31 states incorporate creativity into their state definition of giftedness; however, there is limited information on how creativity is used in the identification process and integrated within gifted programs. Many states use multiple criteria systems (as discussed previously), but how creativity assessments are added into the equation has yet to be fully examined.

1.1.2.1 Creativity Tests

Creativity assessments used within gifted identification include divergent thinking tests (e.g., TTCT, Torrance, 2017; Wallach-Kogan Creativity Test, Wallach & Kogan, 1965), performance-based problem-solving assessments (e.g., DISCOVER, problem-solving tasks; Maker, 2005; VanTassel-Baska et al., 2007), creative products (e.g., Consensual Assessment Technique, Amabile, 1996), gifted rating scales (e.g., GRS, Pfeiffer & Jarosewich, 2003; SRBCSS, Renzulli & Hartman, 1971), and other creativity checklists (e.g., Kaufman et al., 2012; Proctor & Burnett, 2004). Callahan et al. (1995) found creativity was often included within definitions of giftedness, however the creativity assessments used were "fraught with problems" (Kaufman et al., 2008, p. 142) and not aligned with program outcomes. McBee et al. (2016) suggests the sensitivity and accuracy of an assessment can impact who is identified as gifted, specifically related to nomination bias and inconsistent use of teacher rating scales in two-phase systems. Thus, it is vital to fully understand measurement properties of assessments used at any point of the identification process.

Regardless of the measurement issues associated with creativity assessments, Kaufman et al. (2012) recommends that measures of creativity should be used to identify students for gifted services. This is due to the assumption that creativity tests provide less bias in testing comparable to traditional ability and achievement tests and provide a broader picture of the potential of a student. Although Kaufmen et al. caution the use of creativity assessments because of flawed validity and reliability with many creativity assessments, they do advocate for their usage despite their typical secondary role in the evaluation process (e.g., combination of cognitive ability, academic achievement, and creativity tests). If this is the case, then there needs to be more evaluation of widely used tests used for creativity within the gifted identification process.

1.1.2.2 Torrance Test of Creative Thinking

There are a wide range of creativity tests used within gifted identification, but the most widely used norm-referenced test to measure creative potential is the Torrance Test of Creative Thinking (TTCT; Torrance, 1974, 2008, 2017). Since the 1990s, the TTCT has been translated into more than 35 languages internationally and received substantial attention in validation research (e.g., Cramond et al., 2005; Kim, 2006; Kim et al., 2006; Said-Metwaly et al., 2018; Plucker, 2011; Yoon, 2017). Kaufman et al. (2008) even suggests that the TTCT is one of the most influential measures of creativity.

Several scholars have found a wide range of relationships between the TTCT-Figural and academic achievement (e.g., GPA, ACT, ITBS; Gajda et al., 2017), as well as analyzed the relation to intelligence (Kim, 2008; Shi et al., 2017). Although validity studies have found evidence that scores on verbal divergent thinking tests are predictive of longer-term creative achievements (Plucker, 2011; Runco & Acar, 2012), there is little known on the differential predictive validity of using creativity tests for gifted identification across student demographics. Relevant to the current study, Shi et al. (2017) studied the threshold hypothesis with the composite score of the TTCT-Figural and Raven's Standard Progressive Matrices (Raven et al., 1998) and found a moderate relationship ($r = .39$), similar to the correlation estimates of mental ability and creativity used in McBee et al.'s (2014) simulation study. Although intelligence and creativity have been extensively studied, there is little known of the relationship with the TTCT-Figural and the CogAT7 (Lohman, 2012) or NNAT-2 (Naglieri, 2008), all of which are frequently used in identification practices.

Due to psychometric concerns of creativity tests, Barbot et al. (2019) questioned if there should be continued use of the divergent thinking tests (e.g., TTCT). Plucker and Runco (1998) have noted that the researchers and educators have avoided divergent thinking tests due to the "perceived lack of predictive validity" (p. 38) that mainly stems from methodological concerns (e.g., duration of study, normality of data, psychometric properties of the test). These methodological concerns contribute to mixed results of criterion validity for divergent thinking tests that need to be further explored, particularly in terms of the usage within identification systems (Lee & Rinn, 2021). The TTCT continues to receive extensive study on construct validity (Acar et al., 2021; Forthmann et al., 2020; Said-Metwaly et al., 2018). For instance, the factor structure of the figural form of the TTCT has been repeatedly studied, in which evidence of a two-factor structure has been found (Kim et al., 2006; Said-Metwaly et al., 2018; Yoon, 2017). However, the confounding of fluency on other subscale scores continues to perplex researchers (Forthmann et al., 2020).

Grantham (2013) discussed how the rise of Torrance's creativity scholarship coincided with desegregation movements, and how Torrance helped to reframe creativity, amongst other scholars, as inclusive with definitions of giftedness (e.g., Renzulli, 1978; Tannenbaum, 2003; Torrance, 2004). Torrance advocated that the intersection of creativity, equity, and strength-based ideological orientations toward underrepresented groups could help increase access to gifted and talented programs, specifically with a focus on increasing the representation of Black males and economically disadvantaged students from the 1960s and 1970s (Henshon &

Grantham, 2021). However, current research has been limited in supporting Torrance's claims. In a systematic review of the literature, Lee and Rinn (2021) found studies that used the TTCT to have limited information on demographic subgroup differences beyond age, sex/gender, and grade-level in relation to gifted identification. Likewise, Acar et al. (2021) found potential discrepancies in regard to scoring tests taken by African American students when studying the differences in the updated standardized originality scoring (i.e., the standard list provided for scoring originality that tells the scorer what is original or not). Their findings suggest individuals of Black or African American ethnicities could be "adversely affected by the TTCT originality scoring, reducing their overall score and leading to an underrepresentation of such individuals in gifted education programs" (p. 12). On the other hand, Kim (2011, 2017) has advocated the utility of creativity tests in gifted identification, particularly the TTCT, as the creativity tests could identify more underrepresented groups of students within gifted programs. Kim (2006) contends that,

> standardized administration, scoring procedures and norms, and the development and evaluation have made the TTCT especially useful for identifying gifted and talented students. The TTCT-Figural can be fair in terms of gender, race, and community status, as well as for persons with a different language background, socioeconomic status, and culture (p. 8).

Similar to Kim (2006, 2011), Luria et al. (2016) claims creativity could be an "equalizer" to influence how accurately we identify students for gifted services and, subsequently, increase program diversity. Although there are arguments for the use of creativity tests in identification processes and in relation to equitable access, there has been scarce recent research on the TTCT in relation to differences of race/ethnicity and socioeconomic status since the Civil Rights Era (e.g., Torrance, 1971). Further, there is little known on the relationship of creativity tests, namely the TTCT-Figural, and the probability of being identified for gifted services.

1.1.3 The Current Study

Systemic inequities have historically afflicted gifted identification systems, largely due to achievement and excellence gaps that continue to exacerbate underrepresentation of CLED students (Worrell & Dixson, 2018). To increase equity in gifted programs, expanded notions of giftedness and intelligence have helped to change federal and state policies and influenced the adoption of multiple criteria in gifted identification practices to include creativity and amongst other characteristics (e.g., leadership, highly motivated, curiosity) related to the construct of giftedness (Plucker & Callahan, 2014; Marland Report, 1972; Frasier et al., 1995; Renzulli, 1978). However, merely implementing a multiple criteria system inclusive of an array of different tests is far more complex and requires a critical lens to parse out the relative merits and faults of identification systems.

The current analysis will use a quantitative critical perspective (Gillborn et al., 2017; Salban, 2018) to examine an identification system within a midsized diverse school district in the state of Texas. A quantitative criticalist perspective (i.e., quantcrit; Garcia et al., 2018) assesses "educational processes and outcomes to reveal inequities . . . to identify perpetuation of those that were systemic . . . [and to] question models, measures, and analytical practices, in order to ensure equity'' (Stage & Wells, 2014, p. 1; as cited in Sablan, 2018). Quantcrit was derived to support goals of critical race theory (CRT; Delgado & Stefancic, 2000; Ladson-Billings, 1998), and challenge the status quo (often with intentional and unintentional racist undertones) of wholeheartedly assuming statistical research is the only way truth can be derived to inform policy. Gillborn et al. (2017) argues that with appropriate reflexivity, quantitative research can improve equity in education. The authors suggest five general principles to follow when conducting quantitative research that focuses on race in education: (1) acknowledge the centrality of racism in society, (2) recognize data is not neutral and can serve Eurocentric, White assumptions and interests, (3) interrogate the use of categorical variables, (4) realize data can have numerous and conflicting interpretations, and (5) support goals of social justice (Gilborn et al., 2017). Social justice can be defined as "reconstructing society in accordance with principles of equity, recognition, and inclusion" (Bell, 1997, p. 3) and confronting institutional practices that advantage groups comparatively to marginalized groups.

1.1.4 Purpose and Research Questions

This study evaluated student demographic subgroup differences (race/ethnicity, sex, English learner status, and free/reduced lunch status) that may exist when using the TTCT compared to cognitive ability and achievement tests used in identification for gifted services. Influenced by Lakin (2018) and McBee et al. (2014), combination rules (i.e., OR, AND, AVERAGE) that incorporated scores on cognitive ability tests, achievement tests, and the figural form of the TTCT in relation to gifted program diversity were assessed. Additionally, I examined the relationship among student demographics (i.e., race/ethnicity, free/reduced lunch status, English learner status, sex), measures of cognitive ability (CogAT7, NNAT-2), academic achievement (ITBS Reading and Math), and the TTCT Figural Form A or B; and the interaction of student demographics and scores on the TTCT-Figural and the probability of being identified for gifted services. The following research questions were addressed in this study:

1. What demographic subgroup differences (e.g., race/ethnicity, sex, socioeconomic status, English learner status) exist after Phase 1 (i.e., universal screening or referrals) when using the TTCT compared to cognitive ability and academic achievement tests?
2. Using cognitive ability tests, academic achievement tests, and a creativity test (i.e., scores on the TTCT), how do combination rules relate to gifted program diversity?
3. To what extent do student demographics (e.g., race/ethnicity, sex, socioeconomic status, English Learning status), cognitive ability tests, academic achievement tests,

and scores on the TTCT predict the probability of being identified for gifted programs?

4. Do student demographics (e.g., race/ethnicity, sex, socioeconomic status, English Learning status) moderate the relationship among scores on the TTCT and the probability of being identified for gifted programs?

1.2 Method

1.2.1 A Priori Power Analysis

An a priori power analysis for logistic regression (G*Power 3.1; Faul et al., 2007), based on an estimation of a normal distribution, indicated that the minimum sample size would need to be 705 students to have the power to detect small effects (R^2 = .11) at the .05 level. Using the WebPower package (Zhang et al., 2021) in R, an additional a priori power analysis for linear regression, indicated that the minimum sample size would need to be 70 students to have the power to detect small effects (R^2 = .11) at the .05 level. If all student demographic variables are considered within a multiple regression analysis, the minimum sample size would need to be 140 students to have the power to detect small effects (R^2 = .11) at the .05 level.

1.2.2 Sample Characteristics

Rosemary School District (pseudonym) is a large district located in a midsize city near a major metropolitan area in the state of Texas. The National Center of Education Statistics (NCES) defines a midsize city as a "territory inside an urbanized area and inside a principal city with population less than 250,000 and greater than or equal to 100,000" (National Center for Education Statistics, 2015, p. 2). Rosemary School District expands a diverse area and includes 18 of 23 elementary schools that receive funding for Title I status.

In the 2018-2019 school year, Rosemary School District served more than 29,000 students in Grades K-12 and had 3,276 students (10.9%) identified for gifted services. Of the total population of the district, 46.7% identified as White, 31.1% as Hispanic/Latinx, 15.2% as Black, 3.2% as Asian, 1.4% as two or more races, .6% as American Indian/Alaskan Native, and .2% as Hawaiian and Other Pacific Islander. Students with economically disadvantaged status (i.e., students on free/reduced lunch) comprised 45.5% of the student population and English learners comprised 14.8% (Texas Education Agency [TEA], 2019). Although the district and TEA could not provide the total number of males or females in the 2018-2019 school year, the Office of Civil Rights (U.S. Department of Education, 2020) indicated there were 48.6% females and 51.4% males in the previous school year. The sample collected consisted of student-level identification data from the 2018-2019 school year for all students who were referred by parents, teachers, self, or community members, or universally screened with the NNAT-2 in 1st grade. See Table 1.1 for frequency statistics for the total and aggregated samples in the 2018-2019 school year.

1.2.2.1 Rosemary School District Gifted Program and Identification Process

Rosemary School District's gifted program is designed to offer students advanced instruction and specific academic opportunities based on general intellectual ability within individual and collaborative environments (inside or outside the classroom). Gifted specialists are to provide students with instruction to develop "complex thinking, problem-solving" (Rosemary Document, 2016, p. 5), and advanced learning opportunities. The scope and sequence for the district dictates gifted specialists need to introduce, apply, and maintain specific skills. For instance, students' critical thinking skills are developed through cognitive process dimensions (i.e., analyze, evaluate, create), spatial thinking, logical thinking, and dimensions of depth and complexity. Regarding creative thinking, the program incorporates elements of divergent thinking (i.e., fluency, flexibility, originality, and elaboration), problem-solving (e.g., SCAMPER model, the Creative Problem-Solving Model), and morphological forced connections (Rosemary Document, 2016, p. 10). In elementary school (K-5th grade), students are served by gifted specialists in a pull-out class at the campus level through either grade-specific or multi-age classes. Students are also grouped by clusters of 3 to 5 within homeroom classes. In middle school (6th-8th grade), students meet in a daily gifted program class and gifted specialists provide instruction to develop creative and critical thinking skills, in addition to opportunities for student research projects to create "advanced products" (Rosemary Document, 2016, p. 5). High school students are served in the district's gifted program through Advanced Placement courses and extracurriculars. High school student identification for gifted services was not the central focus of this study.

1.2.2.1.1 Identification

Rosemary School District define gifted and talented students as:

> a child or youth who performs at or shows the potential for performing at a remarkably high level of accomplishment when compared to others of the same age, experience, or environment and who (1) exhibits high performance capability in an intellectual, creative, or artistic area; (2) possesses an unusual capacity for leadership; or (3) excels in a specific academic field. (Texas Education Code § 29.121, 2019)

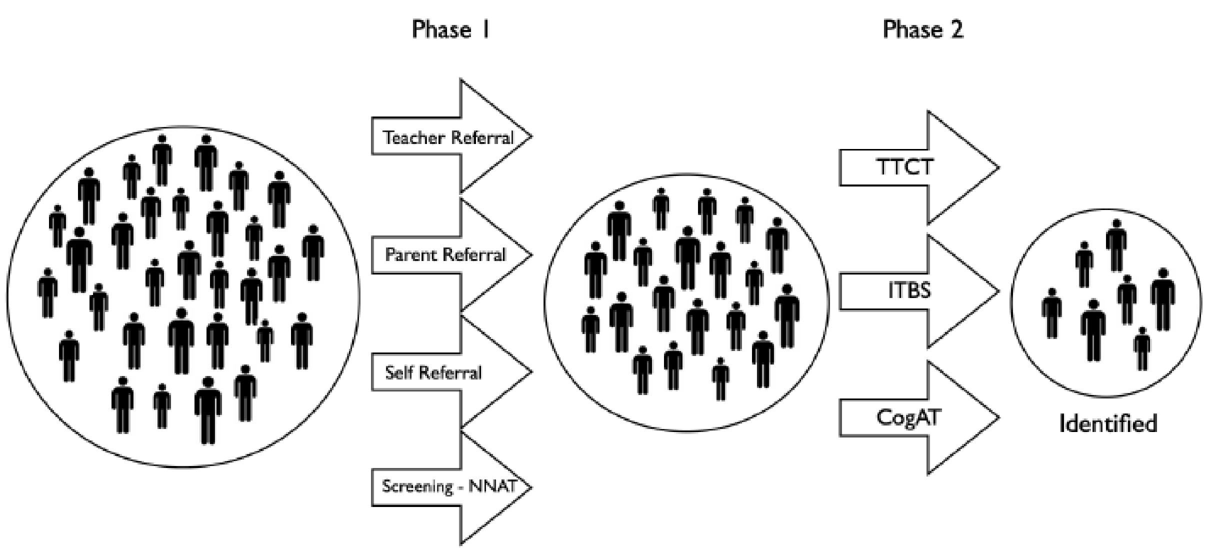
To follow the state definition of gifted students, Rosemary School District's gifted program requires use of multiple criteria specified by the district and their documentation explicitly states that students who are gifted and talented "can come from all races, socioeconomic groups, geographical locations, and environments" (Rosemary Document, 2016, p. 4) and assessed in native languages or with non-verbal ability tests. Likewise, Rosemary School District uses planned experiences in elementary to help find students who would benefit from gifted services and frontload learning experiences based on creativity and critical thinking, planned activities students could encounter within the gifted program (Plucker et al., 2017b).

1.2.2.1.2 Two-Phase Identification System

Rosemary School District designates their identification process as a three-step process: (1) students are referred, (2) they are assessed, and then (3) they go through both campus and district selection committees, however this can be expressed as a two-phase identification system. This is seen through students having to pass through Phase 1 criteria (i.e., referral, universal screening) to be formally evaluated with tests (i.e., Phase 2) to be designated as gifted and served in the gifted program. All kindergarteners and first graders go through a universal screener in the fall semester; in addition, students can be referred by parent and teacher checklists at any grade-level (K-12). The majority of referrals for formal evaluation occur in first grade through eighth grade; There are limited number of referrals that occur in ninth through twelfth grades. For the current study, the focus is only on first grade through eighth grade in regard to formal identification in Phase 2. See Figure 1.1 for a model of Rosemary School District's two-phase identification system.

Figure 1.1

Two-Phase Identification System for Rosemary School District



Students are universally screened with a non-verbal cognitive ability measure in first grade (i.e., NNAT-2; Naglieri, 2008) and, at any time point in elementary or secondary, students can be referred by anyone (i.e., parent, teacher, community member, self) to be formally evaluated for gifted services. In Phase 2, in addition to scores on either a parent and/or teacher referral form, students must score highly on two of three criteria to be identified in first through twelfth grade. The criteria include a combination of cognitive ability tests (i.e., CogAT7 [Non-Verbal, Quantitative, Verbal]), achievement tests (i.e., ITBS Reading, ITBS Math), and a figural divergent thinking test (i.e., TTCT-Figural) to identify students for gifted services. After testing, Rosemary School District designates a campus screening committee that involves a campus administrator, counselor, classroom teacher, and gifted specialist, who are trained in the nature and needs of gifted children, to formally evaluate student results from each semester's testing sessions. The campus committee will recommend students to the district selection committee. Students are only allowed to be tested once per year; however, the district designates an appeal, furlough, and exiting process out of the program. For the current study, the dataset was attenuated among students who were referred in all grades and students universally screened in first grade to better understand student differences for each pathway to formal evaluation for gifted services.

1.2.3 Variables

1.2.3.1 Dependent Variable(s)

For Research Question 1, the dependent variable consists of scores on the TTCT-Figural, CogAT7 (Non-Verbal, Quantitative, Verbal), ITBS (Reading and Math), and the NNAT-2. For questions three and four, the dependent variable is a binary outcome that consists of either being identified as gifted (1) or not identified as gifted (0).

1.2.3.2 Student and School Demographics

Student level data will include race/ethnicity, sex, free/reduced lunch status, and English learner status. Race/ethnicity was operationalized based on the federal racial and ethnic distinctions and combined into a single variable; Race was categorized as Black/African American, Hispanic/Latinx, Asian, Native American (i.e., federal designation as American Indian/Alaska Native), two or more races, Native Hawaiian/Pacific Islander, or White; all were dummy coded within the dataset. Since there were small sample sizes of students in the categories of two or more races and Native Hawaiian/Pacific Islander, they were consolidated into a single category called Other. White students served as a reference group since they represent the largest portion of students referred or universally screened. English learners were dummy coded (1 = English learner; 0 = not an English learner), as was sex (1 = female, 0 = male). Socioeconomic status was operationalized as students with free/reduced lunch status and was dummy coded (1 = free/reduced lunch, 0 = full priced lunch).

1.2.4 Measures

1.2.4.1 Torrance Test of Creative Thinking-Figural

The TTCT is used to evaluate creative thinking skills, specifically divergent thinking in the form of graphical representations (Torrance, 2017). The TTCT-Figural uses parallel forms (Form A or Form B), and each consist of three drawing activities: Activity one consists of one large stimulus; activity two consists of ten incomplete figures; activity three consists of 2-3 pages of repeated figures in either lines (Form A) or circles (Form B). Schools can choose to alternate between using either Form A or Form B within their identification process. All three activities are assessed based on five subscales: fluency, originality, elaboration, abstractness of titles, and resistance to premature closure. These five subscales are described as follows:

1. Fluency: The number of relevant ideas; shows an ability to produce a number of figural images.

2. Originality: The number of statistically infrequent ideas; shows an ability to produce uncommon or unique responses. The scoring procedure counts the most common responses as 0 and all other legitimate responses as 1. The originality lists have been prepared for each item based on normative data, which are readily memorized by scorers.

3. Elaboration: The number of added ideas; demonstrates the subject's ability to develop and elaborate on ideas.

4. Abstractness of Titles: The degree beyond labeling; based on the idea that creativity requires an abstraction of thought. It measures the degree a title moves beyond concrete labeling of the pictures drawn.

5. Resistance to Premature Closure: The degree of psychological openness; based on the belief that creative behavior requires a person to consider a variety of information when processing information and to keep an "open mind." (Kim, 2006, p. 5)

To obtain an overall Creativity Index composite score, the standardized scores of the five subscales are used. This is obtained through converting the raw scores to standard scores with means of 100 and a standard deviation of 20. The range of standard scores include fluency (40-154), originality (40-160), elaboration (40-160), abstractness of titles (40-160), and resistance to premature closure (40-160; Kim, 2006, p. 5). The average of these standard scores yields the Creativity Index. Kuder-Richardson Formula 20 reliability estimates obtained from four studies show the Creativity Index ranges from .87 to .96 (Torrance, 2017, p. 71). Reliability estimates could not be obtained from the current study due to the secondary nature of the data and not having access to item-level data.

For the following study, the TTCT-Figural was used since the district uses the TTCT-Figural Form A or B as criteria for creativity, and because it is the most widely used norm-referenced test used in identification for creativity (Callahan et al., 1995; Kaufman et al., 2008). Although not the focus of this study, it is important to note the ongoing criticisms of construct validity of the TTCT-Figural. The TTCT-Figural has undergone extensive validation studies

(Kim et al., 2006), but many scholars continue to interrogate the dimensionality of the construct (e.g., one-factor or two-factor; Said-Metwaly et al., 2018) and the confounding influence of fluency (Forthmann et al., 2019). More specific to this study, Torrance (1974) cautioned the use of TTCT composite scores, as a composite score may be misleading and does not show strengths within each subscale. However, it is common for composite scores to be used in gifted identification, as seen in Rosemary School District. The district uses a 125 flat cut score for the TTCT-Figural, regardless of the age or grade of the student, which is, on average, within the 83rd percentile across ages.

1.2.4.2 Naglieri Nonverbal Ability Test-2

The Naglieri Nonverbal Ability Test (NNAT-2, Naglieri, 2008) is an individual or group-administered nonverbal cognitive ability test based on analyzing figural matrices made of geometric shapes. The NNAT-2 is characterized as being culturally "neutral" (Mun et al., 2020, p. 13) and perceived as ideal for a diverse student population. The NNAT-2 has seven levels, each level (specific grade intended) of the test consists of 48 items. The NNAT-2 is made up of four formats, formerly known as pattern completion, reasoning by analogy, serial reasoning, and spatial visualization; these are no longer separate scores. From a standardization norming sample of children from age 5 to 14, Kuder-Richardson Formula 20 reliability coefficients (from raw scores) range from .83 to .92 (Naglieri, 2008). Balboni et al. (2010) found evidence of concurrent and predictive validity of the NNAT and Raven's Colored Progressive Matrices (Raven et al., 1998) with achievement scores; specifically, they found the NNAT accounting for 10% more of the variance in math test scores than Raven's Colored Progressive Matrices. The NNAT-2 Naglieri Ability Index (NAI) is a standardized score (M = 100, SD = 16) and based on short intervals of chronological age (3 months). The participating district only administers the NNAT-

2 in the first grade and uses a range of NAI scores in their identification process. In Rosemary School District, according to their documentation, students need to score within the 94th percentile (NAI = 125), however they will consider students who score in the 89th percentile (NAI = 120) as well. Observed scores for students who were universally screened and formally evaluated ranged from 63 to 133. For this study, the blanket cut-off of 125 was used to assess combination rules.

1.2.4.3 Cognitive Abilities Test-7

The Cognitive Ability Test (CogAT7; Lohman, 2012) is a group administered ability test that examines verbal, quantitative, and non-verbal reasoning ability for Grades K-12. Each CogAT full battery consists of three subtests (nine total tests). Each battery is designed to assess both "inductive and deductive reasoning abilities, which Cattell would classify as fluid-analytic abilities" (Lohman, 2012, p. 7). The Non-Verbal Battery is a figural test that consists of three subtests that assess figure classification, figure matrices, and a paper folding subtest. The Quantitative Battery consists of number analogies, number puzzles, and number series subtests. The Verbal Battery consists of picture analogies, picture classification, and sentence completion subtests. Construct and concurrent validation studies have provided support in using the CogAT for gifted identification (Warne, 2015). Scores are standardized by age (M = 100; SD = 16); CogAT 7 adopted age-based level classifications to model the Iowa Assessments Form E (Dunbar et al., 2011). The primary reporting score for the CogAT7 is the Standard Age Scores (SAS) and range from scores of 50 to 160 (Lohman, 2012, p. 44). Split-half reliabilities were used to determine equivalency of forms; Reliability coefficients were obtained across the standardization sample on all three batteries: the reliability estimates for the Verbal Composite ranged between .91 to .95, the Non-Verbal Composite ranged between .91 to .94, and the

Quantitative Composite ranged from .91 to .94. The participating district only administered the non-verbal CogAT in fourth grade. For the current study, the full battery of the CogAT7 (Non-Verbal, Quantitative, Verbal) was used in the formal identification process at Rosemary School District. For each subscale, the cut-score designated by the district was 125, which is at the 94th percentile.

1.2.4.4 Iowa Test of Basic Skills Form A

The Iowa Test of Basic Skills (ITBS; Hoover et al., 2007) is a group administered achievement test battery that examines academic progress within major content areas (e.g., Reading, Math, Social Studies, Science). At Rosemary School District, the ITBS Form A of the Reading and Mathematics portions are used with age-based leveling. ITBS Reading evaluates foundational reading comprehension skills (words and comprehension) at the younger levels (e.g., level six) and as students move to higher levels of reading (i.e., become more independent), the emphasis changes to derived meaning from passages (Levels 7-14). ITBS reading questions consist of reading passages and subsequent questions that either are multiple choice or require a one to two sentence response: these questions scale in complexity. The ITBS Math examines foundational mathematical skills at the lower levels, then moves to assessing math concepts, math problems, and math computation through level seven. Then from levels nine through 14, mathematical concepts and estimation, mathematical problem-solving, and data interpretation, and mathematics computation are assessed. The ITBS Math provides open-ended questions that occasionally require students to analyze/solve problems and "describe their thinking using words, diagrams, graphs, symbols, calculations, and equations, or inequalities" (Iowa Test of Basic Skills: Guide to Research and Development, n.d, p. 154). Concurrent validity was assessed with the CogAT6 and ITBS Form A with the same standardization sample and found moderate to high correlations between each subscale across grade levels. Internal-consistency reliability estimates for Form A for Levels 7-14 for ITBS Reading is α = .89 and ITBS Math with computation is α = .87. For Level 6, Kuder Richardson Formula 20 reliability estimates for ITBS Reading range from .91 to .94 and for ITBS Math ranges from .79 to .82 in the Fall and Spring administration with a standardization sample. For the current study, Rosemary School District established the cut-off score for either ITBS Reading or ITBS Math at the 95th percentile.

1.2.5 Analysis

All analyses were conducted in R Version 4.1.0, RStudio Version 1.2.1335 (RStudio Team, 2020). Prior to analysis, there were 19 cases that were cut for missing demographic information that were unavailable at the school district. After assessing for missingness in the data using the Naniar package (Tierney et al., 2021), 18.9% of the data was missing. To investigate further, I subset the data between students who were referred (i.e., from parents, teachers, self, community members) and students who were universally screened with the NNAT-2 in first grade. I found 6.85% of responses were missing from referrals, whereas there was 25.76% missing from students who were universally screened. Each subset violated Little's (1988) missing completely at random test, however the reason for missingness was systematically known. From the students who were referred, a large portion of missingness was from scores on the NNAT-2 (94.29%), however the majority of students were not ever required to take the NNAT-2 as part of the identification process unless they scored low on the CogAT Non-Verbal Battery. Likewise, there is a large portion of students who are universally screened (n = 2070), and despite all students having scores on the NNAT-2, there was 83.04% missing from taking either form of the TTCT (i.e., Form A or Form B). Upon further examination within the universal screening subset, only 128 students were identified in the dataset. Since the interest is analyzing Phase 2 (i.e., formal evaluation), each subset was analyzed separately all students who were referred, and all students universally screened in first grade.

Due to the known source of missingness, multiple imputation was appropriate and was used to preserve the sample size (Peugh & Enders, 2004). Prior to imputation, outliers within the dataset were assessed using boxplots and extreme cases were marked as not available (i.e., NA) and imputed along with other missing data points. Using the mice package (i.e., Multivariate Imputation by Chained Equations), 100 imputations using predictive mean matching were conducted for continuous variables (Graham & Olchowski, 2007; van Buuren et al., 2020). After initial imputation, the complete function pooled parameter estimates and variance statistics (Rubin, 1996).

1.2.5.1 Research Question 1

For Research Question 1, descriptive statistics (e.g., means, standard deviation, correlation coefficients), bivariate statistics, and data visualization techniques were used to analyze the data. Visual representations showcase differences in distributions comparable to student demographics and measured variables (e.g., density plots). Then, a series of linear regressions analyzed each student demographic subgroup separately (i.e., sex, Black/African American, White/Caucasian, Hispanic/Latinx, Asian, Native American, Hawaiian/Pacific Islander, two or more races, free or reduced lunch, English learners) in relation to the TTCT-Figural, cognitive ability measures (i.e., CogAT7, NNAT-2), and achievement tests (i.e., Iowa Test of Basic Skills, Reading and Math). For example, a simple linear regression analyzed sex differences with the TTCT, then additional regressions will analyze sex differences with the CogAT7 (Non-Verbal, Quantitative, Verbal), NNAT-2, ITBS (i.e., Reading, Math) independently. Further, a multiple regression analyzed race/ethnicity differences in a single model. Each dependent variable (TTCT-Figural, CogAT-7, NNAT-2, ITBS Reading and Math) was standardized prior to any analysis to compare the overall effect of each model (ΔR^2) in comparison to TTCT Form A or B.

Standardized beta coefficients, squared structure coefficients (for multiple regression), confidence intervals, R^2 effect sizes, and the incremental change in R^2 (i.e., ΔR) of each model comparable to the TTCT Form A or B are reported. For each of the regressions, the assumption of linearity was met since the categorical variables were dummy coded. Since there were several regressions analyzed (24 regressions for referrals [6 regressions per demographic subgroup], 28 regressions for universal screening [7 regressions per demographic subgroup]), the Benjamini Hochberg (α_{FDR}) was used to control for a false discovery rate (i.e., proportion of errors that occur when false rejection of the null hypothesis) for multiple comparisons (Benjamini & Hochberg, 1995).

As an additional analysis, a hierarchical regression analysis was used to examine the relationship among student demographics and all tests to each form of the TTCT (i.e., Form A, Form B, highest score of Form A or B). In addition to each form being analyzed, the data was analyzed by students who were referred or universally screened in first grade. For students referred, the relationships among all student demographic variables and tests used (i.e., CogAT7, ITBS Reading, ITBS Math) to the TTCT (Form A, Form B, and Form A or B) were assessed. For students universally screened, the relationships among all student demographic variables and tests used (i.e., NNAT-2, CogAT7, ITBS Reading, ITBS Math) to the TTCT (Form A, Form B, and Form A or B) were assessed.

For multiple regressions, using the performance package in R (Lüdecke et al., 2019), assumptions were assessed (i.e., linearity, autocorrelation, multicollinearity, heteroscedasticity, normality). For Research Question 1, all assumptions were met besides indication of violation of heteroscedasticity (e.g., p = .037) and normality within the hierarchical regressions. Upon reflection of the fitted vs. residual plots, data were observed to have acceptable heteroscedasticity. Also, although the performance package detected non-normality, the shape of the distributions for many of the measures were normally distributed, besides the ITBS Math and ITBS Reading scores. The detection of non-normality could be attributable to the left skew toward the upper end of the distribution and be related to the nature of the data not being randomly sampled. To account and guard against false discovery rate in multiple comparisons, a Benjamini Hochberg (α_{FDR}) was used to conservatively adjust p-values (Benjamini & Hochberg, 1995).

1.2.5.2 Research Question 2

1.2.5.2.1 Demographic Representation

For Research Question 2, frequency counts for each demographic subgroup (i.e., sex, Black/African American, White/Caucasian, Hispanic/Latinx, Asian, Native American, Hawaiian/Pacific Islander, Multiracial, free or reduced lunch, English learners), the total student population, and total students identified as gifted were first analyzed to calculate the percentage identified and then representation indices (Peters et al., 2019; Yoon & Gentry, 2009). Based on the work of Yoon and Gentry (2009), representation indices were calculated as follows:

$$RI = \frac{\%\,Gifted}{\%\,Total\ Population}$$

To assess the demographic representation at Rosemary School District, representation indices were calculated based on all students within the sample who were identified as gifted. Total student counts were obtained by the Texas Education Agency's Texas Performance

Accountability Report (TAPR; Texas Education Agency, 2020) for the 2018-2019 school year in Rosemary School District. For females and males, total population percentages were based on Office of Civil Rights data. The total students identified for gifted services by subgroup only consist of students identified in elementary and middle school across all 31 schools in the district. Proportions per demographic subgroup are reported and proportionality are discussed.

1.2.5.2.2 Combination Rules

Influenced by both Lakin (2018) and McBee et al. (2014) on their work on combination rules with the full battery of the CogAT7, hypothetical combination rules based on the cut scores used by Rosemary School District were applied. To extend Lakin's work, the following analysis added a creativity test (i.e., TTCT) and an achievement test for reading and math (i.e., ITBS Reading, ITBS Math) prior to computing the combination rules. This was to follow what Rosemary School District uses in their second phase of identification, which is the full battery of the CogAT7, ITBS Reading and Math, and TTCT-Figural (TTCT-Figural) Form A or B. The entire sample of students that went through the identification process was used for this portion.

Each student demographic was compared based on the percentage of students who would be hypothetically identified with three specific combination rules. These rules included an OR rule (at least one score above the districts' designated cut score[s]), AND rule (multiple measures at the districts' designated cut scores for each test), or an AVERAGE rule (the average of multiple measures). To compute the OR and AND rule, students were filtered by the designated cut scores of the district (i.e., 95th percentile on ITBS Math & Reading, 94th percentile [SAS score of 125] for CogAT7 [Verbal, Quantitative, Non-Verbal], 83rd percentile for the TTCT Form A or B [125]) and selected by demographic subgroups for comparison. For the average rule, all tests were standardized (i.e., converted to Z-score), then averaged across all tests. Due to sampling effects and shrinkage of variance after standardizing all tests, an adjusted cut-score was computed using the giftedCalc's package (McBee, 2021) to determine the appropriate adjusted cut score that would better represent the 95th percentile across tests.

1.2.5.3 Research Questions 3 and 4

To answer the third and fourth research questions, a hierarchical generalized linear regression analysis was used. For the third research question, a logistic regression analysis was used to investigate the relationship among student demographics (e.g., race/ethnicity, sex, socioeconomic status, English learning status), cognitive ability measures (i.e., NNAT-2, CogAT), achievement tests (i.e., ITBS Reading, Math), and scores on the TTCT to the probability of being identified for the gifted program at Rosemary School District (Gifted Identification = 1, if yes; 0, if no). Assumptions for hierarchical generalized linear regression, more specifically logistic regression, were assessed (i.e., binary dependent variable, observation independence, lack of multicollinearity, linearity related to log odds, outliers, sufficient sample size; Menard, 2002; Pampel, 2000) and found to be acceptable. Contrary to multiple regression, logistic regression coefficients indicate change in expected log odds (i.e., odds are transformed into logits for interpretation). Odds ratios, probability estimates, confidence intervals, and Tjur-R^2 will be reported. Tjur R^2 (i.e., the coefficient of determination) is an analogue to linear regression R^2 effect size estimates and is calculated by computing the mean of the predicted probabilities of the dependent variable and taking the difference between the means (Tjur, 2009).

To answer the fourth research question, a hierarchical generalized linear regression will be used to explore the utility of specific predictor variables in explaining the probability of being identified as gifted. Also, this allowed model comparison with the incremental changes in the Tjur R^2 effect size from the main effects model to secondary effects when interaction terms are added to a subsequent model. Before assessing secondary effects, continuous predictors were mean-centered, and interaction terms were created to assess the moderation of student demographics and scores on the TTCT to the probability of being identified for the gifted program in the same district. For all generalized linear models, a Benjamini Hochberg (α_{FDR}) was used to account for possible Type I error (i.e., false discovery rate; Benjamini & Hochberg, 1995).

1.2.5.3.1 Model Description

For both the referred students and students who were universally screened in first grade, a hierarchical generalized model was used for each subset of the data. In Block 1, I examined the main effects of all the student demographic control variables to the probability of being identified as gifted.

1.2.5.3.2 Referrals

For students referred, the following model was run and transformed to predicted probabilities:

$$\text{Logit}_{\text{Gifted Identification}} = \beta_0 + \beta_{\text{Black}} + \beta_{\text{Hispanic}} + \beta_{\text{Asian}} + \beta_{\text{AmeriInd}} + \beta_{\text{Other}} + \beta_{\text{Female}} + \beta_{\text{Free/reduced lunch}} + \beta_{\text{English learners}} + \varepsilon$$

In Block 2, I examined the main effects of achievement and non-verbal cognitive ability measures to the probability of being identified as gifted. The following model wase run and transformed to predicted probabilities:

$$\text{Logit}_{\text{Gifted Identification}} = \beta_0 + \beta_{\text{Black}} + \beta_{\text{Hispanic}} + \beta_{\text{Asian}} + \beta_{\text{AmeriInd}} + \beta_{\text{Other}} + \beta_{\text{Female}} + \beta_{\text{Free/reduced lunch}} + \beta_{\text{English learners}} + \beta_{\text{Quant CogAT}} + \beta_{\text{Verbal CogAT}} + \beta_{\text{Non-verbal CogAT}} + \beta_{\text{ITBS Math}} + \beta_{\text{ITBS Reading}} + \varepsilon$$

In Block 3, I analyzed the main effects when the TTCT is added with the student demographic variables. The following model was run and transformed to predicted probabilities:

$$\text{Logit}_{\text{Gifted Identification}} = \beta_0 + \beta_{\text{Black}} + \beta_{\text{Hispanic}} + \beta_{\text{Asian}} + \beta_{\text{Native American}} + \beta_{\text{Other}} + \beta_{\text{Female}} + \beta_{\text{Free/reduced lunch}} + \beta_{\text{English learners}} + \beta_{\text{Quant CogAT}} + \beta_{\text{Verbal CogAT}} + \beta_{\text{Non-verbal CogAT}} + \beta_{\text{ITBS Math}} + \beta_{\text{ITBS Reading}} + \beta_{\text{TTCT-Figural}} + \varepsilon$$

In Block 4, secondary effects were examined by adding interaction terms of the TTCT and the student demographic variables. The following model was run and transformed to predicted probabilities:

$$\text{Logit}_{\text{Gifted Identification}} = \beta_0 + \beta_{\text{Black}} + \beta_{\text{Hispanic}} + \beta_{\text{Asian}} + \beta_{\text{Native American}} + \beta_{\text{Other}} + \beta_{\text{Female}} + \beta_{\text{Free/reduced lunch}} + \beta_{\text{English learners}} + \beta_{\text{Quant CogAT}} + \beta_{\text{Verbal CogAT}} + \beta_{\text{Non-verbal CogAT}} + \beta_{\text{ITBS Math}} + \beta_{\text{ITBS Reading}} + \beta_{\text{TTCT-Figural}} + \beta_{\text{TTCT-Figural}} * \beta_{\text{Black}} + \beta_{\text{TTCT-Figural}} * \beta_{\text{Hispanic}} + \beta_{\text{TTCT-Figural}} * \beta_{\text{Asian}} + \beta_{\text{TTCT-Figural}} * \beta_{\text{Native American}} + \beta_{\text{TTCT-Figural}} * \beta_{\text{Other}} + \beta_{\text{TTCT-Figural}} * \beta_{\text{Female}} + \beta_{\text{TTCT-Figural}} * \beta_{\text{Free/reduced lunch}}\ \beta_{\text{TTCT-Figural}} * \beta_{\text{English learners}} + \varepsilon$$

1.2.5.3.3 Universal Screening

For students universally screened in first grade, Native American and students categorized as Other were removed from the model because there were no students identified for gifted services and there were severely inflated standard errors. The following model was run and transformed to predicted probabilities:

$$\text{Logit}_{\text{Gifted Identification}} = \beta_0 + \beta_{\text{Black}} + \beta_{\text{Hispanic}} + \beta_{\text{Asian}} + \beta_{\text{Female}} + \beta_{\text{Free/reduced lunch}} + \beta_{\text{English learners}} + \varepsilon$$

In Block 2, I examined the main effects of achievement and non-verbal cognitive ability measures to the probability of being identified as gifted. The following model was run and transformed to predicted probabilities:

$$\text{Logit}_{\text{Gifted Identification}} = \beta_0 + \beta_{\text{Black}} + \beta_{\text{Hispanic}} + \beta_{\text{Asian}} + \beta_{\text{Female}} + \beta_{\text{Free/reduced lunch}} + \beta_{\text{English learners}} + \beta_{\text{NNAT-2}} + \beta_{\text{Quant CogAT}} + \beta_{\text{Verbal CogAT}} + \beta_{\text{Non-verbal CogAT}} + \beta_{\text{ITBS Math}} + \beta_{\text{ITBS Reading}} + \varepsilon$$

In Block 3, I analyzed the main effects when the TTCT is added with the student demographic variables. The following model was run and transformed to predicted probabilities:

$$\text{Logit}_{\text{Gifted Identification}} = \beta_0 + \beta_{\text{Black}} + \beta_{\text{Hispanic}} + \beta_{\text{Asian}} + \beta_{\text{Female}} + \beta_{\text{Free/reduced lunch}} + \beta_{\text{English learners}} + \beta_{\text{NNAT-2}} + \beta_{\text{Quant CogAT}} + \beta_{\text{Verbal CogAT}} + \beta_{\text{Non-verbal CogAT}} + \beta_{\text{ITBS Math}} + \beta_{\text{ITBS Reading}} + \beta_{\text{TTCT-Figural}} + \varepsilon$$

In Block 4, secondary effects were examined by adding interaction terms of the TTCT and the student demographic variables. The following model was run and transformed to predicted probabilities:

$$\text{Logit}_{\text{Gifted Identification}} = \beta_0 + \beta_{\text{Black}} + \beta_{\text{Hispanic}} + \beta_{\text{Asian}} + \beta_{\text{Female}} + \beta_{\text{Free/reduced lunch}} + \beta_{\text{English learner s}} + \beta_{\text{NNAT-2}} + \beta_{\text{Quant CogAT}} + \beta_{\text{Verbal CogAT}} + \beta_{\text{Non-verbal CogAT}} + \beta_{\text{ITBS Math}} + \beta_{\text{ITBS Reading}} + \beta_{\text{TTCT-Figural}} + \beta_{\text{TTCT-Figural}} * \beta_{\text{Black}} + \beta_{\text{TTCT-Figural}} * \beta_{\text{Hispanic}} + \beta_{\text{TTCT-Figural}} * \beta_{\text{Asian}} + \beta_{\text{TTCT-Figural}} * \beta_{\text{Female}} + \beta_{\text{TTCT-Figural}} * \beta_{\text{Free/reduced lunch}} + \beta_{\text{TTCT-Figural}} * \beta_{\text{English learners}} + \varepsilon$$

1.2.5.4 Additional Analysis

Since there was indication of violation of independence, additional multilevel modeling was used to report differences to account for clustering between schools. More specifically, students were referred from 31 schools, and universally screened in first grade in 23 schools. More specifically, a hierarchical generalized model (i.e., multilevel logistic regression) was used to assess the relationship among student demographics (i.e., race/ethnicity, socioeconomic status, English learner status), cognitive ability measures, achievement scores, and the TTCT-Figural Form A or B with the probability of being identified as gifted across schools.

Prior to testing the multilevel model in Question 3 (i.e., Block 3), Level 1 predictors were explored for fourteen Level 1 variables that consisted of student demographics (i.e., sex, Black/African American, Hispanic, Asian, Other, English learner status, free-reduced lunch), scores on cognitive ability measures (i.e., CogAT7 Non-Verbal, CogAT7 Quantitative, CogAT7 Verbal), achievement scores (i.e., ITBS Reading, ITBS Math), and a single variable for the TTCT with the highest scores on either Form A or B. Again, for universal screening students, Level 1 predictors were explored for thirteen Level 1 variables that consisted of student demographics (i.e., females, Black/African American, Hispanic, Asian, English learner status, free-reduced lunch), scores on cognitive ability measures (i.e., NNAT-2, CogAT7 Non-Verbal, CogAT7 Quantitative, CogAT7 Verbal), achievement scores (i.e., ITBS Reading, ITBS Math), and a single variable for the TTCT with the highest scores on either Form A or B.

Initially, an unconditional model was tested with gifted identification across all 31 schools for the subset of data for students referred and then universally screened in first grade for all 23 elementary schools. Then, separate models using a hierarchical approach were used for conditional models.

1.2.5.4.1 Referrals

For students who were referred, in Model 1, a multilevel random intercepts model was used to assess student demographics (i.e., females, Black/African American, Hispanic, Asian, Other, English learner status, free-reduced lunch) to the probability of being identified for gifted services. In Model 2, a multilevel random intercepts model was used to assess student demographics from Block 1, and scores on cognitive ability measures (i.e., CogAT7 Non-Verbal, CogAT7 Quantitative, CogAT7 Verbal), achievement scores (i.e., ITBS Reading, ITBS Math), and a single variable for the TTCT with the highest scores on either Form A or B.

1.2.5.4.2 Universal Screening

For students who were universally screened, in Model 1, a random intercepts model was used to assess student demographics (i.e., Females, Black/African American, Hispanic, Asian, Other, English learner status, free-reduced lunch) to the probability of being identified for gifted services. In Model 2, a multilevel random intercepts model was used to assess student demographics, and scores on cognitive ability measures (i.e., NNAT-2, CogAT7 Non-Verbal, CogAT7 Quantitative, CogAT7 Verbal), achievement scores (i.e., ITBS Reading, ITBS Math), and a single variable for the TTCT with the highest scores on either Form A or B.

For both multilevel models, Level 2 variables include the schools (universal screening only consisted of 23 schools, whereas referred students consisted of all 31 schools) to solely examine the clusters between schools. Both fixed effects at each level, random effects, standard errors, Akaike Information Criteria (AIC) are reported (Finch et al., 2014; Gelman & Hill, 2007; Raudenbush & Bryk, 2002).

1.3 Results

Initial analyses began with obtaining descriptive and bivariate statistics for the total sample, as well as the subsets of the data for students referred and students universally screened in first grade. Specific descriptive statistics can be seen in Table 1.2 and Table 1.3. Please see Figure 1.2 for a correlation matrix for the entire sample. For the total sample, bar plots were created to see who was identified as gifted by race/ethnicity, sex, free/reduced lunch status, and English learner status (see Figure 1.3 to Figure 1.6).

1.3.1 Referrals

Using the ggplot2 package in R (Wickham et al., 2020), density plots for the TTCT Form A or B, TTCT Form A, TTCT Form B, CogAT7 Non-Verbal, CogAT Quantitative, CogAT Verbal, ITBS Reading, and ITBS Math were analyzed. An examination of the density plot for the TTCT Form A or B and skewness (-.16) provided indication of appropriate symmetry and resemblance of a normal distribution, also kurtosis (-.36) indicated the distribution was marginally mesokurtic; this was analogous to both Form A and B individually. The CogAT7 Verbal, CogAT7 Quantitative, and CogAT7 Non-Verbal Battery resembled a normal distribution. However, ITBS Math and ITBS Reading density plots were moderately skewed left (ITBS Reading = -.60; ITBS Math = - .79) and show indication of a ceiling effect. See Figure 1.7 to Figure 1.12 for density plots

As for bivariate statistics, the TTCT showed weak relationships with CogAT7 Non-Verbal, ($r(7) = .06$, $p < .05$), ITBS Math ($r(7) = .08$, $p < .01$), and being identified for gifted services ($r(7) = .13$, $p < .001$). The TTCT did not relate to scores on the CogAT7 Quantitative ($r(7) = .03$, $p = .378$), CogAT7 Verbal ($r(7) = .04$, $p = .185$), or ITBS Reading ($r(7) = .02$, $p = .457$). Cognitive ability and achievement scores did, however, show moderately strong significant correlations with one another. See Figure 1.13 for more details.

1.3.2 Universal Screening

Density plots for the TTCT Form A or B, TTCT Form A, TTCT Form B, CogAT7 Non-Verbal, CogAT Quantitative, CogAT Verbal, ITBS Reading, and ITBS Math were analyzed. For students who were universally screened in first grade, an examination of the density plot, skewness (-.30), and kurtosis (-.27) for TTCT Form A or B provided indication of symmetry and resemblance of a normal distribution.

Similarly, scores on the NNAT-2, CogAT7 Verbal, CogAT7Quantitative, and CogAT7 Non-Verbal Battery resembled a normal distribution. The ITBS Reading density plots were skewed left (ITBS Reading = -.84) and the ITBS Math was moderately skewed left (-.42). Both ITBS Reading and Math show indication of a ceiling effect (see Figure 1.14 to Figure 1.20).

For universally screened students in first grade, the TTCT Form A or B showed weak relationships with CogAT Non-Verbal ($r(8) =.09$, $p < .001$), ITBS Reading ($r(8) = .07$, $p < .001$), ITBS Math (r(8) = -.05, $p < .05$), NNAT-2 (r(8) = .13, $p < .001$), and being identified for gifted services (r(8) = .09, $p < .001$). The TTCT did not relate to scores on the CogAT7 Quantitative (r(8) = -.03, $p = .132$) or CogAT7 Verbal (r(8) = -.02, $p = .306$). Like the students referred, the cognitive ability and achievement scores did show moderately strong correlations. See Figure 1.21 for more details.

1.3.3 Question 1: Demographic Subgroup Differences by Test

1.3.3.1 Referrals

For students who were referred in Phase 1, a series of multiple regressions and simple linear regressions were conducted by each test after each test was converted to standardized scale (i.e., Z-score) for comparison. I predicted the Z-score of each test from a series of dummy codes for race/ethnicity, sex, English learner status, and free/reduced lunch status.

1.3.3.1.1 Race

For referrals, a multiple regression was used to assess if race/ethnicity predicted scores on the TTCT (Form A or B), CogAT7 Non-Verbal, CogAT7 Quantitative, CogAT7 Verbal, ITBS Reading, and ITBS Math. See Table 1.4 for more details based on student race/ethnicity per test.

- TCT Form A or B: The overall results of the regression indicated only one race/ethnicity explained variance in scores on the TTCT Form A or B, but the overall model showed a small effect ($R^2 = .016$, F[5, 1185] = 3.76, $p < .001$). Students who are Hispanic/Latinx (B = -0.21, 95% CI [-.36, -.07], $p < .01$) scored .21 standard deviations lower than students who are White and accounted for 39.2% of the overall effect (1.6%) in scores.
- CogAT7 Non-Verbal: The overall results of the regression indicated only one race/ethnicity explained variance in scores on the CogAT7 Non-Verbal Battery, however the overall model showed a small effect ($R^2 = .020$, F[5, 1185] = 4.80 , $p < .001$) and only showed

.004 (ΔR^2) increase from the TTCT Form A or B model. Students who are Black/African American (B = -.36, 95% CI [-.19, -.07], $p < .001$) scored .36 standard deviations lower than students who are White and accounted for 70.3% of the (2%) overall effect in scores.

- CogAT7 Quantitative: The overall results of the regression indicated all predictors for race/ethnicity explained variance in scores on the CogAT7 Quantitative Battery, but the overall model showed a small effect (R^2 = .034, F[5, 1185] = 8.227, $p < .001$). Students who are Black/African American (B = -.31, 95% CI [-.47, -.14], $p < .001$), Hispanic/Latinx (B = -.21, 95% CI [-.36, -.07], $p < .01$), Asian (B = .36, 95% CI [.10, .62], $p < .01$), Native American (B = -.56, 95% CI [-0.94, -0.18], $p < .001$), and students combined into the Other category (B = -.45, 95% CI [-.79, -.10], $p < .01$) indicated statistically significant differences in scores on the CogAT7 Quantitative. Students are Black/African American, scored .31 standard deviations lower and Hispanic/Latinx scored .21 standard deviations lower than White students. Similarly, Native American students scored .56 standard deviations lower, and students combined into the Other category scored .45 standard deviations lower than White students. Conversely, students who are Asian scored .36 standard deviations higher than White students. All predictors of race/ethnicity contributed to the overall effect in scores on the CogAT7 Quantitative battery (92.8%) and of the variance (3%) in scores.

- CogAT7 Verbal: The overall results of the regression indicated only one race/ethnicity explained variance in scores on the CogAT7 Verbal Battery, however the overall model showed a small effect (R^2 = .06, F[5, 1185] = 14.14, $p < .001$). Students who are Hispanic/Latinx (B = -.59, 95% CI [-.74, -.45], $p < .001$) scored .59 standard deviations lower than White students and accounted for 79% of the overall effect (6%) in scoresThe ΔR^2 indicated that there was a .04 increase from the model with TTCT Form A or B.

- ITBS Reading: The overall results of the regression indicated all race/ethnicities explained variance in scores on the ITBS Reading, the overall model showed a small effect (R^2 = .04, F[5, 1185] = 9.32, $p < .001$), and showed .022 (ΔR^2) increase from the TTCT Form A or B model. Students who are Black/African American (B = -.16, 95% CI [-0.60, -0.27], $p < .001$), Hispanic/Latinx (B = -.10, 95% CI [-0.39, -0.10], $p < .01$), Asian (B = -.12, 95% CI [-.78, -.27], $p < .001$), Native American (B = -.06, 95% CI [-0.78, -0.03], $p < .05$), and students combined in the Other category (B = -.08, 95% CI [-0.83, -0.14], $p < .01$) indicated statistically significant differences in scores on ITBS Reading. Students who are Black/African American scored .16 standard deviations lower, whereas Hispanic/Latinx students scored .10 standard deviations lower than White students. Similarly, Native American students scored .06 standard deviations lower, students combined into the Other category scored .08 standard deviations lower, and students who are Asian scored .12 standard deviations lower than White students. All predictors of race/ethnicity contributed to variance in scores on the ITBS Reading and accounted for 67.7% of the overall effect (4%) in scores.

- ITBS Math: The overall results of the regression indicated two race/ethnicities explained variance in scores on the ITBS Math, the overall model showed a small effect (R^2 = .03, F[5, 1185] = 6.14, $p < .01$). Students who are Black/African American (B = -.11, 95% CI [-.48, -.15], $p < .001$) and Asian (B = .08, 95% CI [.09, .61], $p < .05$) indicated statistically significant differences in scores on the ITBS Math. Students Black/African American scored .11 standard deviations lower than White students, whereas Asian students scored .08 standard deviations higher than White students. Both Black/African American and Asian students accounted for 79.4% of the overall effect (3%) in scores.

1.3.3.1.2 English Learner

A simple linear regression was used to assess if English learner status predicted scores on the TTCT (Form A or B), CogAT7 Non-Verbal, CogAT7 Quantitative, CogAT7 Verbal, ITBS Reading, and ITBS Math compared to native English speakers. See Table 1.5 for more details based on student English learner status for referrals.

- TTCT Form A or B: The overall results of the regression indicated English learners explained variance in scores on the TTCT Form A or B, but the overall model showed a small effect (R^2 = .007, F[1, 1189] = 0.851, p < .001). English learners (B = -.24, 95% CI [-.40, -.08], p < .01) scored -.24 lower than native English speakers.

- CogAT7 Non-Verbal: The overall results of the regression indicated English learners did not explain variance in scores on the CogAT7 Non-verbal Battery, (R^2 = .0007, F[1, 1189] = .851, p = .357). Thus, English learners (B = -.08, 95% CI [-.24, .09], p = .712) showed no significant differences in scores compared to native English speakers.

- CogAT7 Quantitative: The overall results of the regression indicated English learners did not explain variance in scores on the CogAT7 Quantitative Battery, (R^2 = .002, F[1, 1189] = 2.361, p = .125). English learners (B = -.13, 95% CI [-.29, .03], p = .250) showed no significant differences in scores compared to native English speakers.

- CogAT7 Verbal: The overall results of the regression indicated English learners explained variance in scores on the CogAT7 Verbal Battery, but the overall model showed a small effect (R^2 = .09, F[1, 1189] = 112.2, p < .001). English learners (B = -.83, 95% CI [-.99, -.68], p < .001) scored .83 standard deviations lower than native English speakers. The ΔR^2 indicated a .079 increase from the model with TTCT Form A or B.

- ITBS Reading: The overall results of the regression indicated English learners explained variance in scores on the ITBS Reading, but the overall model showed a small effect (R^2 = .02, F[1, 1189] = 23.1, p < .001). English learners (B = -.39, 95% CI [-.55, -.23], p < .001) scored .39 standard deviations lower than native English speakers.

- ITBS Math: The overall results of the regression indicated English learners did not explain variance in scores on the ITBS Math (R^2 = .001, F[1, 1189] = 1.43, p = .232). English learners (B = .09, 95% CI [-.06, .26], p = .464) showed no significant differences in scores compared to native English speakers.

1.3.3.1.3 Sex

A simple linear regression was used to assess if sex (i.e., male, female) predicted scores on the TTCT (Form A or B), CogAT7 Non-Verbal, CogAT7 Quantitative, CogAT7 Verbal, ITBS Reading, and ITBS Math. See Table 1.6 for more details based on students based on sex.

- TTCT Form A or B: The overall results of the regression indicated females explained variance in scores on the TTCT Form A or B, but the overall model showed a small effect (R^2 = .02, F[1, 1189] = 19.69, p < .001). Females (B = .25, 95% CI [.14, .37], p < .001) scored .25 standard deviations higher than male students.

- CogAT7 Non-Verbal: The overall results of the regression indicated females did not explain variance in scores on the CogAT7 Non-Verbal (R^2 = .0003, F[1, 1189] = .324, p = .570). Females (B = -.03, 95% CI [-.14, .08], p = .692) showed no significant differences in scores compared to male students.

- CogAT7 Quantitative: The overall results of the regression indicated females explained variance in scores on the CogAT7 Quantitative Battery, but the overall model showed a small effect (R^2 = .019, F[1, 1189] = 23.49, p < .001). Females (B = -.28, 95% CI [-.39, -.17], p < .001) scored .28 standard deviations lower than male students.

- CogAT7 Verbal: The overall results of the regression indicated females explained variance in scores on the CogAT7 Non-Verbal, but the overall model showed a small effect (R^2 = .0003, F[1, 1189] = .324, p = .570). Females (B = .13, 95% CI [.02, .24], $p < .05$) scored .13 standard deviations higher than male students.

- ITBS Reading: The overall results of the regression indicated females explained variance in scores on the ITBS Reading, however the overall model showed a small effect (R^2 = .005, F[1, 1189] = 6.20, $p < .05$). Females (B = .14, 95% CI [.03, .26], $p < .05$) scored .14 standard deviations higher than male students. Conversely, male students showed no difference in scores.

- ITBS Math: The overall results of the regression indicated females explained variance in scores on the ITBS Math, however the overall model showed a small effect (R^2 = .014, F[1, 1189] = 16.98, $p < .001$). Females (B = -.24, 95% CI [-.35, -.12], $p < .001$) scored .24 standard deviations lower than male students.

1.3.3.1.4 Free/Reduced Lunch Status

A simple linear regression was used to assess if free/reduced lunch status predicted scores on the TTCT (Form A or B), CogAT7 Non-Verbal, CogAT7 Quantitative, CogAT7 Verbal, ITBS Reading, and ITBS Math compared to students who paid full price for lunch. See Table 1.7 for more details based on students based on free/reduced lunch status.

- TTCT Form A or B: The overall results of the regression indicated students on free/reduced lunch explained variance in scores on the TTCT Form A or B, but the overall model showed a small effect (R^2 = .005, F[1, 1189] = 5.88, $p < .05$). Students on free/reduced lunch (B = -.15, 95% CI [-.27, -.03], $p < .05$) scored .15 standard deviations lower than students who were not on free/reduced lunch. Alternatively, students who pay full lunch showed no difference in scores.

- CogAT7 Non-Verbal: The overall results of the regression indicated student on free/reduced lunch explained variance in scores on the CogAT7 Non-Verbal Battery, however the model showed an overall small effect (R^2 = .009, F[1, 1189] = 10.85, p < .01). Students on free/reduced lunch (B = -.20, 95% CI [-.32, -.08], p < .001) scored .20 standard deviations lower than students who were not on free/reduced lunch.

- CogAT7 Quantitative: The overall results of the regression indicated students on free/reduced lunch explained variance in scores on the CogAT7 Quantitative Battery, but the overall model showed a small effect (R^2 = .025, F[5, 1185] = 30.97 , p < .001). Students on free/reduced lunch (B = -.34, 95% CI [-.45, -.22], p < .001) scored .34 standard deviations lower than students who were not on free/reduced lunch.

- CogAT7 Verbal: The overall results of the regression indicated students on free/reduced lunch explained variance in scores on the CogAT7 Verbal Battery, but the overall model showed a small effect (R^2 = .05, F[1, 1189] = 65.79, p < .001). Students on free/reduced lunch (B = -.48, 95% CI [-.60, -.37], p < .001) scored .48 standard deviations lower than students who were not on free/reduced lunch

- ITBS Reading: The overall results of the regression indicated free/reduced lunch explained variance in scores on the ITBS Reading, but the overall model showed a small effect (R^2 = .021, F[1, 1189] = 25.9, p < .001). Students on free/reduced lunch (B = -.31, 95% CI [-.43, -.19], p < .001) scored .31 standard deviations lower than students who were not on free/reduced lunch.

- ITBS Math: The overall results of the regression indicated free/reduced lunch explained variance in scores on the ITBS Math, however the overall model showed a small effect (R^2 = .006, F[1, 1189] = 6.955, p < .001). Students on free/reduced lunch (B = -.16, 95% CI [-.28, -.04], p < .05) scored .16 standard deviations lower than students who were not on free/reduced lunch.

1.3.3.2 Universal Screening

For students who were universally screened in Phase 1, a series of multiple regressions and simple linear regressions were conducted for each test used in the process after each test was standardized (i.e., converted to Z-score). For conciseness, the NNAT-2 and TTCT Form A or B is reported below per demographic subgroup who were universally screened.

1.3.3.2.1 Race

A multiple regression was used to assess if race/ethnicity predicted scores on the the NNAT-2 and the TTCT Form A or B for students universally screened in first grade. See Table 1.8 for more details based on student race/ethnicity for each test used for formal identification.

- NNAT-2: The overall results of the regression indicated four categories of race/ethnicity explained variance in scores on the NNAT-2. The overall model showed a small effect (R^2 = .08, F[5, 2064] = 33.26, p < .001). Students who are Black/African American (B = -.66, 95% CI [-.78, -54], , p < .001), Hispanic/Latinx (B = -.37, 95% CI [-.47, -.28], p < .001), Asian (B = .30, 95% CI [.09, .52], p < .05), and Native American (B = -.40, 95% CI [-.70, -.10], p < .05) showed statistically significant differences in scores. Students who are Black/African American scored .61 standard deviations lower than White students. Similarly, Hispanic/Latinx students scored .37 standard deviations lower and Native American students scored .40 standard deviations lower than White students. Alternatively, students who are Asian scored .30 standard deviations higher than White students. All four predictors accounted for 82.8% of the overall effect (8% of variance) in scores.

- TTCT Form A or B: The overall results of the regression indicated three categories of race/ethnicity explained variance in scores on the TTCT Form A or B, but the overall model showed a small effect (R^2 = .027, F[5, 2064] = 11.85, p < .001). Students who are Black/African American (B = .24, 95% CI [.12, .36], p < .001) scored .24 standard deviations higher than White students. Conversely, Hispanic/Latinx students (B = -.20, 95% CI [-.30, -.10], p < .01) scored .20 standard deviations lower, as did Asian students (B = -.42, 95% CI [-.65, -.20], p < .01) who scored .42 standard deviations lower than White students. All three predictors accounted for 137.3% of the overall effect (2%) in scores.

1.3.3.2.2 English Learner

A simple linear regression was used to assess if English learner status predicted scores on the NNAT-2 and the TTCT Form A or B. See Table 1.9 for more details for English learner status for students universally screened in first grade.

- NNAT-2: The overall results of the regression indicated English learners explained variance in scores, however the overall model showed a small effect (R^2 = .007, F[1, 2068] = 71.19, p < .001). English learners (B = -.23, 95% CI [-.35, -.11], p < .001) scored .23 standard deviations lower than native English speakers.

- TTCT Form A or B: The overall results of the regression indicated English learners explained variance in scores, but the overall model showed a small effect (R^2 = .033, F[1, 2068] = 71.19, p < .001). English learners (B = -.50, 95% CI [-.62, -.38], p < .001) scored .50 standard deviations lower than native English speakers.

1.3.3.2.3 Sex

A simple linear regression was used to assess if sex (i.e., male, female) predicted scores on NNAT-2 and the TTCT Form A or B. See Table 1.10 for more details based on students based on sex.

- TTCT Form A or B. The overall results of the regression indicated females explained variance in scores, but the overall model showed a small effect (R^2 = .22, F[1, 2068] = 25.93, $p <$ .001). Females (B = .22, 95% CI [.14, .31], $p < .001$) scored .22 standard deviations higher than male students.

- NNAT-2. The overall results of the regression indicated females did not explain variance in scores (R^2 = .001, F[1, 2068] = 2.11, $p = .147$). There was no statistically significant difference for female students (B = .06, 95% CI [-.02, .15], $p = .292$) compared to male students.

1.3.3.2.4 Free/Reduced Lunch Status

A simple linear regression was used to assess if free/reduced lunch status predicted scores on the NNAT-2 and the TTCT Form A or B. See Table 1.11 for more details on free/reduced lunch status for students universally screened.

- NNAT-2. The overall results of the regression indicated only one race/ethnicity explained variance in scores, however the overall model showed a small effect (R^2 = .06, F[1, 2068] = 142.4, $p < .001$). Students on free/reduced lunch (B = -.52, 95% CI [-.61, -.44], $p < .001$) scored .52 standard deviations lower than students not on free/reduced lunch.

- TTCT Form A or B. The overall results of the regression indicated students on free/reduced lunch did not explain variance in scores (R^2 = .0005, F[1, 2068] = 1.193, $p = .275$). There was no statistically significant difference in students on free/reduced lunch (B =.05, 95% CI [-.04, .14], $p = .496$) compared to students not on free/reduced lunch.

1.3.3.3 Additional Analysis: Hierarchical Regression for TTCT by Form

1.3.3.3.1 Referrals

For students who were referred, a hierarchical regression analysis was conducted to analyze the relationship among student demographics (i.e., Black/African American, Hispanic/Latinx, Asian, American Indian/Alaska Native, Other, sex, English learner status, and free/reduced lunch), scores on the CogAT7 (i.e., Non-Verbal, Quantitative, Verbal), and scores on the ITBS (i.e., Reading, Math) to scores on the TTCT by Form A, Form B, or a combined column of the highest scores on either Form A or Form B. For each regression, squared structure coefficients were used, in addition to standardized β coefficients to evaluate the overall amount of variance that predictors explain of the overall R^2 effect size without the effect of other variables within the model (Yeatts et al., 2017). See Table 1.12 for details.

- TTCT Form A. In Block 1, all demographic control variables were added and the main effects of Black/African American, Hispanic/Latinx, Asian, American Indian/Alaska Native, Other, sex, English learner status, and free/reduced lunch were assessed. The overall results of the regression indicated one predictor explained variance in scores over and above the other predictors; however, the overall model showed a small effect (R^2 = .032, F[8, 1182] = 4.904, p < .001). Students who are female (β = .11, p < .001) showed a weak positive relationship and explained 64.6% (r^2 = .646) of the overall effect size obtained (3% of variance).

In Block 2, in addition to the demographic control variables, the CogAT7 (i.e., Non-Verbal, Quantitative, Verbal) and ITBS (Reading, Math) were added to the model. The overall results of the regression indicated four predictors explained variance in scores. The overall model showed a small effect (R^2 = .042, F[13, 1177] = 3.972, p < .001, ΔR^2 = .01) and showed marginal improvement to the first model with demographic control variables. Students who are female (β = .12, p < .01) and scores on the ITBS Math (β = .11, p < .05) showed a weak positive relationship with scores, and contributed 44.4% to the overall obtained effect (4.2% variance.

- TTCT Form B. In Block 1, all demographic control variables were added and Black/African American, Hispanic/Latinx, Asian, American Indian/Alaska Native, Other, sex, English learner status, and free/reduced lunch were assessed. The overall results of the regression indicated one predictor explained variance in scores over and above the other predictors; however, the overall model showed a small effect (R^2 = .031, F[8, 1182] = 4.687, p < .001). Similar to Form A, students who are female (β = .13, p < .001) showed a weak positive relationship with scores. Additionally, females (r^2 = .569) explain 56.9% of the overall effect size obtained (3.1% variance).

In Block 2, in addition to the demographic control variables, the CogAT7 (i.e., Non-Verbal, Quantitative, Verbal) and ITBS (Reading, Math) were added to the model. The overall results of the regression indicated two predictors explained variance in scores over and above the other predictors. The overall model showed a small effect (R^2 = .046, F[13, 1177] = 4.403, p < .001, ΔR^2 = .015), and showed negligible improvement to the first model with demographic control variables. Students who are female (β = .15, p < .001) and scores on the ITBS Math (β = .13, p < .01) showed a weak positive relationship and explained 64.4% of the overall effect obtained (4.6% of variance).

- TTCT Form A or B. In Block 1, all demographic control variables were added and the main effects of Black/African American, Hispanic/Latinx, Asian, American Indian/Alaska Native, Other, sex, English learner status, and free/reduced lunch were assessed. The overall results of the regression indicated one predictor explained variance in scores over and above the other predictors; however, the overall model showed a small effect (R^2 = .035, F[8, 1182] =5.408, p < .001). Again, students who are female (β = .12, p < .001) had a weak positive relationship and explained 46.1% of the overall effect size obtained (3.5% of variance).

In Block 2, in addition to the demographic control variables, the CogAT7 (i.e., Non-Verbal, Quantitative, Verbal) and ITBS (Reading, Math) were added to the model. The overall results of the regression indicated two predictors explained variance in scores over and above the other predictors. The overall model showed a small effect (R^2 = .046, F[13, 1177] = 4.413, $p <$ .001, ΔR^2 = .011) and showed marginal improvement to the first model with demographic control variables. Students who are female (β = .12, $p < .001$) and scores on the ITBS Math (β = .11, $p <$ .001) showed a weak positive relationship and explained 49.5% of the overall effect obtained (4.6% of variance).

1.3.3.3.2 Universal Screening

To assess differences for students who were universally screened, a hierarchical regression analysis was conducted to analyze the relationship among student demographics (i.e., Black/African American, Hispanic/Latinx, Asian, American Indian/Alaska Native, Other, sex, English learner status, and free/reduced lunch), scores on the CogAT7 (i.e., Non-Verbal, Quantitative, Verbal), scores on the ITBS (i.e., Reading, Math), and scores on the NNAT-2 to scores on the TTCT by Form A, Form B, or a combined column of the highest scores on either Form A or Form B. See Table 1.13.

- TTCT Form A: In Block 1, all demographic control variables were added and the main effects of Black/African American, Hispanic/Latinx, Asian, American Indian/Alaska Native, Other, sex, English learner status, and free/reduced lunch were assessed. The overall results of the regression indicated four predictors explained variance in scores over and above the other predictors, however the overall model showed a small effect (R^2 = .056, F[8, 2061] = 15.16, $p < .001$). Students who are female (β = .09, $p < .001$), Asian (β = -.05, $p < .001$), English learners (β = -.19 , $p < .001$), and on free/reduced lunch (β = .10, $p < .001$) showed weak positive relationships and explain 84.6% of the overall effect size obtained (5.6% of variance).

In Block 2, in addition to the demographic control variables, the CogAT7 (i.e., Non-Verbal, Quantitative, Verbal) and ITBS (Reading, Math) were added to the model. The overall results of the regression indicated nine predictors explained variance in scores over and above the other predictors. The overall model showed a small effect (R^2 = .12, F[14, 2055] = 20.09, $p < .001$, ΔR^2 = .064) and showed improvement to the first model with demographic control variables. Students who are female (β = .10, $p < .001$), Black/African American students (β = .06, $p < .05$), on free/reduced lunch (β = .12, $p < .001$), as well as scores on the CogAT Non-Verbal (β =.05, $p < .001$), ITBS Reading (β = .05, $p < .05$), and the NNAT-2 (β = .25, $p < .001$) all showed positive relationships to scores. Conversely, Asian (β = -.07, $p < .01$), English learners (β = -.18, $p < .001$), and scores on the CogAT Verbal (β = -.11, $p < .001$) and CogAT Quantitative (β = -.09, p<.001) showed a weak negative relationship to scores. These nine predictors explained 85.5% of the overall effect size obtained (12% of variance).

- TTCT Form B: In Block 1, all demographic control variables were added and the main effects of Black/African American, Hispanic/Latinx, Asian, American Indian/Alaska Native, Other, sex, English learner status, and free/reduced lunch were assessed. The overall results of the regression indicated five predictors explained variance in scores over and above the other predictors. The overall model showed a small effect (R^2 = .117, F[8, 2061] = 34.19, $p < .001$). Students who are female (β = .09, $p < .001$), on free/reduced lunch (β = .10, $p < .001$), and combined into the category of Other (β = .07, $p < .001$) showed weak positive relationships. Alternatively, students who are Hispanic/Latinx (β = -.21, $p < .001$) and English learners (β = -.19, $p < .001$) showed negative relationships. Together, all five predictors explained 88.5% of the overall effect size obtained (11.7% of variance).

In Block 2, in addition to the demographic control variables, the CogAT7 (i.e., Non-Verbal, Quantitative, Verbal) and ITBS (Reading, Math) were added to the model. The overall results of the regression indicated nine predictors explained variance in scores over and above the other predictors. The overall model showed a large effect (R^2 = .596, F[14, 2055] = 216.8, p < .001, ΔR^2 = .479), and showed a drastic improvement to the first model with demographic control variables. Students who are Black/African American (β = .17, p < .001), English learners (β = .11, p < .001), on free/reduced lunch (β = .07, p < .001), as well as scores on the CogAT Non-Verbal (β = .39, p < .001), and the NNAT-2 (β = .59, p < .001) all showed a positive relationship to scores. Conversely, students who are female (β = -.28, p < .001), Hispanic/Latinx (β = -.09, p < .001), as well as scores on the CogAT Verbal (β = .39, p < .001), CogAT Quantitative (β = -.16, p < .001) and ITBS Reading (β = -.48, p < .001) showed negative relationships to scores. These nine predictors explained 81.5% of the overall effect size obtained (59.6% of variance).

- TTCT Form A or B: In Block 1, all demographic control variables were added and the main effects of Black/African American, Hispanic/Latinx, Asian, American Indian/Alaska Native, Other, sex, English learner status, and free/reduced lunch were assessed. The overall results of the regression indicated four predictors explained variance in scores over and above the other predictors, however the overall model showed a small effect (R^2 = .066, F[8, 2061] = 18.11, p < .001). Students who are female (β = .11, p < .001), Black/African American (β = .07, p < .05), or on free/reduced lunch (β = .09, p < .01) showed weak positive relationships, whereas students who are English learners (β = -.19, p < .001) showed negative relationships. All five predictors explain 94.2% of the overall effect size obtained (6.6% of variance).

In Block 2, in addition to the demographic control variables, the CogAT7 (i.e., Non-Verbal, Quantitative, Verbal) and ITBS (Reading, Math) were added to the model. The overall results of the regression indicated eight predictors explained variance in scores over and above the other predictors. The overall model showed a small effect (R^2 = .110, F[13, 2056] = 18.12, $p < .001$, ΔR^2 = .04), and showed marginal improvement to the first model with demographic control variables. Students who are female (β = .12, $p < .001$), Black/African American (β = .10, $p < .001$), on free/reduced lunch (β = .11, $p < .01$), as well as scores on the NNAT-2 (β = .21, $p < .001$) all showed a weak positive relationship to scores. Conversely, Asian (β = -.07, $p < .01$), English learners (β = -.17, $p < .001$), and scores on the CogAT Verbal (β = -.11, $p < .001$) and CogAT Quantitative (β = -.09, $p < .001$) showed a weak negative relationship to scores. These nine predictors explained 79.2% of the overall effect size obtained (11% of variance).

From model comparisons of the TTCT Form A, Form B, and Form A or B, there were notable differences between Form A and Form B. The Rosemary School District predominantly uses TTCT Form A, and will give Form B as an alternative to Form A. Since the district would use either Form A or B in their decision-making process for identification, the remainder of analyses use the TTCT Form A or B as a predictor variable.

1.3.4 Question 2: Representation Indices and Combination Rules

1.3.4.1 Demographic Proportionality

Within the 2018-2019 school year, Native Americans were represented more than 1.5 times as frequent compared to the total population at Rosemary School District, whereas Native Hawaiian/Pacific Islander students are equally proportional (i.e., 1.0). Unsurprisingly, White and Asian students are well-represented in gifted programs compared to the total population. However, there are still areas for improvement. Black/African American, Hispanic/Latinx, students on free/reduced lunch, and English learners are disproportionately represented within Rosemary School District's gifted program. For example, English learners are about half as frequently represented in gifted programs compared to the total population of English learners in the district. This is also the case with Hispanic/Latinx students and students on free/reduced lunch. More drastically, students who are two or more races have a Representation index (RI) of .14, which shows severe disproportionality to total students in the district who identify as two or more races. See Table 1.14 for all representation indices.

1.3.4.2 Combination Rules

Using all students who were universally screened or referred for the formal identification process (i.e., using the CogAT7 [Non-Verbal, Quantitative, Verbal], ITBS [Reading, Math], TTCT [Form A or B]), hypothetical combination rules were applied to account for how the district could combine multiple criteria (see Table 1.15). Upon initial evaluation, the AND rule was the most restrictive rule since it only identified a single White female student from this sample (n = 1). Unsurprisingly, after filtering the sample for students who would meet an OR rule based on the district's pre-determined cut-scores, this created a larger diverse pool of students who could be identified for gifted services (n = 1175). For instance, 36% of students who were tested would be identified with the implementation of an OR rule. The number of Hispanic/Latinx students would be 24% identified for gifted services if an OR rule was used, this would be an improvement from the 18% Hispanic/Latinx students identified in the 2018-2019 identification process. Also, there would be higher representation of Native American students, students in the Other category (i.e., Native Hawaiian/Pacific Islander, Two or more races), English learners, and students on free/reduced lunch. In applying an AVERAGE rule at the 95th percentile cut score, this created a pool of students of (n = 119). The average rule helped to increase the number of Black/African American students that would be identified (12%), but subsequently, also increased the number of White students (73%) and Asian (6.72%). The AVERAGE rule shows increased improvement from the AND rule but remains too stringent at the 95th percentile compared to the OR rule.

1.3.5 Question 3 and 4: Hierarchical Generalized Linear Regression Analysis

1.3.5.1 Referrals

In Block 1, a logistic regression analysis was conducted to investigate the overall main effects of race, sex, English learner status, and free/reduced lunch status to the probability of being identified for gifted services at Rosemary School District among students who were referred. Students who are Asian or who are English learners were found to contribute to the model ($\chi^2(6, N = 1191) = 97.5, p < .001$), and the overall model found a small effect, Tjur R^2 = .039. English learners had a lower probability (30%) of being identified for gifted services. Analogous to probability, odds ratio estimates show English learners (Exp(b) = .43, *SE* = .27, Wald = -3.084, 95% CI [.25, .73]) had .43 lower odds in being identified for gifted services than students who were White, male, paid full price lunch, and were native English speakers. Conversely, Asian students (Exp(b) = 2.63, *SE* = .29, Wald = 3.297, 95% CI [1.48, 4.66]) were more likely to be identified with over double the odds. Students who are Black/African American, Hispanic/Latinx, female, or on free/reduced lunch showed no differences in the probability of being identified for gifted services.

In Block 2, in addition to the demographic variables, cognitive ability (i.e., CogAT7 [Verbal, Quantitative, Non-Verbal]) and achievement tests (i.e., ITBS Reading, ITBS Math) were added to control for the effects of tests on the probability of being identified for gifted services. Students who are Asian, and student performance on the CogAT Non-Verbal, CogAT

Quantitative, CogAT Verbal, ITBS Reading, and ITBS Math, were found to contribute to the model ($\chi^2(13, N = 1191) = 248.7, p < .001$), and the overall model found a large effect, Tjur R^2 = .457. After controlling for tests, Asian students had four and half times the odds (Exp(b) = 4.55, *SE* = .42, Wald = 3.638, 95%CI [2.01, 10.28]) in being identified for gifted services than students who were White, male, paid full price lunch, and were native English speakers. Students who are English learners, Black/African American, Hispanic/Latinx, female, or on free/reduced lunch showed no differences in the probability of being identified for gifted services.

In Block 3, in addition to the student demographic variables, cognitive ability indices (i.e., CogAT7 [Verbal, Quantitative, Non-Verbal]), achievement test scores (i.e., ITBS Reading, ITBS Math), and scores on the TTCT Form A or B were added to the model. Students who are Asian and student performance on the TTCT Form A or B, CogAT Non-Verbal, CogAT Quantitative, CogAT Verbal, ITBS Reading, and ITBS Math were found to contribute to the model ($\chi^2(14, N = 1191) = 247.6, p < .001$), and the overall model found a large effect, Tjur R^2 = .476. Analogous to the previous model, Asian students still had four times the odds (Exp(b) = 4.40, *SE* = .42, Wald = 3.507, 95%CI [1.92, 10.07]) of being identified for gifted services than students who were White, male, paid full price lunch, and were native English speakers. Students who are Black/African American, Hispanic/Latinx, female, or on free/reduced lunch showed no differences in the probability of being identified for gifted services. All of the tests used for formal identification contributed a little over 50% to the probability of being identified. After adding the TTCT Form A or B, the model improved marginally ($\Delta R^2 = .019$). See Figure 1.22 to Figure 1.27 for the probabilistic relationship by each test.

In Block 4, continuous variables were centered (i.e., all tests) and interaction terms with demographic variables and the TTCT were added to the model. No demographic differences were found. However, students who took the CogAT7 Non-Verbal, CogAT Quantitative, CogAT Verbal, ITBS Reading, and ITBS Math were found to contribute to the model ($\chi^2(23, N = 1191) = 264.5, p < .001$), and the overall model found a large effect, Tjur $R^2 = .483$. The model showed that scores on the TTCT Form A or B showed no significant difference in the probability of being identified, but the conditional effects of mean-centered scores on the CogAT7 Non-Verbal, CogAT Quantitative, CogAT Verbal, ITBS Reading, and ITBS Math contributed over 50% to the probability of being identified for gifted services. There were no interaction effects found. See Table 1.16 for more details.

1.3.5.2 Universal Screening

In Block 1, a logistic regression analysis was conducted to investigate the overall main effects of race/ethnicity, sex, English learner status, free/reduced lunch status to the probability of being identified for gifted services at Rosemary School District for students who were universally screened in first grade. Students on free/reduced lunch were found to contribute to the model ($\chi^2(6, N = 2070) = 697.8, p < .001$), and the overall model found a small effect, Tjur $R^2 = .023$. Students on free/reduced lunch had .36 lower odds (Exp(b) = .36, *SE* =.27, Wald = -3.783, 95%CI [.21, .61]) than students who were White, male, paid full price lunch, and were native English speakers of being identified for gifted services. Students who are Black/African American, Asian, Hispanic/Latinx, female, or English learners showed no differences in the probability of being identified for gifted services.

In Block 2, in addition to the demographic variables, cognitive ability (i.e., NNAT-2, CogAT7 [Verbal, Quantitative, Non-Verbal]) and achievement tests (i.e., ITBS Reading, ITBS Math) were added to control for the effects of tests on the probability of being identified for gifted services. Performance on the NNAT-2, CogAT Quantitative, CogAT Verbal, and ITBS

Reading, as well as female students contributed to the overall model ($\chi^2(12, N = 2070) = 207.7$, $p < .001$), and found a large effect, Tjur $R^2 = .395$. After controlling for tests, female students had .48 lower odds (Exp(b) = .48, *SE* = .24, Wald = -2.985, 95% CI [.30, .78]) than students who were White, male, paid full price lunch, and were native English speakers of being identified for gifted services. Students who are Black/African American, Asian, Hispanic/Latinx, on free/reduced lunch, English learners, as well as scores on the CogAT Non-Verbal and ITBS Math, showed no differences in the probability of being identified for gifted services.

In Block 3, in addition to the student demographics variables, cognitive ability indices (i.e., NNAT-2, CogAT7 [Verbal, Quantitative, Non-Verbal]), achievement test scores (i.e., ITBS Reading, ITBS Math), and scores on the TTCT Form A or B were added to the model. Performance on the TTCT Form A or B, NNAT-2, CogAT Quantitative, CogAT Verbal, ITBS Reading, as well as students who are Asian or female were found to contribute to the model ($\chi^2(13, N = 2070) = 198.4$, $p < .001$), and overall found a large effect, Tjur $R^2 = .417$. After controlling for tests, Asian students had nearly four times the odds (Exp(b) = 3.98, *SE* = .50, Wald = 2.778, 95% CI [1.50, 10,52]) than students who were White, male, paid full price lunch, and were native English speakers of being identified for gifted services. Female students had a .40 lower odds (Exp(b) = .40, *SE* = .25, Wald = -3.567, 95%CI[.24, .66]) than students who were White, male, paid full price lunch, and were native English speakers of being identified for gifted services. Students who are Black/African American, Hispanic/Latinx, on free/reduced lunch, English learners, as well as scores on the CogAT Non-Verbal and ITBS Math showed no differences in the probability of being identified for gifted services. Scores on the TTCT Form A or B, NNAT-2, CogAT Quantitative, CogAT Verbal, and ITBS Reading showed roughly a 50% probability of being identified for gifted services. Again, after adding the TTCT Form A or B, the model marginally improved ($\Delta R^2 = .022$). See Figure 1.28 to Figure 1.34 for the probabilistic relationship by each test.

In Block 4, continuous variables were centered (i.e., all tests) and interaction terms with demographic variables and the TTCT were added to the model. After controlling for the effects of the TTCT Form A or B, NNAT-2, CogAT Quantitative, CogAT Verbal, and ITBS Reading, students who were female were found to contribute to the model ($\chi^2(19)$, $N = 2070$) = 316.1, $p <$ .001), and overall found a large effect, Tjur $R^2 = .420$. Similar to the previous model, the model indicated females had .39 lower odds (Exp(b) = .39, $SE = 0.26$, Wald = -3.524, 95%CI [.19, .40]) than students who were White, male, paid full price lunch, and were native English speakers; this was a conditional effect after each test was mean-centered. Students who are Black/African American, Asian, Hispanic/Latinx, on free/reduced lunch, English learners, as well as scores on the CogAT Non-Verbal and ITBS Math showed no differences in the probability of being identified for gifted services. There were no interaction effects found. See Table 1.17 for more details.

1.3.6 Additional Analysis: Multilevel Generalized Linear Models

To determine if schools vary with regard to who is identified for gifted services, multilevel generalized models were used as an extension to the generalized linear models. Prior to adding demographic variables and controlling for tests at Level 1, the proportion of variance explained by schools was 11% (ICC = .11) for students referred and was 7% (ICC = .07) for students who were universally screened.

1.3.6.1 Referrals

For students who were referred, results showed that, on average, students had a 6.5% probability of being identified for gifted services (Exp(b) = .07, $SE = .23$, Wald = -11.611, probability = .07 / [1 + .07]), with a R^2 = .72 marginal effect across schools (*n* = 31 schools). The variability in intercepts from school to school indicate little variance from each other, specifically schools vary .09 standard deviations. After controlling for the effects of each test, the fixed effects indicated, on average, Asian students who are referred had four times the odds (Exp(b) = 4.23, *SE* = .43, Wald = 3.374) of being identified for gifted services within elementary and middle schools. Analogous to the generalized linear regression above for students who were referred, students who are female, Black/African American, Asian, Hispanic/Latinx, on free/reduced lunch, and English learners showed no differences in the probability of being identified for gifted services. See Table 1.18 for more details.

1.3.6.2 Universal Screening

For students who were universally screened in first grade, results showed that, on average, students had a 1% probability of being identified for gifted services (Exp(b) = .01, *SE* = .38, Wald = -13.135, probability = .01/ [1 + .01]), with a R^2 = .72 marginal effect across schools (*n* = 23 schools). The variance in intercepts from school to school indicate schools vary .26 standard deviations. After controlling for the effects of each test, the fixed effects indicated, on average, Asian students who were universally screened had four times the odds (Exp(b) = 4.42, *SE* = .51, Wald = 2.914) of being identified for gifted services within elementary and middle schools. Additionally, females had .36 lower odds (Exp(b) = .39, *SE* = .26, Wald = -3.632) of being identified than students who were White, male, paid full price lunch, and were native English speakers to be identified for gifted services. Similarly found in the generalized linear regression for students who were referred, students who are Black/African American, Hispanic/Latinx, on free/reduced lunch, and English learners showed no differences in the probability of being identified for gifted services. See Table 1.19 for more details.

1.4 Discussion

Within this study, the relationship among student demographics, scores on the TTCT Form A or B, cognitive ability tests (i.e., CogAT7 Non-Verbal, Quantitative, Verbal; NNAT-2), achievement tests (ITBS Reading and Math), and the probability of being identified as gifted were analyzed for students who were referred and students who were universally screened within a midsize diverse school district in the state of Texas. The findings could generalize to other diverse school districts in midsized urban areas that receive funding for Title I status, as well as districts who use similar tests in their identification system.

Initial descriptive statistics for the total sample found the TTCT weakly correlated with most of the cognitive ability and achievement tests, besides the CogAT7 Quantitative and the ITBS Reading. Although there were differences in correlation coefficients among students who were referred or universally screened (see Figure 1.13 and Figure 1.21), cognitive ability and achievement tests were overall weakly related to the TTCT (or not at all). It is also important to note that the TTCT had the weakest relationship with gifted identification for both students who were universally screened ($r(8) = .09$, $p < .001$) and referred ($r(7) = .14$, $p < .001$), compared to academic achievement and cognitive ability test scores. Unsurprisingly, measures of cognitive ability had higher relationships with being identified for gifted services. This coincides with the value placed on measures of intelligence over creativity in schools (Kaufman & Plucker, 2011), regardless of the emphasis of the interplay between the two (i.e., threshold theory; Guilford, 1967; Jauk et al., 2013; Shi et al., 2017). Further, what is known about threshold theory is that the relationship with creativity tends to diminish with higher cognitive ability. Thus, the results are also likely expected based on threshold theory in that students are universally screened with a cognitive ability measure and required to meet a high threshold (e.g., 94th-95th percentile), then score lower on a measure of creativity (i.e., TTCT) in the formal evaluation process. An alternative explanation could be due to state definitions not always emphasizing creativity (Rinn et al., 2020), thus the emphasis on cognitive ability and academic achievement are likely influenced by the ramifications of interpretations of the Texas State Plan for the Education of Gifted Students that give unclear direction on how creativity should be served in gifted programs (Texas Education Agency, 2019).

Within the initial simple regressions, cognitive ability tests and achievement tests explained marginally more variance than scores on the TTCT Form A or B for all demographic subgroups. More specifically, there were significant differences found between racial/ethnic categories on the TTCT Form A or B, CogAT7 (Non-Verbal, Quantitative, Verbal), ITBS (Reading and Math), and NNAT-2, and there were significant differences for sex, English learners, and students on free/reduced lunch when assessed individually. However, when added to a multiple regression with student demographics, cognitive ability tests, and achievement tests to the TTCT (i.e., TTCT Form A, TTCT Form B, and TTCT Form A or B), differences diminished for students who were referred, but were still apparent in students who were universally screened in first grade. For students universally screened, there were discrepancies found by TTCT Form B.

Upon evaluation of the representation indices, there were areas of strength for underrepresented groups and opportunities for improvement. Additionally, hypothetical combination rules favored either an OR rule or an AVERAGE rule, compared to an AND rule. For Question 3 and four, after controlling for tests, the generalized linear model found student subgroup differences in identification per referrals and students universally screened. The hierarchical generalized model found demographic subgroup differences for students who are female, Asian, on free/reduced lunch, or designated as an English learner with differences regarding whether students were referred or universally screened. Similarly, in the multilevel generalized linear regressions, there were significant differences found for students who are female or Asian in identification for gifted services across schools.

1.4.1 QuantCrit Lens

Before a further discussion of my results, it is vital to practice reflexivity (Gillborn et al., 2017) and reflect on my own positionality. I am a doctoral candidate at the University of North Texas and former advanced placement educator. I am also a White female, first-generation college student, who, now in adulthood, is married and resides in an upper-middle class neighborhood. Through my schooling as well as experiences as an advanced placement educator and educational researcher, I have become deeply invested in developing creativity, identifying students for appropriate services, and helping provide learning opportunities for marginalized students to achieve at advanced levels. I recognize the advantages that I have as a White woman and I can never truly know firsthand the experience of a person of color or multilingual individual (or the intersectionalities that exist within diverse experiences), however I aim to use my privilege to support equity and help schools mitigate barriers to access to gifted programs. Although race/ethnicity was only a portion of this study (i.e., sex, English learner status, free/reduced lunch were also analyzed), a critical lens was used in reflection of the findings. For the current study, five general principles for quantitative critical analysis guided my reflection: (1) acknowledgement of the centrality of racism in society, (2) recognizing data as not neutral and can preserve White interests and beliefs, (3) interrogate the use of categorical variables, (4) data can have numerous conflicting interpretations, and (5) overall, to support goals of social justice (Gillborn et al., 2017). The goals of social justice in education include creating a society in which “the distribution of resources is equitable and ecologically sustainable, all members are physically and psychologically safe and secure, recognized, and treated with respect” (Bell, 2016, p. 3).

1.4.2 Demographic Subgroup Differences per Test

Within the series of regressions, there were notable demographic differences in comparison to scores on each of the tests used at Rosemary School District. The data showed significant differences based on categories of race/ethnicity, sex, English learner status, and free/reduced lunch status across the TTCT Form A or B, CogAT7 (Non-Verbal, Verbal, Quantitative), ITBS (Reading and Math), and NNAT-2.

1.4.2.1 Race/Ethnicity

For students referred, the TTCT displayed significant differences for Hispanic/Latinx students with a decrease in scores, and Black/African American, Asian, Native American, and Other students showed no differences in performance on the TTCT. This differed from students universally screened in first grade in that Black/African American students scored higher on the TTCT, whereas Hispanic/Latinx and Asian students scored lower than students who are White (White was the reference group). For students who were referred, Black/African American, Native American, Asian, and Other students showed no significant differences in scores on the TTCT. For students universally screened, only Native American and students categorized as Other exhibited no differences; there were not many Native American or students categorized as Other universally screened, thus impacting their representation.

For students referred and universally screened, group mean differences varied for students classified as Black/African American, Hispanic/Latinx, Native American, Asian, and students categorized as Other across cognitive ability (CogAT7 Non-Verbal, Quantitative, Verbal; NNAT-2) and achievement tests (ITBS Reading and Math) compared to White students. Comparatively, the CogAT7 (Verbal, Non-Verbal, Quantitative) and ITBS (Reading and Match) had a marginally higher effect than the TTCT in both students who were universally screened in first grade and those referred in first through eighth grade (see Table 1.4 and Table 1.8 for more detailed results on each test). Analogous to group differences found in excellence gaps research, achievement test scores showed significant differences for students classified as Black/African American, Hispanic/Latinx, Native American, and students categorized as Other (Plucker & Peters, 2016; Rambo-Hernandez et al., 2019) for both students who were referred and universally screened; however they were not likely the sole contributing factor to underrepresentation of racially diverse groups at Rosemary School District (de Brey et al., 2019; Worrell & Erwin, 2012), as that was not supported by the generalized linear regressions that controlled for tests with this sample.

Some caution should be exercised in the interpretation of dummy-coded race/ethnicity variables, as the combined category of Other is difficult to interpret and needs to be disaggregated to fully interpret. Further, the comparison within each multiple regression to White students would change with a different reference group and could be further explored. Likewise, the significant group differences found can have conflicting interpretations based on if an author holds the belief that all standardized tests (e.g., cognitive ability, academic achievement) have historically shown racial test bias (Ford et al., 2020; Gould, 1996), or the opposite, that the author takes the stance that group differences on standardized tests are not indicative of test bias (Peters, 2021), but present differential prediction and violate other types of validity (Messick, 1989; Warne et al., 2013). There are aspects to both perspectives that should be recognized.

Racism is centralized within American society and anti-racist practices should be implemented especially within gifted programs. Mun et al. (2020) discusses how in order to disrupt institutional racism, leaders need to remove barriers that "perpetuate deficit thinking about CLED students, their families, and the communities in which they live" (p. 134; also see Ford, 2014). Although the measurement of intelligence has been viewed, by some, as a significant human achievement, it is historically fraught with controversy (Neisser et al., 1996; Nisbett et al., 2012). Borland (2003) argues that measurements used in the identification process, particularly standardized tests, could be viewed as a means of surveillance that reinforces power over students and teachers (e.g., hierarchical observation; Foucault, 1995) and normalizes judgments (p. 109) to separate students based on ability found on standardized tests (e.g., IQ). As such, there are many that view standardized testing, inclusive of intelligence tests, as culturally and racially biased (Baldwin, 2002; Ford & Whiting, 2007; Mansfield, 2015) and call for their removal from gifted identification. For example, Ford et al. (2020) said identification systems are:

> ...grounded in the belief that gifted is synonymous with intelligence and achievement, and that both can be measured validly and reliably with tests and checklists, regardless of culture, exposure, opportunity, language proficiency, and income. (p. 30)

There are scholars who agree with the notion that there has been an overreliance on cognitive ability and academic achievement tests, and that more non-traditional tests should be used in the identification process, particularly for underrepresented student groups (Naglieri & Ford, 2003, 2005; Mun et al., 2016). As Ford et al. (2020) has expressed, there are various possible reasons for differences in scores that prevent a child from being identified for gifted services. It is imperative to note that unequal outcomes expressed in a single test score could have multiple intersectional reasons for differences (e.g., unfair penalization due to race, English learner status, gender, socioeconomic status, motivation, or specific environmental reasons;

Popsham, 2012) and/or be more indicative of larger societal inequality beyond the test itself (Peters, 2021).

The results of this study show that measures of cognitive ability and academic achievement do explain marginally more of an effect in demographic subgroup differences than a creativity test (i.e., TTCT) in the current sample, similar to how cognitive ability and academic achievement contribute more to the probability of being identified for gifted services. Similar demographic score differences found within this study have been historically found with cognitive ability tests across multiple fields of study (Flynn, 2009) and, almost indistinguishably, also across measures of academic achievement (e.g., achievement and excellence gaps; Plucker & Peters, 2016; Worrell & Dixson, 2018).

To take another perspective, the results from this current study show indication of differential prediction (i.e., predictive bias), but do not provide a definitive indication of test bias (Berry, 2015). This study is not a signal to vilify all standardized tests as racially biased (Worrell & Dixson, 2020). All standardized measures that were used within this study went through substantial standardization procedures at an item-level (and by form) to guard against measurement bias (i.e., test bias; American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). Mean differences were found on each test, but the potential causes of group-specific differences from each regression equation could be attributable to some other variable not accounted for. At least for the TTCT, when more demographic variables were added to a multiple regression, many of the significant race/ethnicity differences dissipated (see below Additional Analysis of TTCT by Form), providing a conflicting interpretation. Also, the scores obtained at Rosemary School District are not reflective of national demographics; thus, score differences found within this study are more likely an indication of local sampling variability.

That said, this study analyzed aspects of criterion validity (i.e., concurrent validity with other identification measures, differential predictive validity of gifted identification), and not the specific construct validity inclusive of the specific internal differences of the measurements themselves (i.e., measurement invariance, item-response theory, differential item functioning). Rather, what may be more concerning is consequential validity (Messick, 1989; Popsham, 1997; Warne et al., 2013). Consequential validity "requires evaluation of the intended or unintended social consequences of test interpretation and use" (Messick, 1989, p. 84) and is concerned with whether the ends justify the means of using a specific test, despite unintended social side effects. For example, racial/ethnic and/or sex differences in scores could be an unintended side effect of using a creativity test and could possibly create an adverse impact if used in identification for gifted services without careful consideration of what is contributing to the differences in performance (e.g., construct validity, criterion-contamination). Additionally, it is unknown how a child's observed score obtained in the identification process compares to a students' score without measurement error (i.e., true score in Classical Test Theory; Lord, 1980). Peters (2021) argues that "the cause of the true score difference is not relevant to the question of whether or not an assessment is biased. A test with reliable scores that yield valid inferences will result in unequal outcomes in the context of an unequal society" (p. 2). That is not to say that the observed group differences should go without scrutiny for prospective bias (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). In the context of gifted identification, if scores obtained on tests used in identification are severely disparate and contribute to unequal outcomes, these differences should encourage concern and urge administrators to re-evaluate existing systems.

Again, it is important to acknowledge that institutional racism does exist and permeates throughout many economic, political, and educational structures (amongst others). Racial inequality seeps through many interconnected components of our educational systems; however, more information is needed to determine the types of bias potentially exhibited and should be explored in future studies using appropriate statistical analyses. For Rosemary School District, there were racial differences found on each test in the process, but this is only one school district with specific demographics and intricacies of the surrounding community that need to be locally explored. Beyond Asian students scoring higher on specific tests, the racial/ethnic differences for underrepresented demographic groups found on the tests did not translate to the probability of identification for gifted services; though, consistent with the extant literature, Asian students were more likely to be identified. More information in relation to race/ethnicity is discussed below in the section on the generalized linear regressions.

1.4.2.2 Sex

For both referrals and students who were universally screened, females scored higher than males on the TTCT-Figural, CogAT Verbal, and the ITBS Reading. For students who were referred, females scored significantly lower on the CogAT Quantitative and ITBS Math, but no differences were found on the CogAT Non-Verbal. Additionally, for students who were universally screened in first grade, there was no significant difference in scores between males and females on the NNAT-2, CogAT Quantitative, or the ITBS Math. The universal screening findings for females indicate early quantitative cognitive ability or academic achievement in mathematics are indistinguishable to males in this sample. Be that as it may, for students referred, females scored lower than males on measures of quantitative cognitive ability and academic achievement in mathematics for students referred across grade-levels. Thus, this could be an indication of female internalization of stereotypical conceptions that ‘males are better at math’ and influencing their math self-concept within elementary school and impacting their performance in later grades (Cvencek et al., 2011). Females could also be more likely to be referred because of early talent in reading or writing. For students referred and universally screened, females scored higher than males on measures of verbal cognitive ability and academic achievement in reading. Similar score difference have been found in prior studies on sex differences in verbal cognitive ability (Wai et al., 2010) and academic achievement in reading and writing (Pargulski & Reynolds, 2016; Reilly et al., 2019).

These findings support Kim (2017) and Bart et al. (2015) who found females scored significantly higher than males on all subscales of the TTCT-Figural. The notable differences found on the TTCT-Figural could be attributable to females performing better on drawing tasks that require sustained attention (Stewart, 2007), or females developing fine motor skills more rapidly than males (Kokštejn et al., 2017). Historically, one of the influences for the early iterations of the TTCT-Figural was a projective drawing test (using stimulus drawings in activity two) designed to capture expressions of masculinity-femininity (i.e., sex role stereotypes; Franck & Rosen, 1949). Franck and Rosen (1949) also found females to score higher on the test, were more likely to draw open figures, and were more likely to show more internal elaboration with their sample. All of these are part of scoring higher on the TTCT-Figural today. Although females outperform males within the current study, their performance on the TTCT-Figural did not contribute to the probability of being identified for gifted services, as females were less likely to be identified for gifted services overall.

1.4.2.3 English Learner Status

English learners significantly scored lower on the TTCT than native English speakers regardless of if they were universally screened in first grade or referred at any grade. For students who were referred, English learners scored lower on the CogAT Quantitative, CogAT Verbal, and the ITBS Reading, but there were no significant differences for students who took the CogAT Non-Verbal or the ITBS Math. There were some notable differences among students who were universally screened; English learners scored lower on the NNAT-2, CogAT Non-Verbal, and the ITBS Math.

The findings for the non-verbal cognitive measures resemble similar findings by Peters and Engerrand (2016) who found English learners scored lower on the NNAT-2 and the CogAT Nonverbal-Form 7. Similarly, Carman et al. (2018, 2020) found English learners scored lower on the CogAT7 Non-Verbal, as well as the NNAT-2. Thus, it is not surprising to find English learners to obtain lower scores on a non-verbal creativity measure as well. The lower scores obtained on the TTCT-Figural could be an indication of predictive bias (or measurement bias) for English learners, as the specific subscales of the TTCT-Figural require students to create abstract titles (i.e., Abstractness of Titles subscale), which might be difficult for a child with limited proficiency in English.

Mun et al. (2016) posit inconsistent identification practices implemented for English learners could contribute to underrepresentation. For instance, differences for English learners could be due to English learners not having a translated version of the test, nor the instructions given in their native language. Rosemary School District's gifted identification and program documentation says students are provided tests in languages they understand; however, test administration methods for the TTCT are not clear. Even if the TTCT was administered in English, the district does provide an additional point to students who are English learners if they were close to the designated cut-score to buffer point differences that would prevent students from reaching the 125 threshold for any battery on the CogAT7, ITBS (Reading or Math), or TTCT-Figural Form A or B. Mun et al. (2016), Naglieri and Ford (2003), and VanTassel-Baska (2008), amongst others, have advocated for the use of non-verbal cognitive ability tests, and creativity tests to help identify more English learners, however there were significant differences found on the TTCT-Figural Form A or B and the NNAT-2 with this sample. This could be an indication that English learners may need to be re-assessed with a translated version of the TTCT-Figural, as well as administration of the test in their native language. Further, the mere usage of non-verbal tests has not been shown to increase proportional representation (Carman & Taylor, 2010); as such, similar districts should be wary of differences in scores and not assume using a non-verbal test will solve issues of inequity within gifted programs.

1.4.2.4 Free/Reduced Lunch

For students who were referred, the students who were on free/reduced lunch status scored lower on the TTCT-Figural than students not on free/reduced lunch. Conversely, there were no significant differences in scores on the TTCT-Figural or the ITBS Math found for students who were universally screened. Comparatively, the CogAT7 (Verbal, Non-Verbal, Quantitative) and ITBS (Reading and Math) also showed lower scores for students on free/reduced lunch who were referred. For students universally screened, students had a much lower NNAT-2 score compared to scores on the CogAT7 (Verbal, Non-Verbal, Quantitative), but higher scores on the ITBS Reading.

The findings for the TTCT-Figural were similar to Hermon et al. (2018) in that students on free/reduced lunch scored lower than students not on free/reduced lunch, as well as confirmed by an older study by Bashaw and White (1979) that found kindergarteners who were economically disadvantaged to have lower scores on the TTCT-Figural. Likewise, similar score differences for students on free/reduced lunch have been found on the CogAT7 (Verbal, Non-Verbal, Quantitative) by Peters and Engerrand (2016) and Carman et al. (2018, 2020), as well as for the NNAT-2 (Carman & Taylor, 2010). Carman and Taylor found students who were from high socioeconomic backgrounds had higher scores than students from low socioeconomic backgrounds. Similar findings have been found on other cognitive ability tests and measures of academic achievement (e.g., Measures of Academic Progress [MAP], Northwest Evaluation Association, 2011; Plucker & Peters, 2016) for students who are living in poverty (Olszewski-Kubilius & Corwith, 2018). The disparities found are likely consequences of students from more advantageous backgrounds being provided opportunities to learn at younger ages (e.g., greater access to reading books and educational computer games, more time spent with parents inside/outside the home, Bassok et al., 2016; access to center-based preschools; Riccardi et al., 2020). Thus, differential economic, social, and cultural capital impacts performance on all tests used in the gifted identification process, and, notably, these differences can impact the likelihood of being identified (see Grissom et al., 2016 for more details). The same could be said for the disparities on the TTCT-Figural. Although Torrance (1971) propagated the TTCT as helpful for students from disadvantaged backgrounds, there were similar discrepancies for students on free/reduced lunch across all tests that should be further examined.

1.4.2.5 Additional Analysis of TTCT by Form

Although there were significant differences found when each test was assessed separately, when added to a single model there were disparities between student demographics, cognitive ability, academic achievement, and scores on the TTCT-Figural Form A, Form B, and Form A or B. Data can have conflicting interpretations when considering influences of other variables, especially across different paths to the formal evaluation for gifted services (e.g., referrals, universal screening). For both students referred and universally screened, a series of hierarchical regressions examined the relationship among student demographics, cognitive ability tests, and achievement tests to Form A, Form B, and the combination of Form A or B (see Table 1.12 and Table 1.13).

For students who were referred, there were similar differences found across both forms, as well as considering the top scores from either form. Analogous to the simple regressions, females scored higher than males across all forms of the TTCT. In the second block of each regression, not only did females score higher than males, but there was a positive relationship with the ITBS Math. Upon further examination of the type of questions offered on the ITBS Mathematics portion, there are items that ask test takers to interpret charts or tables, and "students' response options are pictorial, numbers, or words" (Dunbar et al., 2015, p. 29). The TTCT-Figural requires students to draw their responses, so although the correlations found were relatively small, the relationship could be due to an overlap in pictorial versus figural responses. Alternatively, it could be due to the subscale of fluency. This could be due to a relationship on the ITBS Math where each pictorial item you get correctly relates to an item that would be counted as a fluent idea on the TTCT Figural.

For students universally screened, there were several notable inconsistencies found among Form A and B. For TTCT Form A, females and students on free/reduced lunch tended to score higher, whereas English learners scored lower. When all the tests were added in the second regression, students who are Asian or English learners, as well as scores on the CogAT Verbal and Quantitative batteries, had a negative relationship with the TTCT Form A. Females, students on free/reduced lunch, and scores on the NNAT had a positive relationship with the TTCT Form A. This was the complete opposite from the results found with TTCT Form B. Students who are

Hispanic/Latinx, female, or on free/reduced lunch scored lower, whereas students categorized as Other or English learners scored higher on the TTCT Form B. After adding the NNAT-2, CogAT7 (Verbal, Quantitative, Non-Verbal), and ITBS (Reading and Math) to the model, there were further discrepancies. Students who are Black/African American, on free/reduced lunch, and English learners scored higher, whereas Hispanic/Latinx and females scored lower on the TTCT Form B. The relationships changed again for the combined Form A or B. For more details, see Table 1.13.

The findings from the students universally screened in first grade showed Black/African American students to score higher on the TTCT-Figural compared to the reference group (i.e., White students) and support similar findings from a review by Torrance (1971). This specific result regarding students who are Black/African American, on free/reduced lunch, and English learners scoring higher than other students supports the notion that the TTCT-Figural could help increase representation (Henshon & Grantham, 2021). This partially supports the overt recommendation offered by Kim (2011), Luria et al. (2016), and Kaufman (2010) regarding using creativity tests to identify more racially diverse groups, specifically the TTCT-Figural (Kim, 2011). Nevertheless, there are mixed findings, and potential consequences of usage for other student groups (e.g., Hispanic/Latinx).

Something vital to note is that there is no information regarding race/ethnicity or socioeconomic status provided in the TTCT-Figural technical manual (Torrance, 2017) and neither the Torrance Center at the University of Georgia or Scholastic Testing Services collect information on race/ethnicity or socioeconomic status in their standardization samples (S. Sumners, Personal Communication, July 7, 2020). Thus, this study could not compare results to a standardized sample based on a range of demographic information. Specific to this study, the mixed mean differences on the TTCT-Figural Form A or B found at Rosemary School District suggest that closer attention should be placed on racial/ethnic differences when re-norming each form of the test. For example, the universally screened students who are Black/African American had an increase in scores on the TTCT Form A or B, however students who were Hispanic/Latinx had a decrease in scores (also the differences for Hispanic/Latinx students were seen in the referred students). Findings seen in this study suggest racial differences should also be further studied regarding the age and/or grade for test administration when using the TTCT as well as intersectionalities that exist amongst students in the process.

All that said, the discrepancies seen for universal screening students should be further explored for Form B, especially for students who are universally screened in earlier grades (K-2nd). Although there has been evidence of equivalent forms of the TTCT-Figural Form A and B (Ball & Torrance, 1980; Clapham, 1998), there may be additional inconsistency that should be addressed by looking at differential predictions by different forms of the TTCT-Figural with younger children. Alternatively, these inconsistencies could be in part due to the large amount of missing data that was imputed with predictive mean matching. More replication is needed with a larger sample of younger children (K-1st grade) who took both Forms A and B for further determination.

1.4.3 Demographic Proportionality

Within the 2018-2019 school year, there was an increase from the 2017-2018 school year (OCR, 2018) in terms of proportionality of students identified for gifted services who are Native American (RI = 1.50) and Native Hawaiian/Pacific Islander (RI = 1.00) in Rosemary School District. In line with scholarship on representation indices, students who are Asian and White continue to be well-represented (Yoon & Gentry, 2009). However, students who are

Black/African American, Hispanic/Latinx, of two or more races, on free/reduced lunch, and English learners remained disproportionate within Rosemary School District's gifted program. Although not proportional, Rosemary School District had higher representation of Black/African American students (RI = .60) than comparative to a school with similar demographics (e.g., Cypress School District had .36 in the 2016-2017 school year; Mun et al., 2021). Something to note is that these estimates for representation are based on the total population reported to the Texas Education Association (TEA) for the 2018-2019 school year and are inclusive of the entire district, not just the elementary and middle schools (see Table 1.14).

Peters et al. (2019a) found students designated with Limited English Proficiency (LEP) to be identified 41% less frequently across the state of Texas. Comparatively, Rosemary School District had higher representation of students who are Black/African American, Native American, and English learners compared to numbers across the state of Texas in 2015-2016. This increased representation in English learners identified at Rosemary School District could be attributable to the majority of their elementary campuses being designated as Title I, with more focused attention on students from economically disadvantaged backgrounds and/or underrepresented groups (Gentry et al., 2021). It is important to note that we live in an unequal society and proportionality across all demographics is not a viable goal (Peters, 2021), however schools can still mitigate disproportionate representation in gifted programs.

1.4.4 Combination Rules

In applying a hypothetical AVERAGE, AND, and OR rule, it was found that the OR rule and the AVERAGE rule were the least restrictive in application. Similar to McBee et al. (2014) and Lakin (2018), the AND rule was the most restrictive rule and only produced one student that met the hypothetical criteria (i.e., score at the 95th percentile on all measures). Conversely, the

OR rule was more relaxed and allowed more students to be identified (n = 1175). Although the OR rule broadened diversity across all demographics (with a larger group of students who would be identified), the heterogeneity of the group could pose challenges in providing services for the sheer number of students (e.g., teachers, gifted specialists, curriculum materials, funding) and not having the adequate resources to serve all 1175 students. In an ideal educational setting, schools would be able to serve all of these students if strength-based, domain-specificity was supported in gifted programs. However, this is not the reality in many schools. Alternatively, the AVERAGE rule identified 119 students and provided improved access for Black/African American students (12%) but increased the number of White and Asian students as well.

The findings here echo similar work from Lakin (2018), McBee et al. (2014), and Lohman (2008) in that there are advantages to implementing an AVERAGE or an OR rule. The current study did not hold program size constant, though, so little comparison can be made to Lakin's (2018) results. However, the inclusion of creativity tests, measures of cognitive ability, and achievement tests likely limited the pool of students who would be identified with an AVERAGE, and especially an AND, rule. Like findings from McBee et al. (2014), uncorrelated tests used in the identification process likely contributed to less students being identified with the AVERAGE and AND rules (see Table 1.15). Further, considerations of the impact of two-phase identification systems with uncorrelated tests should be further explored in gifted identification research with actual district data. Future research should aim to replicate Lakin's (2018) study, inclusive of holding program size constant, and extend with additional counterfactual analyses of different test combination pathways, as well as explore test combination pathways through the lens of sensitivity (McBee et al., 2014).

1.4.5 Hierarchical Generalized Linear Regressions

To examine how student demographics, measures of creativity, cognitive ability tests, and academic achievement tests relate to the probability of being identified for gifted services, generalized linear regressions were used to assess both students who were referred and students who were universally screened at Rosemary School District (see Table 1.16 and Table 1.17 and Figure 1.22 to Figure 1.34).

For students referred to formal evaluation, prior to controlling for tests used in Rosemary School District's identification system, students who are Asian were 52% more likely to be identified and students who are English learners were 15% less likely to be identified for gifted services. After controlling for cognitive ability and academic achievement tests used in formal evaluation (Phase 2), English learners no longer showed a difference in the probability of being identified and showed improved overall model fit (Tjur R^2 = .457). This is likely partially attributable to the additional points English learners earn if they are close to the cut-score on any of the tests. Further, Asian students remained more likely to be identified compared to students who are White, male, paid full price for lunch, and were native English speakers. All remaining student demographics showed no difference in the probability of being identified. Similarly, when the TTCT was added the model, Asian students had four times the odds of being identified, but the addition of the TTCT only explained roughly 2% more variance in being identified (ΔR^2 = .019) compared to the model with only cognitive ability and achievement tests. Consistent with extant gifted education literature, Asian students frequently outperform other students in measures of cognitive ability and academic achievement, in turn, becoming well-represented in gifted programs across the nation (Peters et al., 2019; Plucker & Peters, 2016). Future research should disaggregate the students categorized as Asian to better understand the differences within this specific racial/ethnic category and better express the heterogeneity that exists, as well as include interaction terms specific to demographics (e.g., race/ethnicity interacting with socioeconomic status or English learner status) (Peters, 2021). More interestingly, when interactions were considered, there were no demographic differences found, nor did the conditional effect of the TTCT Form A or B show a significant difference in the probability of being identified for gifted services. There were also no interactions among student demographics and the TTCT found. However, student scores on the CogAT7 (Non-Verbal, Quantitative, Verbal) and ITBS (Reading and Math) still contributed to the probability of being identified.

There were differences for students universally screened in first grade. When only considering student demographics, students on free/reduced lunch were 3% less likely to be identified for gifted services. This is likely reflective of variability in early interventions provided to students who are on free/reduced lunch (Grissom et al., 2016) and, moreover, systemic inequality that impacts students prior to even entering elementary school (Peters, 2021). After controlling for cognitive ability and achievement tests (which also contributed to the model), the model showed a large effect (Tjur R^2 = .395) of female students had .43 lower odds of being identified for gifted services and students who are on free/reduced lunch status no longer show a difference in how likely they are to be identified compared to students who were White, male, paid full price lunch, and a native English speaker. When adding the TTCT to the model, Asian students had nearly four times the odds in being identified for gifted services and female students had were .40 lower the odds and slightly less likely to be identified (see Table 1.17). Further study is needed to understand the contradictory relationship with females who are universally screened to be less likely to be identified despite having higher scores on the TTCT-Figural, ITBS Reading, and CogAT Verbal. As seen with the correlations between the NNAT-2, TTCT-Figural, ITBS Reading, and CogAT Verbal, this might be simply an indication that the

NNAT-2 is not related to performance on the TTCT-Figural, nor to the other tests in formal evaluation (McBee et al., 2014). Almost identical to the prior models after testing for interactions, the conditional effect for females remained slightly less likely to be identified for gifted services, but Asian students showed no differences in being identified and there were no interactions found. Also, students who are Black/African American, Hispanic/Latinx, on free/reduced lunch, English learners, as well as mean-centered scores on the CogAT Non-Verbal and ITBS Math showed no differences in the probability of being identified for gifted services.

Creativity has been integrated within many theories of giftedness for decades (DMGT, Gagné, 2017; Renzulli's Three-Ring Conception, Renzulli, 1978; Starfish Model, Tannenbaum, 2003), however little research has reported the relationship regarding how creativity tests compare to measures of cognitive ability and achievement tests in the probability of being identified. The findings from this study offer a few perplexing interpretations for including the TTCT-Figural. First, the TTCT-Figural, cognitive ability tests, and academic achievements all contribute to the probability of being identified, thus the inclusion of the TTCT-Figural does show some promise.

For both students referred or universally screened, generalized linear models that included cognitive ability and academic achievement were found to have a slightly larger effect on decisions to identify than when the TTCT Form A or B are included (see Table 1.16 and Table 1.17). However, the models that only included cognitive ability and academic achievement had a greater overall effect. Glancing upon Figure 1.22 to Figure 1.27,, the plots show how a high creativity score (i.e., TTCT-Figural) does not equate to an increased prediction of identification at Rosemary School District, whereas high scores of cognitive ability and academic achievement increase the probability of being identified for students who are referred.

For students universally screened in first grade, more emphasis was on cognitive ability measures (see Figure 1.28 to Figure 1.34). These results coincide with the emphasis on cognitive ability and academic achievement in gifted identification practices, as well as the continued focus on differential predictions of student demographics with their usage (e.g., Carman et al., 2018, 2020; Lohman et al., 2008; McBee et al., 2014, 2016 Olszewski-Kubilius & Corwith, 2018; Peters & Engerrand, 2016).

Secondly, it is still unknown if the inclusion of the TTCT-Figural is what is influencing no differences being found for underrepresented groups, as the inclusion of only cognitive ability and academic achievement displays similar results across all demographics. Although creativity was part of their scope and sequence, creativity as measured by the TTCT-Figural is not helping to identify more students overall at Rosemary School District. Creativity scholars have recommended creativity tests as a “equalizer” and that they could influence accuracy and program diversity (Luria et al., 2016), and Kim (2011, 2017) even suggests that the TTCT should be used in the identification process to identify more underrepresented groups of students for gifted programs. Kim (2006) stated that the “TTCT-Figural can be fair in terms of gender, race, and community status, as well as for persons with a different language background, socioeconomic status, and culture” (p. 8). At Rosemary School District, the inclusion of the TTCT-Figural did not further equalize the likelihood of being identified compared to cognitive ability and academic achievement tests (see Figure 1.22 and Figure 1.28). This could be showing how creativity, as measured by the TTCT, is only one of the three criteria that students can score highly on, whereas students could score highly on any of the different aspects of cognitive ability (CogAT7 [Non-Verbal, Quantitative, or Verbal]) or academic achievement (ITBS [Reading or Math]) to be identified. Thus, cognitive ability and academic achievement weigh more in the identification process. This could be merely indicative of threshold theory, whereas students who score higher on cognitive ability measures do not necessarily score highly on creativity measures (Guilford, 1967; Jauk et al., 2013; Shi et al., 2017). Alternatively, these results could be showing individual values at either the campus or district committees about their own conception of giftedness, whereby creativity is of secondary value compared to cognitive ability. Comparative usage of using various creativity tests (or alternative assessments) with cognitive ability and academic achievement tests in identification decisions (beyond validation of rating scales), is severely limited in the gifted identification literature and should be further explored (Lee & Rinn, 2021; Lee et al., 2021; McBee & Makel, 2019; McBee et al., 2014).

Lastly, the results could be influenced by Rosemary School District attempting to frontload specific planned experiences and train their teachers and specialists to act as talent scouts at their respective campuses. Although teacher training and specific targeted interventions were not included in this study, these could have influenced the differences in referrals and universal screening at Rosemary School District. Subsequently, other aspects of Rosemary School District's vision, mission, goals, gifted education policies, targeted inventions, and specific decisions made in the formal evaluation process at both the campus and district committees should be evaluated to determine how they relate the probability of being identified (e.g., Gubbins et al., 2021; Lee et al., 2020; Mun et al., 2020; Mun et al., 2021).

1.4.6 Additional Multilevel Analysis

To examine how student demographics, measures of creativity, cognitive ability tests, and academic achievement tests relate to the probability of being identified across schools, multilevel generalized linear regressions were used to assess both students who were referred and students who were universally screened across schools at Rosemary School District.

For students who were referred, students, on average, had a 6.5% probability of being identified for gifted services across schools. There was little variability across schools (*SD* = .09). After controlling for the effects of each test, the fixed effects indicated, on average, Asian were 22% more likely to be identified for gifted services within elementary and middle schools compared to students who are White, male, pay full priced lunch, and are native English speakers. However, students who are female, Black/African American, Asian, Hispanic/Latinx, on free/reduced lunch, and English learners indicated no differences in the probability of being identified for gifted services between schools (see Table 1.18).

For students who were universally screened in first grade, students had a 1% probability of being identified, with a little more variation across schools (*SD* = .26). Again, after controlling for the effects of each test, Asian students, on average, were 1.17% more likely to be identified for gifted services within elementary and middle schools compared to students who are White, male, pay full priced lunch, and are native English speakers. Additionally, females were .0009% less likely to be identified for gifted services compared to students who are White, male, pay full priced lunch, and are native English speakers. Also, students who are Black/African American, Hispanic/Latinx, on free/reduced lunch, and English learners showed no differences in the probability of being identified for gifted services across schools. Thus, students who are universally screened in first grade show more variability across schools compared to students who are referred. Further research should look at other school-level characteristics (e.g., Title I status) in relation to the variability in students universally screened in first grade (see Table 1.19).

Overall, scores on the TTCT-Figural, as well as cognitive ability tests, and academic achievement tests had fairly equal contributions to the probability of being identified, thus having a large marginal effect between schools (R^2 = .72). Despite only comprising 3% of the total population at Rosemary School District, Asian students were more likely to be identified across schools, no matter if they are universally screened or referred; this resembles what was found within the generalized linear regressions. Similarly, females were, on average, less likely to be identified. Although differences for Asian and females were found, there were no significant differences for students who are Black/African American, Hispanic/Latinx, on free/reduced lunch, or English learners across schools.

1.4.7 Implications for Practice

As the United States continues to rapidly become more diverse, it is vital for educational systems to grapple with the increased diversity and address how to overcome policies that are restrictive for students of color, multilingual students, and students who are considered economically disadvantaged. Overall, administrators need to support goals of social justice and consider the equitable distribution of educational resources to recognize student strengths and maximize learning for underrepresented students in gifted programs (Bell, 2016; Dixson et al., 2020). Gifted program policies at Rosemary School District indicate that supporting goals of social justice and removing barriers to access is of vital importance. The insights gained from Rosemary School District show promising implications for similar districts and for test manufacturers (i.e., Scholastic Testing Service).

First off, using the strategy of finding the right test to identify students will minimally impact rates of identification. Thus, it is more important to determine why certain tests are used and how they are being used for gifted identification. District administrators should use the most valid and reliable tests available that measure constructs that align with policy definitions, goals, and outcomes of a specified gifted program (Gubbins et al., 2020; Lee et al., 2020; Peters et al., 2020). Also, administrators need to consider how these policy definitions relate to combination decisions in a multiple criteria system (McBee et al., 2014; McBee & Makel, 2019). It was found that all the tests roughly contribute equally to the likelihood of being identified for gifted services. However, when tests were assessed individually, cognitive ability and academic achievement contributed more to the probability of being identified. Thus, it does not matter if a creativity test is used, it matters how it is used in the process to impact identification.

When considering combination rules used at Rosemary school district, the number of students identified could be indicative of the TTCT-Figural being weakly correlated or uncorrelated with the other tests in the identification process. In a two-phase system, it is of utmost importance to make sure the tests used at Phase 1 are correlated with the tests students will encounter at Phase 2 (McBee et al., 2014, 2016). Thus, it is important for districts to evaluate their identification protocols, why they are using specific tests, how related each test is within the process and unravel individual differences in relation to the probability of being identified.

Coupled with how tests are related, administrators should consider lowering the threshold (cut-score) when combining multiple criteria for identification decisions. The 95th percentile may be too restrictive of a cut-score across student demographics at the district, especially in consideration of an AVERAGE rule. Further, the flat-cut score of 125 is inconsistent across nationally normed age and grade level scores in the TTCT-Figural Norms-Technical Manual (Torrance, 2017). Rosemary School District should consider a locally determined percentile rank (i.e., local norms; Peters et al., 2019) or use a specified national percentile rank (e.g., students in the 75th percentile) based on age or grade from the technical manual.

In regard to usage of the TTCT-Figural, there are a few other aspects of the tests that should be considered. Specifically, translated versions of the TTCT are available and should be used concurrently with test instructions provided in a student's native language whenever possible. There were notable demographic differences found when using the TTCT-Figural, however this could be an indication of issues with the underlying factor structure (Acar et al., 2021; Forthmann et al., 2019). The reliance on the Creativity Index is a common practice but should cause concern because it disregards the other subscales that make-up the composite score. Torrance (2004) and Treffinger (2004) both encouraged administrators that a single indicator for creativity does not provide enough context of student strengths. Treffinger (2004) urged creativity assessments to be used diagnostically to guide instruction in the classroom, rather than "merely as criteria for selecting students" (p. 89). Prior to any decisions made for identification, there needs to be more recent research (and re-norming) regarding the subscales of the TTCT-Figural, namely the originality scores (Acar et al., 2021), with different demographic subgroups (e.g., race/ethnicity, socioeconomic status, English learners). The current study did not confirm an unfavorable overall creativity index score, but there was little known on how students scored on each subscale. Also, Scholastic Testing Service should re-evaluate norming techniques for the TTCT in regard to race/ethnicity, socioeconomic status, English learner status, and sex, and should be reported results in the technical manual, as the results from this study could not be nationally compared.

If administrators are wanting to identify specific aspects of creativity, this might be an indication for administrators to look at the subscales of the TTCT instead of solely basing their decision on a single creativity index composite score. In practice, there needs to be consideration of looking at student subscales for the TTCT to provide additional insight on individual strengths that make-up the Creativity Index score (e.g., fluency, originality, elaboration, abstractness of titles, resistance to premature closure), especially if the program assesses the individual components of the TTCT within the gifted program. Although there was indication of direct alignment for what is assessed in identification to the curriculum scope and sequence, there is still room for improvement. For example, the scope and sequence at Rosemary School District includes flexibility within skills for creativity, when that is not measured on the updated TTCT-Figural (only now on the TTCT-Verbal). This is an indication that the district needs to update their direct alignment of program goals, updated identification measures, and program outcomes (Callahan et al., 1995; Lakin, 2021; Lee et al., 2020).

1.4.8 Limitations and Future Directions

The current study has notable limitations. This study is a cross-sectional study of one school district and may not be generalizable across school districts that are not midsized urban districts. The cross-sectional nature of the study cannot imply causality of what impacts the probability of identification for gifted services, nor do the overall results justify implications of all standardized tests being biased and the cause of disproportionality. Future research should consider longitudinal designs or regression discontinuity designs to better grasp changes in policy implementation and targeted interventions of specific gifted identification practices (e.g., planned experiences; Mun et al., 2021) and the relationship to student demographics across time. Additionally, qualitative, or mixed method inquiry, informed by the results of this study, could provide needed insight to the nuances of the identification process at Rosemary School District. Future research should consider the lived experiences of students, parents, teachers, gifted specialists, and the administrators (e.g., gifted coordinators, advanced academic director, assistant superintendent) who have experience with this identification system to further understand the decision-making within the campus and district committees.

Moreover, the current study only looked at demographic and test differences to the probability of being identified; there could be other variables to be considered. For instance, special education and 504 statuses were not explored as another demographic predictor to assess identification of twice-exceptionality, nor was the intersectionality of demographic variables (e.g., students who are both Hispanic/Latinx and English learners) explored within this study. Furthermore, the combined category of Other (i.e., two or more races, Native Hawaiian/Pacific Islander) made it difficult to interpret which demographic subgroup contributed to the findings and caution should be exercised in interpretation or generalization of findings related to this variable. It is also vital to interrogate the use of the categorical variable of "Other", as the heterogeneity it may express is muddled by multiple variables and not a useful predictor.

Future studies should include a wider range of demographic subgroups, inclusive of a larger population of Native Hawaiian/Pacific Islanders and multiracial individuals and should examine the intersectionality of demographic variables to the probability of identification for gifted services. Also, another perspective would be to understand the heterogeneity within each categorical race/ethnic distinction with tests used in the identification process, particularly the TTCT (Peters, 2021). More specifically, there should be additional analyses that are inclusive of disaggregation by race/ethnicity, socioeconomic status, and geographic region (Hodges et al., 2021).

Caution should be exercised in interpreting the results for students who were universally screened as this was the entire sample of students who received the NNAT-2 whether they moved on to formal identification or not, and predictive mean matching was used to account for a large portion of missingness. For example, the results of the universal screening students showed significant demographic differences by Form A and Form B on the TTCT, but these results could merely be an error in estimation due to the limited number of students who received the TTCT Form B; this portion needs further evaluation and replication with an elementary school sample (e.g., first grade). Moreover, Forthmann et al. (2020) argues that fluency could act as a confounding influence to originality scores; future research could aim to address this relationship when moderated by form (Form A or B), specific activities within the test (Activity 1-3), or by specific student demographics (e.g., age, grade, sex, race/ethnicity, socioeconomic status, English learners), which could add in efforts to update and re-validate the TTCT.

Further, confirmatory factor analyses using item-level data need to examine the factor structure of each test and examine if tests themselves are invariant per demographic subgroup (i.e., measurement invariance). For example, there needs to be continued research on analyzing the specific subscales of the TTCT to evaluate demographic differences in performance of each subscale (i.e., fluency, originality, abstractness of title, elaboration, resistance to premature closure). As an extension of work from the 1960s and 1970s (Torrance, 1971), studies could be more inclusive of a broader range of demographics (e.g., disaggregated categories of race/ethnicity, socioeconomic status, twice exceptionality, English learners, non-binary conceptions of gender) beyond Black and White distinctions.

Although the combination rules assessed within this study partially emulated Lakin's (2018) paper, program size was not held constant, and a more direct replication and extension is still needed. For instance, the 95th percentile was chosen as the cut-off for the AVERAGE rule, however this rigidly high cut off, coupled with uncorrelated tests in the process, most likely lessened the number of students identified within each demographic subgroup for both the AND and AVERAGE rule. Future studies should apply different cut-scores (e.g., 85th or 90th percentile) and use additional counterfactual analyses to compare student demographics who would hypothetically be identified. Additionally, future research should synthesize the work of Lakin (2018) and McBee et al. (2014), and carefully consider correlation between tests used in the process, as well as analyze combination rules based on sensitivity/specificity with different test combinations to be more generalizable to actual district applications of multiple criteria.

1.4.9 Conclusion

Many school districts across the United States identify students for gifted programs (Rinn et al., 2020), and scholars continue to decry systemic problems (Gentry et al., 2020) for underrepresented students in gifted programs (Ford et al., 2020; Peters, 2021). Findings from this study indicate there are other facets of the educational system that should be considered in the probability of being identified for gifted services beyond the tests that are administered (Worrell & Dixson, 2020). There is no perfect system of identification that will ever find proportionality, but systems can be improved to provide more access equitably and efficiently. Reflections on the consequential uses of tests in the process and the differential predictions of local student demographics should be used to evaluate gifted programs. Districts should use the most valid and reliable tests available that relate across both phases of identification, as well as consider lowering arbitrarily high cut-scores across tests. The results across different tests, showed lower differences, or no differences across racial/ethnic demographics, when using the TTCT-Figural, thus showing promise in usage for gifted identification. Nevertheless, the TTCT-Figural showed minimal influence on the probability of being identified for services compared to cognitive ability and academic achievement tests at Rosemary School District. Re-consideration for which phases of identification the TTCT-Figural is used is needed, especially if a specified cognitive ability score is a threshold before administering a creativity test and if high-levels of creativity is said to be supported at the school district. As Torrance (2004) has expressed in the past, there is value in including creativity in the identification process, not as the sole criterion but should be included. Consequentially, the inclusion of a creativity test should be driven by the relation to the program (Gubbins et al., 2021; Lee et al., 2020). When considering using the TTCT-Figural, districts need to be cognizant of when it is administered, how it relates to the other measures used, and how the TTCT-Figural relates with program goals and outcomes.

1.5 References

Acar, S., Sen, S., & Cayirdag, N. (2016). Consistency of the performance and nonperformance methods in gifted identification: A multilevel meta-analytic review. *Gifted Child Quarterly, 60*(2), 81-101. https://doi.org/10.1177/0016986216634438

Acar, S., Branch, M. J., Burnett, C., & Cabra, J. F. (2021). Assessing the universality of the zero originality lists of the Torrance Tests of Creative Thinking (TTCT)-Figural: An examination with African American college students. *Gifted Child Quarterly*. https://doi.org/10.1177/00169862211012964

Amabile, T. M. (1996). *Creativity in context.* Westview Press.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Balboni, G., Naglieri, J. A., & Cubelli, R. (2010). Concurrent and predictive validity of the Raven progressive matrices and the Naglieri nonverbal ability test. *Journal of Psychoeducational Assessment, 28*(3), 222-235. https://doi.org/10.1177/0734282909343763

Baldwin, A. Y. (2002). Culturally diverse students who are gifted. *Exceptionality, 10*(2), 139–147. https://doi.org/10.1207/S15327035EX1002_7

Ball, O. E., & Torrance, E. P. (1980). Effectiveness of new materials developed for training the streamlined scoring of the TTCT, figural A and B forms. *The Journal of Creative Behavior, 14*(3), 199-203.

Barbot, B., Hass, R. W., & Reiter-Palmon, R. (2019). Creativity assessment in psychological research: (Re)setting the standards. *Psychology of Aesthetics, Creativity, and the Arts, 13*(2), 233–240. https://doi.org/10.1037/aca0000233

Bart, W. M., Hokanson, B., & Can, I. (2017). An investigation of the factor structure of the Torrance Tests of Creative Thinking. *Educational Sciences: Theory & Practice*, *17*(2), 515–528. http://dx.doi.org/10.12738/estp.2017.2.0051

Bashaw, W. L., & White W. F. (1979). *Figural creativity and convergent thinking among culturally deprived kindergarten children* [Paper presentation]. American Education Research Association, New York City, NY.

Bassok, D., Finch, J. E., Lee, R., Reardon, S. F., & Waldfogel, J. (2016). Socioeconomic gaps in early childhood experiences: 1998 to 2010. *AERA Open, 2(*3), 1-22. https://doi.org/10.1177/2332858516653924

Bell, L. A. (1997). *Teaching for diversity and social justice: A sourcebook.* Routledge.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289-300. http://www.jstor.org/stable/2346101

Berry, C. M. (2015). Differential validity and differential prediction of cognitive ability tests: Understanding test bias in the employment context. *Annual Review of Organizational Psychology and Organizational Behavior, 2*(1)*,* 435–463. https://doi.org/10.1146/annurev-orgpsych-032414-111256

Borland, J. (2003). The death of giftedness: Gifted education without gifted children. In J. H. Borland (Ed.), *Rethinking gifted education* (pp. 105-124). Teachers College Press.

Borland, J. (2004). *Issues and practices in the identification and education of gifted students from under-represented groups*. U. S. Department of Education. https://eric.ed.gov/?id=ED485164

Brown v. Board of Education, 347 U.S. 483 (1954). https://www.oyez.org/cases/1940-1955/347us483

Callahan, C. M., Tomlinson, C. A., Hunsaker, S. L., Bland, L. C., & Moon, T. (1995). *Instruments and evaluation designs used in gifted programs.* University of Connecticut, National Research Center on the Gifted and Talented.

Capps, J. (2020). Truth and the goldilocks principle. *Think, 19*(55), 65-74. https://doi.org/10.1017/S1477175620000068

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 of the United States of America, 113*(48)*,* 13678-13683. https://doi.org/10.1073/pnas.160504311

Carman, C. A., & Taylor, D. K. (2010). Socioeconomic status effects on using the Naglieri Nonverbal Ability Test (NNAT) to identify the gifted/talented. *Gifted Child Quarterly*, *54*(2), 75–84. https://doi.org/10.1177/0016986209355976

Carman, C. A., Walther, C. A. P., & Bartsch, R. A. (2018). Using the Cognitive Abilities Test (CogAT) 7 Nonverbal Battery to identify the gifted/talented: An investigation of demographic effects and norming plans. *Gifted Child Quarterly, 62*(2), 193–209. https://doi.org/10.1177/0016986217752097

Carman, C. A., Walther, C. A. P., & Bartsch, R. A. (2020). Differences in using the cognitive abilities test (CogAT) 7 nonverbal battery versus the Naglieri nonverbal ability test (NNAT) 2 to identify the gifted/talented. *Gifted Child Quarterly, 64*(3), 171–191. https://doi.org/10.1177/0016986220921164

Clapham, M. M. (1998). Structure of figural forms A and B of the Torrance Tests of Creative Thinking. *Educational and Psychological Measurement*, *58*(2), 275–283. https://doi.org/10.1177/0013164498058002010

Clapham, M. M., Cowdery, E. M., King, K. E., & Montang, M. A. (2005). Predicting work activities with divergent thinking tests: A longitudinal study. *The Journal of Creative Behavior*, *39*(3), 149-166. https://doi.org/ 10.1002/j.2162-6057.2005.tb01256.x

Cramond, B., Matthews-Morgan, J., Bandalos, D., & Zuo, L. (2005). The Torrance Tests of Creative Thinking: Alive and well in the new millennium. *Gifted Child Quarterly, 49*(4), 283-291. https://doi.org/10.1177/001698620504900402

Cramond, B., & Kim, K. H. (2008). The role of creativity tools and measures in assessing potential and growth. In J. L. VanTassel-Baska (Ed.), *Alternative assessments with gifted and talented students* (pp. 203-225). Prufrock Press.

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child Development, 82*(1), 766-779. https://doi.org/10.1111/j.1467-8624.2010.01529.x

de Brey, C., Musu, L., McFarland, J., Wilkinson-Flicker, S., Diliberti, M., Zhang, A., Branstetter, C., & Wang, X. (2019). *Status and trends in the education of racial and ethnic groups 2018* (NCES 2019-038). U.S. Department of Education, National Center for Education Statistics. https://nces.ed.gov/pubs2019/2019038.pdf

Delgado, R., & Stefancic, J. (Eds.). (2013). *Critical race theory: The cutting edge* (3rd ed.). Temple University Press.

Desmet, O. A., van Weerdenburg, M., Poelman, M., Hoogeveen, L., & Yang, Y. (2021). Validity and utility of the Test of Creative Thinking Drawing Production for Dutch adolescents. *Journal of Advanced Academics*. https://doi.org/10.1177/1932202X21990099

Dixson, D. D., Peters, S. J., Makel, M. C., Jolly, J. L., Matthews, M. S., Miller, E. M., Rambo-Hernandez, K. E., Rinn, A. N., Robins, J. H., & Wilson, H. E. (2020). A call to reframe gifted education as maximizing learning. *Phi Delta Kappan*, *102*(4), 22–25. https://doi.org/10.1177/0031721720978057

Dunbar, S. B., Welch, C. J., Hoover, H. D., & Frisbie, D. A. (2011). *Iowa Assessments, Form E.* Riverside Publishing.

Dunbar, S. B., Welch, C., Hoover, H. D., Forsyth, D. A., Frisbie, D. A., & Ansley, T. N. (2015). *Forms E and F: Research and development guide Iowa Assessments.* University of Iowa.

Ericsson, K. A. (2007). Deliberate practice and the modifiability of body and mind: Toward a science of the structure and acquisition of expert and elite performance. *International Journal of Sport Psychology, 38*(1), 4–34.

Erwin, J. O., & Worrell, F. C. (2012). Assessment practices and the underrepresentation of minority students in gifted and talented education. *Journal of Psychoeducational Assessment, 30*(1), 74-87. https://doi.org/10.1177/0743282911428197

Fair Housing Act, 42 U.S.C. § 3601 (1968). https://www.justice.gov/crt/fair-housing-act-2

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Feldhusen, J. F. (2005). Giftedness, talent, expertise, and creative achievement. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (p. 64–79). Cambridge University Press. https://doi.org/10.1017/CBO9780511610455.006

Finch, W. H., Bolin, J. E., & Kelley, K. (2014). *Multilevel modeling using R.* CRC Press.

Flynn, J. R. (2007). *What is intelligence? Beyond the Flynn effect.* Cambridge University Press. https://doi.org/10.1017/CBO9780511605253

Ford, D. Y. (2014). Segregation and the underrepresentation of Blacks and Hispanics in gifted education: Social inequality and deficit paradigms. *Roeper Review, 36*(3), 143-154. https://doi.org/10.1080/02783193.2014.919563

Ford, D. Y., Howard, T. C., Harris, J. J., & Tyson, C. A. (2000). Creating culturally responsive classrooms for gifted African American students. *Journal for the Education of the Gifted, 23*(4), 397–427. https://doi.org/10.1177/016235320002300405

Ford, D. Y., & Grantham, T. C. (2003). Providing access for culturally diverse gifted students: From deficit to dynamic thinking. *Theory Into Practice, 42*(3), 217-225. https://doi.org/10.1207/s15430421tip4203_8

Ford, D. Y., & Whiting, G. W. (2007). A mind is a terrible thing to erase: Black students' underrepresentation in gifted education. *Multiple Voices for Ethnically Diverse Exceptional Learners, 10*(1–2), 28–44. https://multiplevoicesjournal.org/doi/pdf/10.5555/muvo.10.1-2.604461l211051037

Ford, D. Y., Wright, B. L., & Scott, M. T. (2020). A matter of equity: Desegregating and integrating gifted and talented education for underrepresented students of color. *Multicultural Perspectives, 22*(1), 28-36. https://doi.org/10.1080/15210960.2020.1728275

Ford, D. Y., Grantham, T. C., & Whiting, G. W. (2008). Culturally and linguistically diverse students in gifted education: Recruitment and retention issues. *Exceptional Children, 74*(3)*,* 289-306. https://doi.org/10.1177/001440290807400302

Forthmann, B., Szardenings, C., & Holling, H. (2018). Understanding the confounding effect of fluency in divergent thinking scores: Revisiting average scores to quantify artifactual correlations. *Psychology of Aesthetics, Creativity, and the Arts, 14*(1), 94-112. http://dx.doi.org/10.1037/aca0000196

Foucault, M. (1995). *Discipline and punish: The birth of the prison.* Vintage Books.

Franck, K., & Rosen, E. (1949). A projective test of masculinity-femininity. *Journal of consulting psychology*, *13*(4), 247–256. https://doi.org/10.1037/h0057315

Frasier, M. M. (1997). Multiple criteria: The mandate and the challenge. *Roeper Review, 20*(2), A4-A6. https://doi.org/10.1080/02783199709553868

Frasier, M. M., García, J. H., & Passow, A. H. (1995). *A review of assessment issues in gifted education and their implications for identifying gifted minority students* (RM95204). Storrs: The National Research Center on the Gifted and Talented, University of Connecticut. http://nrcgt.uconn.edu/research-based_resources/frasgarc/

Gajda, A., Karwowski, M., & Beghetto, R. A. (2017). Creativity and academic achievement: A meta-analysis. *Journal of Educational Psychology, 109*(2), 269–299. https://doi.org/10.1037/edu0000133

Gagné, F. (2017). The integrative model of talent development (IMTD): From theory to educational applications. In J. A. Plucker, Rinn, A. N., & Makel, M. C. (Eds.),*From Giftedness to Gifted Education: Reflecting Theory in Practice* (pp. 149-182). Prufrock Press.

Garcia, N. M., Lopez, N., & Vélez, V. N. (2018). Quantcrit: Rectifying quantitative methods through critical race theory. *Race, Ethnicity, & Education, 21*(2), 149-157. https://doi.org/10.1080/13613324.2017.1377675

Gelman, A., & Hill, J. (2017). *Data analysis using regression and multilevel/hierarchical models.* Cambridge University Press.

Gentry, M., Gray, A., Whiting, G. W., Maeda, Y., & Pereira, N. (2020). *Gifted education in the United States: Laws, access, equity, and missingness across the country by locale, Title I school status, and race.* https://www.dropbox.com/s/0lxzznnyh5u0jj1/Access%20Denied.pdf

Gillborn, D., Warmington, P., & Demack, S. (2017). Quantcrit: Education, policy, 'big data' and principles for critical race theory of statistics. *Race, Ethnicity, and Education*, *21*(2), 158-179. https://doi.org/10.1080/13613324.2017.1377417

Gould, S. J. (1996). *The mismeasure of man.* W. W. Norton & Company, Inc.

Graham, J. W., & Olchowski, A. E. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science, 8*(3), 206-213. https://doi.org/10.1007/s11121-007-0070-9

Grantham, T. C. (2013). Creativity and equity: The legacy of E. Paul Torrance as an upstander for gifted Black males. *The Urban Review, 45*(4), 518–538. https://doi.org/10.1007/s11256-013-0257-2

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

Gubbins, E. J., Siegle, D., Ottone-Cross, K., McCoach, D. B., Langley, S. D., Callahan, C. M., Brodersen, A. V., & Caughey, M. (2021). Identifying and serving gifted and talented students: Are identification and services connected? *Gifted Child Quarterly, 65*(2), 115–131. https://doi.org/10.1177/0016986220988308

Guilford, J. P. (1967). Creativity: Yesterday, today, and tomorrow. *The Journal of Creative Behavior, 1*(1), 3–14. https://doi.org/10.1002/j.2162-6057.1967.tb00002.x

Hamilton, R., McCoach, D. B., Tutwiler, M. S., Siegle, D., Gubbins, E. J., Callahan, C. M., . . . Mun, R. U. (2017). Disentangling the roles of institutional and individual poverty in the identification of gifted students. *Gifted Child Quarterly*, *62*(1), 6-24. *https://doi.org/10.1177/0016986217738053*

Harrington, D. M., Block, J., & Block, J. H. (1983). Predicting creativity in preadolescence from divergent thinking in early childhood. *Journal of Personality and Social Psychology, 45*(3), 609–623. *https://doi.org/10.1037/0022-3514.45.3.609*

Henshon, S. E., & Grantham, T. C. (2021). Looking toward the future: An interview with Tarek C. Grantham. *Roeper Review, 43*(1), 3-6, https://doi.org/10.1080/02783193.2021.1840241

Hermon, A., Palomera, B., & Hokanson, B. (2018, October). *Post hoc analysis of TTCT results for creative ability* [Paper presentation]. Association for Educational Communications & Technology, Kansas City, MO. https://members.aect.org/pdf/Proceedings/proceedings18/2018/18_10.pdf

Hodges, J., Crutchfield, K., & Lee, L. E. (2021). Equitable access, locale, funding, and the alignment to performance ratings in Texas gifted programs. *Journal for the Education of the Gifted*, *44*(1), 3–30. https://doi.org/10.1177/0162353220978306

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.org/10.1177/0016986217752107

Hodges, J., Mun, R. U., & Rinn, A. N. (2021). Disentangling inequity in gifted education: The need for nuance in racial/ethnic categories, socioeconomic status, and geography

[Manuscript submitted for publication]. Department of Educational Psychology, University of North Texas.

Hoover, H. D., Dunbar, S. B., Frisbie, D. A., Oberley, K. R., Bray, G. B., Naylor, R. J., Lewis J. C., Ordman, V. L., & Qualls, A. L. (2007). *Iowa tests of basic skills: Norms and score conversions.* Riverside Publishing.

Iowa Test of Basic Skills: Guide to Research and Development (n.d). *Iowa Test of Basic Skills: Guide to research and development.* https://itp.education.uiowa.edu/ia/documents/itbs-research-guide.pdf

Jauk, E., Benedek, M., Dunst, B., & Neubauer, A. C. (2013). The relationship between intelligence and creativity: New support for the threshold hypothesis by means of empirical breakpoint detection. *Intelligence*, *41*(4), 212–221. https://doi.org/10.1016/j.intell.2013.03.003

Jolly, J. L., & Robins, J. H. (2018). History of gifted education. In J. L. Roberts, T. F. Inman, & J. H. Robins (Eds.), *Introduction to Gifted Education* (pp. 17-30). Prufrock Press.

Karwowski, M., Dul, J., Gralewski, J., Jauk, E., Jankowska, D. M., Gajda, A., Chruszczewski, M. H., & Benedek, M. (2016). Is creativity without intelligence possible? A necessary condition analysis. *Intelligence, 57,* 105–117. https://doi.org/10.1016/j.intell.2016.04.006

Kaufman, J. C. (2010). Using creativity to reduce ethnic bias in college admissions. *Review of General Psychology, 14*(3), 189-203. https://doi.org/10.1037/a0020133

Kaufman, J., & Plucker, J. (2011). Intelligence and creativity. In R. Sternberg & S. Kaufman (Eds.), *The Cambridge Handbook of Intelligence* (pp. 771-783). Cambridge University Press. https://doi.org/10.1017/CBO9780511977244.039

Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment.* John Wiley & Sons, Inc.

Kaufman, J. C., Plucker, J. A., & Russell, C. M. (2012). Identifying and assessing creativity as a component of giftedness. *Journal of Psychoeducational Assessment, 30*(1), 60-73. https://doi.org/10.1177/0734282911428196

Kim, K. H. (2006). Can we trust creativity tests? A review of the Torrance Test of Creative Thinking (TTCT). *Creativity Research Journal, 18*(1), 3-14. https://doi.org/10.1207/s15326934crj1801_2

Kim, K. H. (2008). Meta-analyses of the relationship of creative achievement to both IQ and divergent thinking test scores. *The Journal of Creative Behavior, 42*(2), 106–130. https://doi.org/10.1002/j.2162-6057.2008.tb01290.x

Kim, K. H. (2011) The creativity crisis: The decrease in creative thinking scores on the Torrance Tests of Creative Thinking. *Creativity Research Journal, 23*(4), 285-295. https://doi.org/10.1080/10400419.2011.627805

Kim, K. H. (2017). The Torrance Tests of Creative Thinking - Figural or Verbal: Which one should we use? *Creativity. Theories – Research - Applications, 4*(2), 302-321. https://doi.org/10.1515/ctra-2017-0015

Kim., K. H., Cramond, B., & Bandalos, D. L. (2006). The latent structure and measurement invariance of scores on the Torrance Tests of Creative Thinking-Figural. *Educational and Psychological Measurement, 66*(3), 459-477. https://doi.org/10.1177/0013164405282456

Kokštejn, J., Musálek, M., Tufano, J.J. (2017). Are sex differences in fundamental motor skills uniform throughout the entire preschool period? *PLOS ONE 12*(4): e0176556. https://doi.org/10.1371/journal.pone.0176556

Ladson-Billings, G. (1998). Just what is critical race theory and what's it doing in a nice field like education? *International Journal of Qualitative Studies in Education, 11*(1), 7-24. https://doi.org/10.1080/095183998236863

Lakin, J. (2018). Making the cut in gifted selection: Score combination rules and their impact on program diversity. *Gifted Child Quarterly, 62*(2), 210-219. https://doi.org/10.1177/0016986217752099

Lakin, J. (2021). Service-identification alignment: Why the services you offer must drive identification decisions. [Manuscript submitted for publication]. *Teaching for High Potential.* https://drive.google.com/file/d/17qoHlyrX_8_GZAow9EqFm4hweWromqQI/view

Lee, L. E., & Peters, S. J. (2021). Universal screening: A process to promote equity [Manuscript submitted for publication]. In S. Johnsen & J. VanTassel-Baska (Eds.), *Handbook on Assessments for Gifted Learners.* Prufrock Press.

Lee, L. E., & Rinn, A. N. (2021). The search for creativity and talent in diverse populations: A systematic review [Manuscript submitted for publication]. Department of Educational Psychology, University of North Texas.

Lee., L. E., Ottwein, J. K., & Peters, S. J. (2020). Eight universal truths of identifying students for advanced academic interventions. In J. L. Jolly & J. H. Robins (Eds.), *Methods & Materials for Teaching the Gifted* (5th ed., pp. 61-79). Prufrock Press.

Lee, H., Karakis, N., Olcay Akce, B., Azzam Tuzgen, A., Karami, S., Gentry, M., & Maeda, Y. (2021). A meta-analytic evaluation of Naglieri Nonverbal Ability Test: Exploring its validity evidence and effectiveness in equitably identifying gifted students. *Gifted Child Quarterly, 65*(3), 199–219. https://doi.org/10.1177/0016986221997800

Little, R. J. A. (1988). Missing data adjustments in large surveys (with discussion). *Journal of Business Economics and Statistics, 6*(1), 287-296. https://doi.org/10.1080/07350015.1988.10509663

Lohman, D. F. (2005a). An aptitude perspective on talent: Implications for identification of academically gifted minority students. *Journal for the Education of the Gifted, 28*(3-4), 333-360. https://doi.org/10.4219/jeg-2005-341

Lohman, D. F. (2005b). *Identifying academically talented minority students*. https://nrcgt.uconn.edu/wp-content/ uploads/sites/953/2015/04/rm05216.pdf

Lohman, D. F. (2009). Identifying academically talented students: Some general principles, two specific procedures. In L. V. Shavinina (Ed.), *International handbook on giftedness* (pp. 971-997). Springer Science + Business Media.

Lohman, D. F. (2012). *Cognitive Abilities Test Form 7: Research and development guide*. Riverside Assessments, LLC.

Lohman, D. F., & Gambrell, J. L. (2012). Using nonverbal tests to help identify academically talented children. *Journal of Psychoeducational Assessment, 30*(1), 25-44. https://doi.org/10.1177/0734282911428194

Lohman, D. F., Korb, K. A., & Lakin, J. M. (2008). Identifying academically gifted English-Language Learners using nonverbal tests: A comparison of the Raven, NNAT, and CogAT. *Gifted Child Quarterly, 52*(4), 275–296. https://doi.org/10.1177/0016986208321808

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Lawrence Erlbaum Associates, Inc.

Lüdecke, D., Makowski, D., Waggoner, P., & Patil, I. (2019). Performance: Assessment of regression models performance (R package version 0.4.7). https://cran.r-project.org/web/packages/performance/performance.pdf

Luria, S. R., O'Brien, R. L., & Kaufman, J. C. (2016). Creativity in gifted identification: Increasing accuracy and diversity. *Annals of the New York Academy of Sciences, 1377*(1), 44-52. https://doi.org/10.1111/nyas.13136

Maker, C. J. (2005). *The DISCOVER Project: Improving assessment and curriculum for diverse gifted learners* (RM05206). University of Connecticut, The National Research Center on the Gifted and Talented.

Mansfield, K. C. (2015). Giftedness as property: Troubling whiteness, wealth, and gifted education in the United States. *International Journal of Multicultural Education, 17*(1), 1–18. https://doi.org/10.18251/ijme.v17i1.841

Marland, S. P. (1972). Education of the gifted and talented: Report to the Congress of the United States by the U.S. Commissioner of Education and background papers submitted to the U.S. Office of Education. 2 vols (Government Documents, Y4.L 11/2: G36). Government Printing Office.

McBee, M. T. (2006). A descriptive analysis of referral sources for gifted identification screening by race and socioeconomic status. *Journal of Secondary Gifted Education*, *17*(2), 103–111. https://doi.org/10.4219/jsge-2006-686

McBee, M. (2010). Examining the probability of identification for gifted programs for students in Georgia elementary schools: A multilevel path analysis study. *Gifted Child Quarterly, 54*(4), 283-297. https://doi.org/10.1177/0016986210377927

McBee, M. (2021). giftedCalcs: R package for gifted identification psychometrics. R version 0.1.3.

McBee, M. T., & Makel, M. C. (2019). The quantitative implications of definitions of giftedness. *AERA Open*. *https://doi.org/10.1177/2332858419831007*

McBee, M. T., Peters, S. J., & Waterman, C. (2014). Combining scores in multiple-criteria assessment systems: The impact of combination rule. *Gifted Child Quarterly*, *58*(1), 69-89. https://doi.org/10.1177/0016986213513794

McBee, M. T., Peters, S. J., & Miller, E. M. (2016). The impact of the nomination stage on gifted program identification: A comprehensive psychometric analysis. *Gifted Child Quarterly, 60*(4), 258-278. https://doi.org/10.1177001169886216656256

Menard, S. (2002). *Applied logistic regression analysis* (2nd ed). Sage University Press.

Messick, S. (1989). *Validity.* In R. L. Linn (Ed.), *The American Council on Education/Macmillan series on higher education. Educational measurement* (p. 13–103). Macmillan Publishing Co, Inc.; American Council on Education.

Mun, R. U., Langley, S. D., Ware, S., Gubbins, E. J., Siegle, D., Callahan, C. M., McCoach, D. B., & Hamilton, R. (2016). *Effective practices for identifying and serving English Learners in gifted education: A systematic review of the literature*. National Center for Research on Gifted Education (NCRGE). https://ncrge.uconn.edu/wp-content/uploads/sites/982/2016/01/NCRGE_EL_Lit-Review.pdf

Mun, R. U., Ezzani, M., & Lee, L. E. (2020). Culturally relevant leadership in gifted education: A systematic review. *Journal for the Education of the Gifted*, *43*(2), 108-142. https://doi.org/10.1177/0162353220912009

Mun, R. U., Ezzani, M. D., Lee, L. E., & Ottwein, J. K. (2021). Building systemic capacity to improve identification and services in gifted education: A case study of one district. *Gifted Child Quarterly*, *65*(2), 132–152. https://doi.org/10.1177/0016986220967376

Naglieri, J. A. (2008). *Naglieri Nonverbal Ability Test-2.* Pearson.

Naglieri, J. A., & Ford, D. Y. (2003). Addressing underrepresentation of gifted minority children using the Naglieri Nonverbal Ability Test (NNAT). *Gifted Child Quarterly*, *47*(2), 155-160. https://doi.org/10.1177/001698620304700206

Naglieri, J. A., & Ford, D. Y. (2005). Increasing minority children's participation in gifted classes using the NNAT: A response to Lohman. *Gifted Child Quarterly*, *49*(1), 29-36. https://doi.org/10.1177/001698620504900104

National Association of Gifted Children and Council of State Program Directors for the Gifted (2015). *2014-2015 State of the states in gifted education: Policy and practice data*. https://www.nagc.org/sites/default/files/key%20reports/2014-2015%20State%20of%20the%20States%20(final).pdf

National Center for Education Statistics. (2015). *Education demographic and geographic estimates program (EDGE): Locale boundaries user's manual.* https://nces.ed.gov/programs/EDGE/docs/NCES_LOCALE_USERSMANUAL_2016012.pdf

National Center for Education Statistics. (2019). *Fast facts: Back to school statistics.* https://nces.ed.gov/fastfacts/display.asp?id=372#PK12_enrollment

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbrina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*(2), 77-101. https://doi.org/10.1037/0003-066X.51.2.77

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist, 67*(2), 130–159. https://doi.org/10.1037/a0026699

Northwest Evaluation Association. (2011). *MAP® V3 for Colorado—Instructional statements.* http://www.nwea.org

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

Pampel, P. C. (2000). *Logistic regression: A primer.* Sage University Press. https://dx.doi.org/10.4135/9781412984805

Pargulski, J. R., & Reynolds, M. R. (2016). Sex differences in achievement matter. *Personality and Individual Differences, 104*(1), 272-278. https://doi.org/10.1016/j.paid.2016.08.016

Peters, S. J. (2021). The challenges of achieving equity within public school gifted and talented programs. *Gifted Child Quarterly*. https://doi.org/10.1177/00169862211002535

Peters, S. J., Carter, J., & Plucker, J. A. (2020a). Rethinking how we identify "gifted" students. *Phi Delta Kappan, 102(4), 8–13*. https://doi.org/10.1177/0031721720978055

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.org/10.1177/0016986216643165

Peters, S. J., Gentry, M., Whiting, G. W., & McBee, M. T. (2019a). Who gets served in gifted education? Demographic representation and a call for action. *Gifted Child Quarterly, 63*(4), 273-287. https://doi.org/10.1177/0016986219833738

Peters, S. J., & Pereira, N. (2017). A replication of the internal validity structure of three major teaching rating scales. *Journal of Advanced Academics*, *28*(2), 101–119. https://doi.org/10.1177/1932202X17701940

Peters, S. J., Rambo-Hernandez, K., Makel, M. C., Matthews, M. S., & Plucker, J. A. (2019b). Effect of local norms on racial and ethnic representation in gifted education. *AERA Open, 5*(2), 1-18. https://doi.org/10.1177/2332858419848446

Peters, S. J., Ottwein, J. K., Lee, L. E., & Matthews, M. S. (2020b). Identification. In J. A. Plucker & C. M. Callahan (Eds.), *Critical Issues and Practices in Gifted Education* (3rd ed., pp. 261-272). Prufrock Press.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74*(4), 525-556. http://www.jstor.org/stable/3515980

Pfeiffer, S. I., & Jarosewich, T. (2003). *Gifted Rating Scales.* Harcourt Assessment

Plucker, J. A., & Runco, M. A. (1998). The death of creativity measurement has been greatly exaggerated: Current issues, recent advances, and future directions in creativity assessment. *Roeper Review, 21*(1), 36-39. https://doi.org/10.1080/02783199809553924

Plucker, J. A. (2011). Reanalysis of student responses to creativity checklists: Evidence of content generality. *Journal of Creative Behavior*, *33*(2), 136–137. https://doi.org/10.1002/j.2162-6057.1999.tb01042.x

Plucker, J. A., & Callahan, C. M. (2014). Research on giftedness and gifted education: Status of the field and considerations for the future. *Exceptional Children, 80*(4), 390-406. https://doi.org/10.1177/0014402914527244

Plucker, J. A., Makel, M. C., Matthews, M. S., Peters, S. J., & Rambo-Hernandez, K. E. (2017a). Blazing new trails: Strengthening policy research in gifted education. *Gifted Child Quarterly, 61*(3), 210-218. https://doi.org/10.1177/0016986217701838

Plucker, J. A., Peters, S. J., & Schmalensee, S. (2017b). Reducing excellence gaps: A research-based model. *Gifted Child Today, 40*(4), 245-250. https://doi.org/10.1177/1076217517723949

Plucker, J. A., & Peters, S. J. (2016). *Excellence gaps in education: Expanding opportunities for talented students.* Harvard Education Press.

Plucker, J. A., & Peters, S. J. (2018). Closing poverty-based excellence gaps: Conceptual, Measurement, and educational issues. *Gifted Child Quarterly, 62*(1), 56-67. https://doi.org/10.1177/0016986217738566

Popham, W. J. (1997), Consequential validity: Right Concern-Wrong Concept. *Educational Measurement: Issues and Practice, 16*(2), 9-13. https://doi-org.libproxy.library.unt.edu/10.1111/j.1745-3992.1997.tb00586.x

Popham, W. J. (2012). *Assessment bias: How to banish it* (2nd ed.). Pearson.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage Publications.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales: Section 1. General overview.* Oxford Psychologists Press.

Reilly, D., Neumann, D. L., & Andrews, G. (2019). Gender differences in reading and writing achievement: Evidence from the National Assessment of Educational Progress (NAEP). *American Psychologist, 74*(4), 445–458. https://doi.org/10.1037/amp0000356

Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan, 60*(3), 180-184, 261.

Renzulli, J. S. (2016). The three-ring conception of giftedness: A developmental model for promoting creative productivity. In S. M. Reis (Ed.), *Reflections on gifted education: Critical works by Joseph S. Renzulli and colleagues* (pp. 55–90). Prufrock Press Inc.

Renzulli, J. S., & Hartman, R. (1971). Scale for rating behavioral characteristics of superior students. *Exceptional Children, 38*(1), 243–248. https://www.nzcer.org.nz/node/45917

Ricciardi, C., Haag-Wolf, A., & Winsler, A. (2020). Factors associated with gifted identification for ethnically diverse children in poverty. *Gifted Child Quarterly, 64*(4), 243–258. https://doi.org/10.1177/0016986220937685

Rinn, A. N., Mun, R. U., & Hodges, J. (2020). *2018-2019 State of the states in gifted education*. National Association of Gifted Children and the Council of State Directors of Programs for the Gifted. https://www.nagc.org/2018-2019-state-states-gifted-education

RStudio Team. (2020). RStudio: Integrated development for R, Version 1.2.1335. RStudio, Inc. https://rstudio.com

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of American Statistical Association, 91*(434), 473-489. https://doi.org/10.1080/01621459.1996.10476908

Runco, M. A. (1993). Divergent thinking, creativity, and giftedness. *Gifted Child Quarterly, 37*(1), 16–22. https://doi.org/10.1177/001698629303700103

Runco, M. A., & Selcuk, A. (2012) Divergent thinking as an indicator of creative potential. *Creativity Research Journal, 24*(1), 66-75. https://doi.org/10.1080/10400419.2012.652929

Ryser, G. R., & McConnell, K. (2004). Scales for Identifying Gifted Students. Prufrock Press.

Sablan, J. R. (2018). 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

Said-Metwaly, S., Fernández-Castilla, B., Kyndt, E., & Van den Noortgate, W. (2018). The factor structure of the Figural Torrance Tests of Creative Thinking: A meta-confirmatory factor analysis. *Creativity Research Journal, 30*(4), 352-360. https://doi.org/10.1080/10400419.2018.1530534

Shi, B., Wang, L., Yang, J., Zhang, M., & Xu, L. (2017). Relationship between divergent thinking and intelligence: An empirical study of the threshold hypothesis with Chinese children. *Frontiers in psychology*, *8*(254), 1-9. https://doi.org/10.3389/fpsyg.2017.00254

Siegle, D., Gubbins, E. J., O'Rourke, P., Langley, S. D., Mun, R. U., Luria, S. R., Little, C. A., McCoach, D. B., Knupp, T., Callahan, C. M., & Plucker, J. A. (2016). Barriers to underserved students' participation in gifted programs and possible solutions. *Journal for the Education of the Gifted*, *39*(2), 103–131. https://doi.org/10.1177/0162353216640930

Stage, F. K., & Wells, R. S. (2014). Critical quantitative inquiry in context. *New Directions for Institutional Research, 2013*(158), 1–7. https://doi.org/10.1002/ir.20041

Sternberg, R. J. (1983). Components of human intelligence. *Cognition, 15*(1-3), 1–48. https://doi.org/10.1016/0010-0277(83)90032-X

Sternberg, R. J. (2018) Creative giftedness is not just what creativity tests test: Implications of a triangular theory of creativity for understanding creative giftedness, *Roeper Review, 40*(3), 158-165, http://doi.org/10.1080/02783193.2018.1467248

Stewart, R. A., Rule, A. C., & Giordano, D. A. (2007). The effect of fine motor skill activities on kindergarten student attention. *Early Childhood Education Journal, 35*(2), 103–109 (2007). https://doi.org/10.1007/s10643-007-0169-4

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.org/10.1177/1529100611418056

Tannenbaum, A. J. (2003). Nature and nurture of giftedness. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (3rd ed., pp. 45-59). Allyn & Bacon.

Texas Education Code § 29.121 (2019). https://statutes.capitol.texas.gov/Docs/ED/htm/ED.29.htm

Texas Education Agency. (2019). Texas state plan for the education of gifted/talented students (Rev.). https://tea.texas.gov/Academics/Special_Student_Populations/Gifted_and_Talented_Education/Gifted_Talented_Education

Texas Education Agency. (2020). 2018-2019 Texas academic performance reports. https://rptsvr1.tea.texas.gov/perfreport/tapr/2019/index.html

Tjur, T. (2009) Coefficients of determination in logistic regression models—A new proposal: The coefficient of discrimination. *The American Statistician, 63*(4), 366-372, https://doi.org/10.1198/tast.2009.08210

Torrance, E. P. (1971). Are the Torrance Tests of Creative Thinking biased against or in favor of the "disadvantaged" groups? *Gifted Child Quarterly, 15*(2)*,* 75–80. https://doi.org/10.1177/001698627101500201

Torrance, E. P. (1974). *The Torrance Tests of Creative Thinking—Norms, Technical Manual Research Edition—Verbal Tests, Forms A and B—Figural Tests, Forms A and B*. Personnel Press.

Torrance, E. P. (2004). The role of creativity in identification of the gifted and talented. In D. J. Treffinger (Ed.), *Creativity and Giftedness* (Vol. 10, pp. 79-86). Corwin Press.

Torrance, E. P. (2008). *Torrance Tests of Creative Thinking norms-technical manual Figural (Streamlined) Forms A & B.* Scholastic Testing Service.

Torrance, E. P. (2017). *Torrance Tests of Creative Thinking norms-technical manual for Figural Forms A and B.* Scholastic Testing Service.

Treffinger, D. J. (2004). Research on creativity. In D. J. Treffinger (Ed.), *Creativity and Giftedness* (Vol. 10, pp. 87-96). Corwin Press.

Van Buuren, S., Groothuis-Oudshoorn, K., Vink, G., Schouten, R., Robitzsch, A., Doove, L., Jolani, S., Moreno-Betancur, M., White, I., Gaffert, P., Meinfelder, F., & Gray, B. (2020). Mice: Multivariate imputation by chained equations (Version 3.9.0). https://cran.r-project.org/web/packages/mice/mice.pdf

VanTassel-Baska, J. (2008). An overview of alternative assessment measures for gifted learners and the issues that surround their use. In J. L. VanTassel-Baska (Ed.), *Alternative assessments with gifted and talented students* (pp. 1-15). Prufrock Press.

VanTassel-Baska, J., Feng, A. X., & Evans, B. L. (2007). Patterns of identification and performance among gifted students identified through performance tasks: A three-year analysis. *Gifted Child Quarterly*, *51(3)*, 218-231. https://doi.org/10.1177/0016986207302717

Urban, K. K., & Jellen, H. G. (2010). Test for Creative Thinking–Drawing Production (TCT-DP): Manual. Pearson Assessment & Information.

U. S. Department of Education. (2020). *Civil rights data collection: 2017-18 State and national estimations.* https://ocrdata.ed.gov/estimations/2017-2018

Wai, J., Cacchio, M., Putallaz, M., & Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30-year examination. *Intelligence, 38*(4), 412-423. https://doi.org/10.1016/j.intell.2010.04.006

Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity-intelligence distinction.* Holt, Rinehart, & Winston.

Warne, R. T. (2015). Test review: Cognitive abilities test, Form 7 (CogAT7). *Journal of Psychoeducational Assessment, 33*(2), 188–192. https://doi.org/10.1177/0734282914548324

Warne, R. T., Anderson, B., & Johnson, A. O. (2013). The impact of race and ethnicity on the identification process for giftedness in Utah. *Journal for the Education of the Gifted*, *36*(4), 487–508. https://doi.org/10.1177/0162353213506065

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D. (2020). Ggplot2: Create elegant data visualisations Using the Grammar of Graphics. R Package version 3.3.3. https://cran.r-project.org/web/packages/ggplot2/index.html

Worrell, F. C., & Dixson, D. D. (2018). Recruiting and retaining underrepresented gifted students. In S. I. Pfeiffer (Ed.), *Handbook of Giftedness in Children* (2nd ed., pp. 209-226). American Psychological Association. https://doi.org/10.1007/978-3-319-77004-8_13

Worrell, F. C., & Dixson, D. D. (2020). Diversity and gifted education. In J. A. Plucker & C. M. Callahan (Eds.), *Critical Issues and Practices in Gifted Education* (3rd ed., pp. 169-183). Prufrock Press.

Yoon, C-H. (2017). A validation study of the Torrance Tests of Creative Thinking with a sample of Korean elementary school students. *Thinking Skills and Creativity, 26*(2017), 38-50. https://doi.org/10.1016/j.tsc.2017.05.004

Yoon, S., & Gentry, M. (2009). Racial and ethnic representation in gifted programs: Current status of and implications for gifted Asian American students. *Gifted Child Quarterly*, *53*(2), 121-136. https://doi.org/10.1177/0016986208330564

Zhang, Z., Mai, Y., & Yang, M. (2021). WebPower: Basic and advanced statistical power analysis. R Package version 0.6. https://cran.r-project.org/web/packages/WebPower/WebPower.pdf

1.6 Data Tables and Figures

Table 1.1

Frequency and Proportions of the Total and Disaggregated Student Samples

		Total Sample	Referrals	Universal Screening
Age[a]		7.31%	9.31%	6.17%
Grade	1	2089(64.06%)	19 (1.6%)	2070 (100%)
	2	277 (8.49%)	277 (23%)	
	3	228 (6.99%)	228 (19%)	
	4	219 (6.72%)	219 (18%)	
	5	185 (5.67%)	185 (16%)	
	6	127 (3.89%)	127 (11%)	
	7	107 (3.28%)	107 (9.0%)	
	8	29 (0.89%)	29 (2.4%)	
Race/Ethnicity	Black/African American	535 (16%)	177 (14.86%)	358 (17%)
	Hispanic/Latinx	857 (26%)	254 (21.33%)	603 (29%)
	Asian	142 (4.4%)	61 (5.12%)	81 (3.9%)
	Native American	69 (2.1%)	27 (2.27%)	42 (2.0%)
	White	1,611 (49%)	639 (53.65%)	972 (47%)
	Other[b]	47 (1.4%)	33 (2.77%)	14 (0.7%)
Sex	Female	1,589 (49%)	579 (48.61%)	1,010 (49%)
	Male	1672 (51.27%)	612 (51.39%)	1060 (51.21%)
Free/Reduced Lunch Status	Yes	1,212 (37%)	409 (34.34%)	803 (39%)
	No	2049 (62.83%)	782 (65.66%)	1267 (61.21%)
English Learner Status	Yes	501 (15%)	173 (14.53%)	328 (16%)
	No	2760 (84.64%)	1018 (85.5%)	1742 (84.15%)
GT ID	Yes	401 (12%)	273 (22.92%)	128 (6.2%)
	No	2860 (87.70%)	918 (77.1%)	1942 (93.8%)
Total Sample		3261	1191	2070

Note. GT ID = Identified for gifted services. a = mean age. b = Other comprises both students categorized as two or more races and Native Hawaiian/Pacific Islanders.

Table 1.2

Means and Standard Deviations Per Student Demographic Group for Referrals

Group	TTCT A or B		CogAT Non-Verbal		CogAT Quantitative		CogAT Verbal		ITBS Reading		ITBS Math	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Black/African American	104.14	13.44	106.60	11.67	106.81	11.65	107.70	10.65	64.11	20.5	61.49	27.72
Hispanic/Latinx	102.90	14.64	109.48	11.07	107.88	10.94	103.11	11.74	68	21.25	67.63	25.56
Asian	109.28	16.22	112.52	10.28	114.52	10.37	107.09	11.30	62.18	23.40	77.72	19.31
American Indian/ Alaska Native	99.44	13.68	106.70	11.79	103.89	13.68	105.70	11.16	64.78	19.56	60.26	28.30
White	105.94	14.40	110.71	11.64	110.37	11.54	110.07	11.60	72.92	18.71	69.17	23.11
Other	106.70	14.65	110.52	8.66	105.21	10.26	105.46	8.72	63.15	18.28	59.27	23.01
Female	106.96	13.88	109.64	11.17	107.58	10.81	108.64	11.51	71.05	19.91	64.66	24.52
Male	103.28	14.63	110.02	11.78	110.80	12.04	107.11	11.86	68.16	20.22	70.50	24.33
Free/Reduced Lunch	103.67	14.26	108.33	11.45	106.68	11.64	104.15	11.30	65.51	21.28	65.07	25.28
Full-Priced Lunch	105.80	14.40	110.63	11.43	110.56	11.30	109.79	11.46	71.69	19.15	69.02	24.12
English Learner	102.14	13.58	109.09	11.14	107.98	11.28	99.51	9.40	62.83	23.52	69.73	24.15
Native English	105.57	14.46	109.96	11.54	109.44	11.60	109.43	12.31	70.71	19.26	67.30	24.66
Total	105.07	14.38	109.84	11.48	109.23	11.56	107.85	11.71	69.57	20.11	67.66	24.59

Note. These estimates are after multiple imputation. This is comprised of Grades 1st through 8th (n = 1191). First graders included in this group were not administered the NNAT. TTCT Form A had a total mean of 104.68 and standard deviation of 14.30. TTCT Form B had a total mean of 104.01 and standard deviation of 14.26. CogAT7 = Cognitive Abilities Test. ITBS = Iowa Test of Basic Skills. TTCT = Torrance Test of Creative Thinking. Other category consists of two or more races and Native Hawaiian/Pacific Islander.

Table 1.3

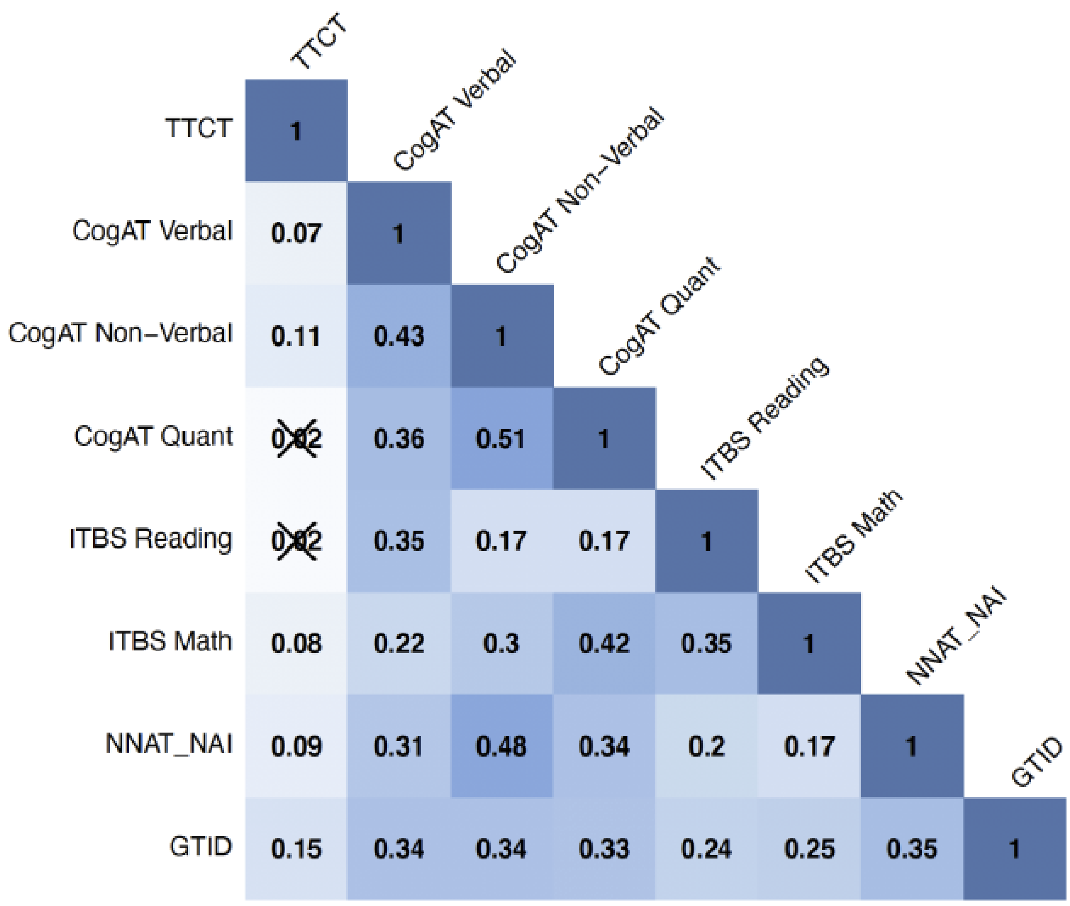
Means and Standard Deviations Per Student Demographic Group for Universal Screening

Group	TTCT A or B		CogAT Non-Verbal		CogAT Quantitative		CogAT Verbal		ITBS Reading		ITBS Math		NNAT	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Black/African American	106.63	13.38	95.85	10.82	101.62	12.21	104.73	10.67	85.36	13.59	57.85	28.97	92.22	14.07
Hispanic/Latinx	100.48	14.00	95.81	11.16	102.29	12.53	100.66	11.04	82.32	15.25	57.99	28.54	96.18	13.19
Asian	97.36	16.55	100.23	12.20	111.60	11.65	100.60	12.12	86.04	14.66	68.68	24.12	105.63	13.50
American Indian/ Alaska Native	101.62	15.04	101.05	10.84	93.57	10.94	115.52	8.30	95.14	5.23	80.43	19.26	95.83	12.40
White	103.28	13.58	101.66	12.97	106.79	12.74	105.24	10.85	82.48	15.42	66.65	26.23	101.40	13.37
Other	106.07	13.64	113.29	11.10	100.21	10.94	112.50	8.76	62.86	13.68	83.50	17.14	97.93	11.45
Female	104.39	13.77	101.10	12.18	104.51	12.71	107.17	9.93	86.38	13.70	62.10	28.00	98.78	13.76
Male	101.27	14.07	96.92	12.26	104.42	13.03	100.77	11.46	80.16	15.77	64.00	27.30	97.89	14.08
Free/Reduced Lunch	103.22	14.22	97.53	11.60	101.09	11.77	103.12	11.01	85.44	14.24	62.12	27.30	93.88	13.52
Full-Priced Lunch	102.53	13.87	99.87	12.80	106.60	13.09	104.38	11.31	81.78	15.48	63.68	27.87	101.13	13.44
English Learner	96.91	14.49	95.13	11.12	101.89	12.26	103.10	11.08	83.47	14.65	72.71	23.69	95.61	12.91
Native English	103.90	13.64	99.68	12.50	104.95	12.93	104.04	11.23	83.15	15.20	61.26	27.98	98.83	14.06
Total	102.80	14.01	98.96	12.40	104.46	12.88	103.89	11.21	83.20	15.12	63.08	27.65	98.32	13.93

Note. These estimates are after multiple imputation. This only includes students in 1st grade (*n* = 2070). TTCT Form A had a total mean of 101.97 and standard deviation of 13.62. TTCT Form B had a total mean of 105.95 and standard deviation of 8.74. CogAT7 = Cognitive Abilities Test. ITBS = Iowa Test of Basic Skills. TTCT = Torrance Test of Creative Thinking. Other category consists of two or more races and Native Hawaiian/Pacific Islander.

Figure 1.2

Correlation Matrix for Total Sample



Note. Includes total sample after imputation. X indicates the correlation coefficient was $p > .05$.

Figure 1.3

Bar Plot of Total Students Identified by Race/Ethnicity

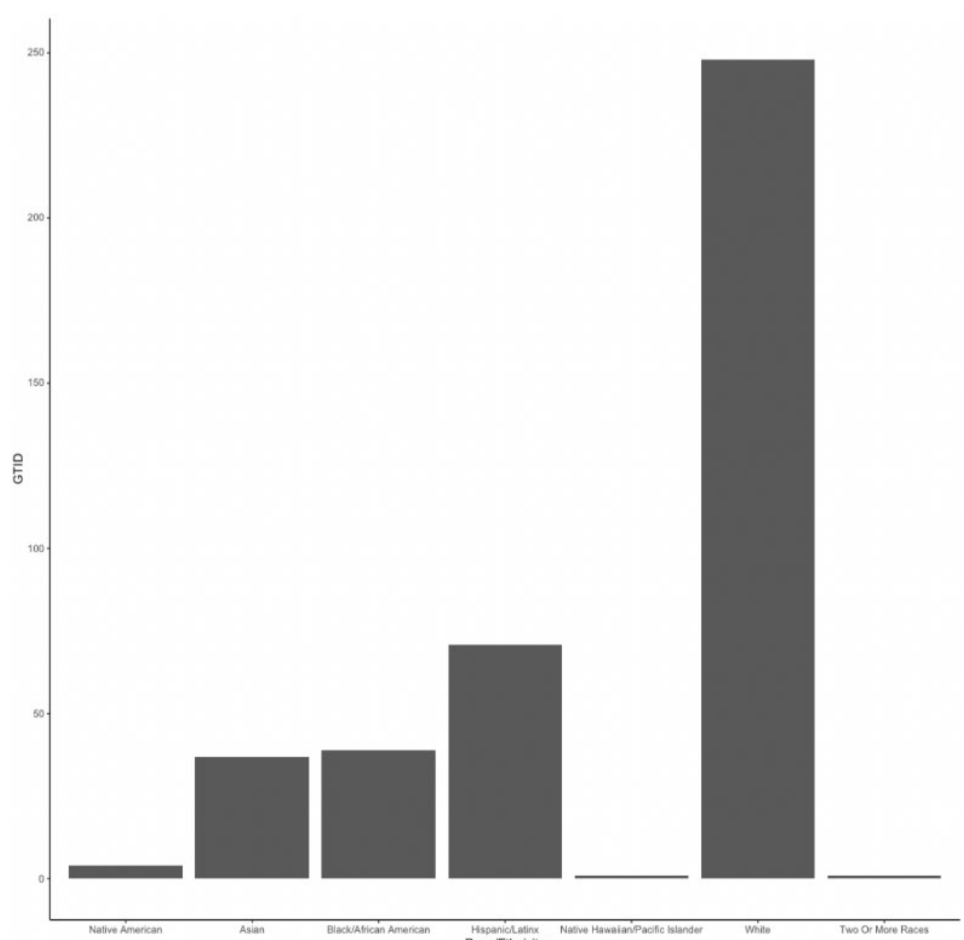


Figure 1.4

Bar Plot of Total Students Identified by Sex

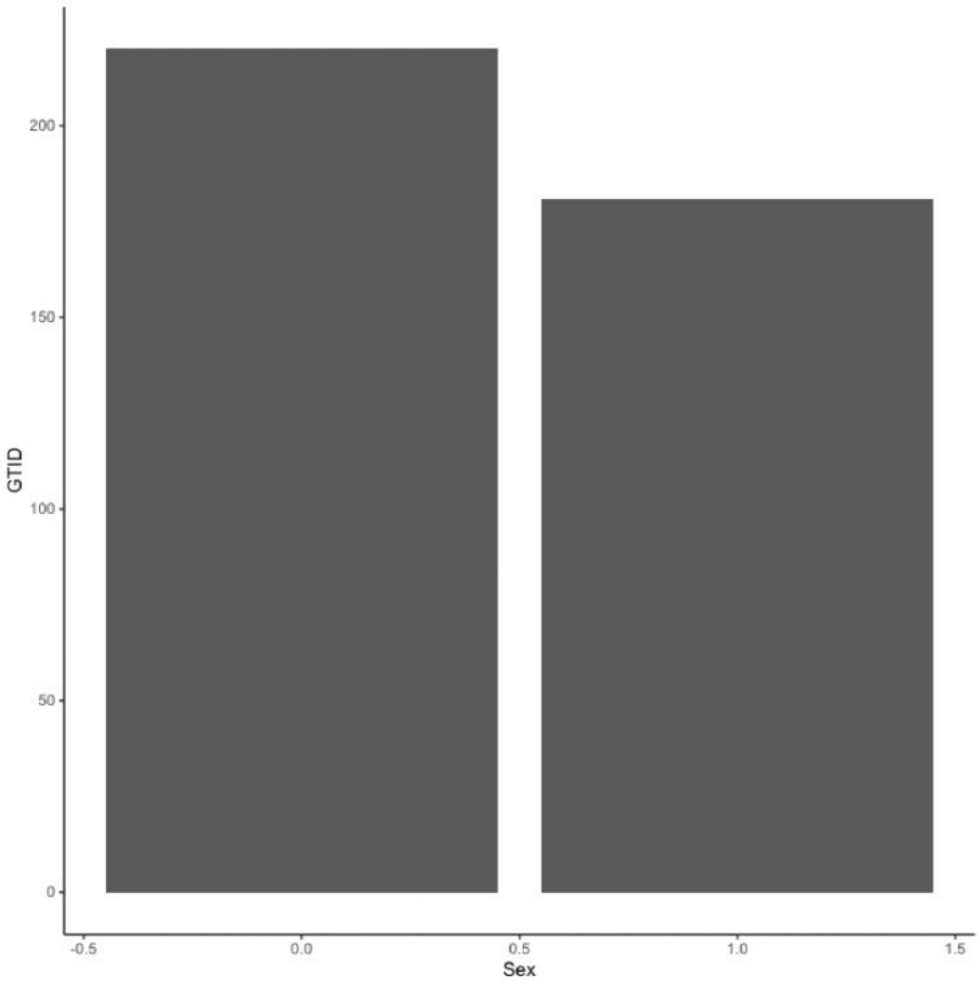


Figure 1.5

Bar Plot of Total Students Identified by Free/Reduced Lunch Status

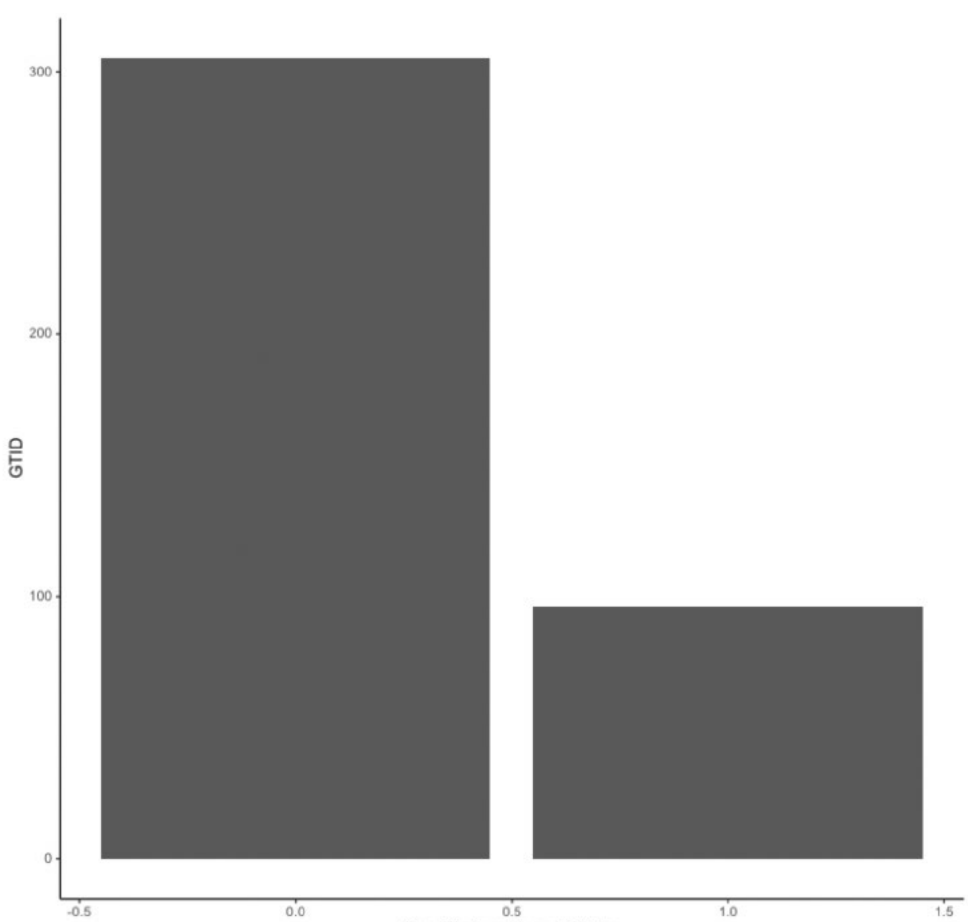


Figure 1.6

Bar Plot of Total Students Identified by English Learner Status

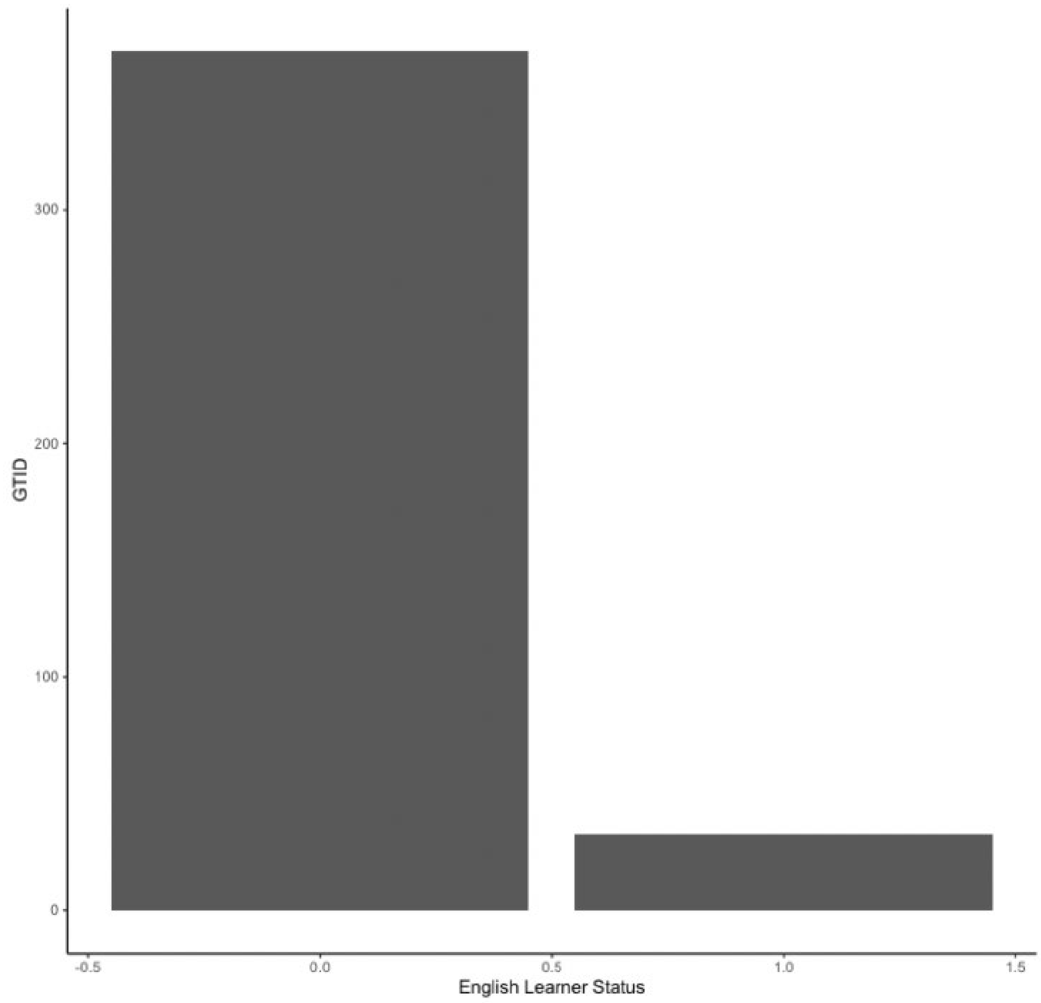
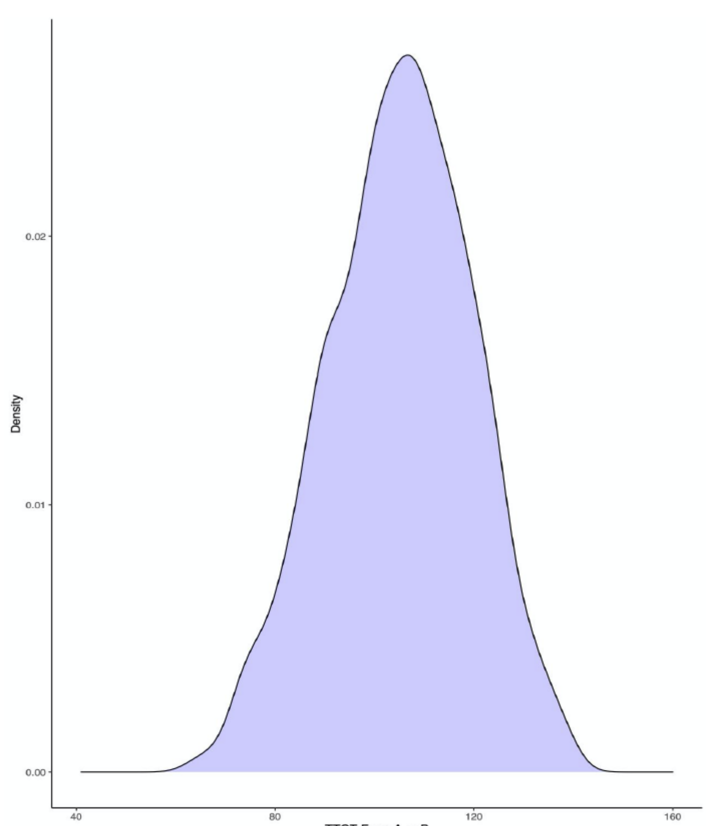


Figure 1.7

Density Curve Plot of Scores on the TTCT Form A or B for Referrals Across First through Eighth Grade



Note. This represents the distribution of scores of TTCT Form A or B for students who were referred.

Figure 1.8

Density Curve Plot of Scores on the CogAT Non-Verbal Referrals Across First through Eighth Grade

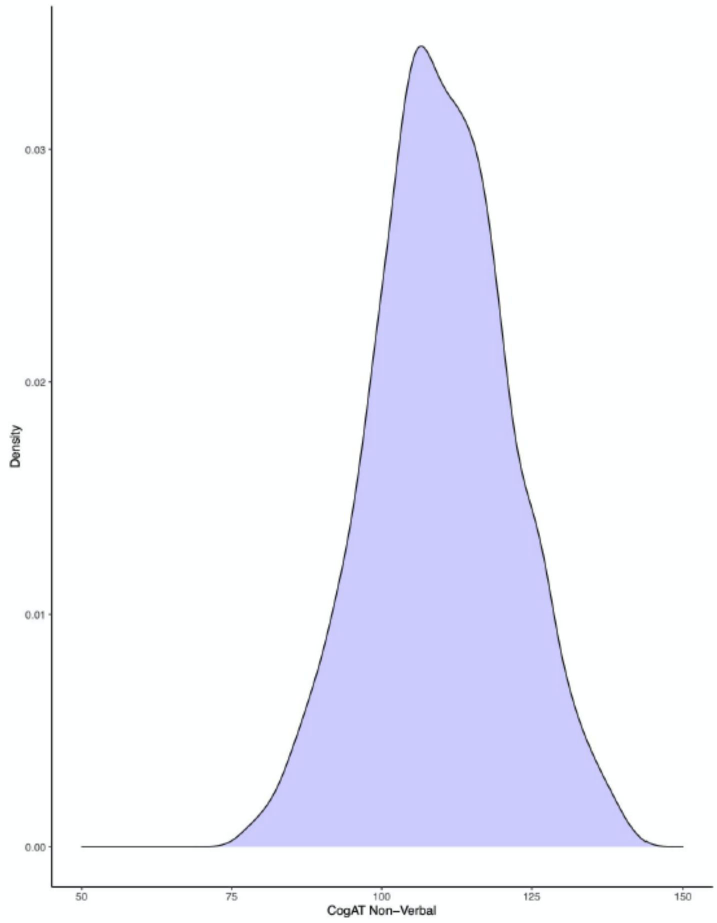


Figure 1.9

Density Curve Plot of Scores on the CogAT Quantitative Referrals Across First through Eighth Grade

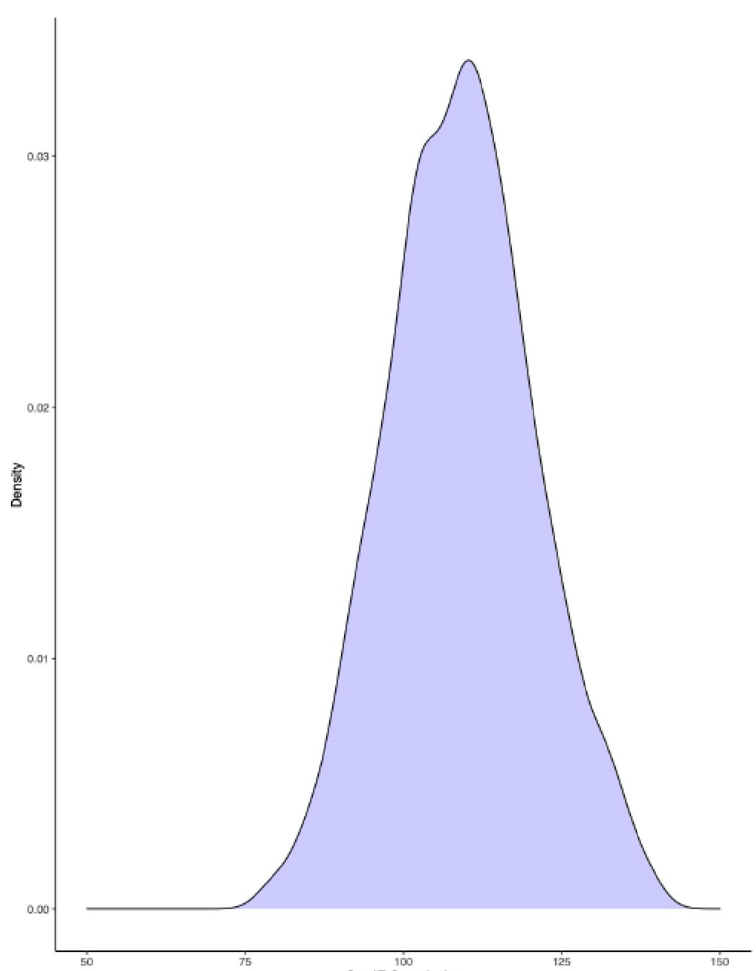


Figure 1.10

Density Curve Plot of Scores on the CogAT Verbal for Referrals Across First through Eighth Grade

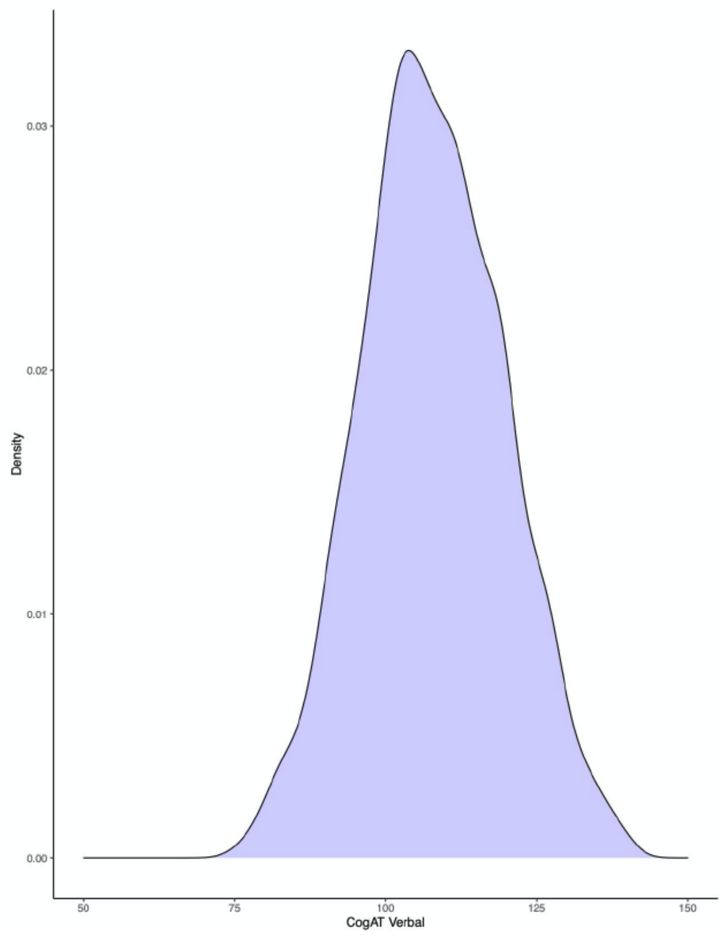


Figure 1.11

Density Curve Plot of Scores on the ITBS Reading for Referrals Across First through Eighth Grade

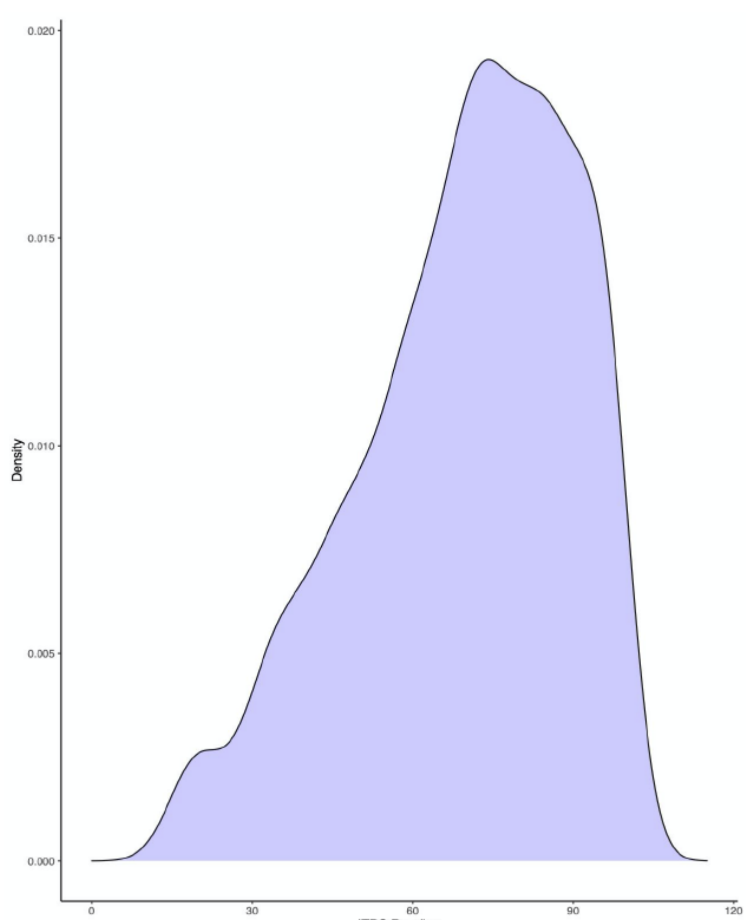


Figure 1.12

Density Curve Plot of Scores on the ITBS Math for Referrals Across First through Eighth Grade

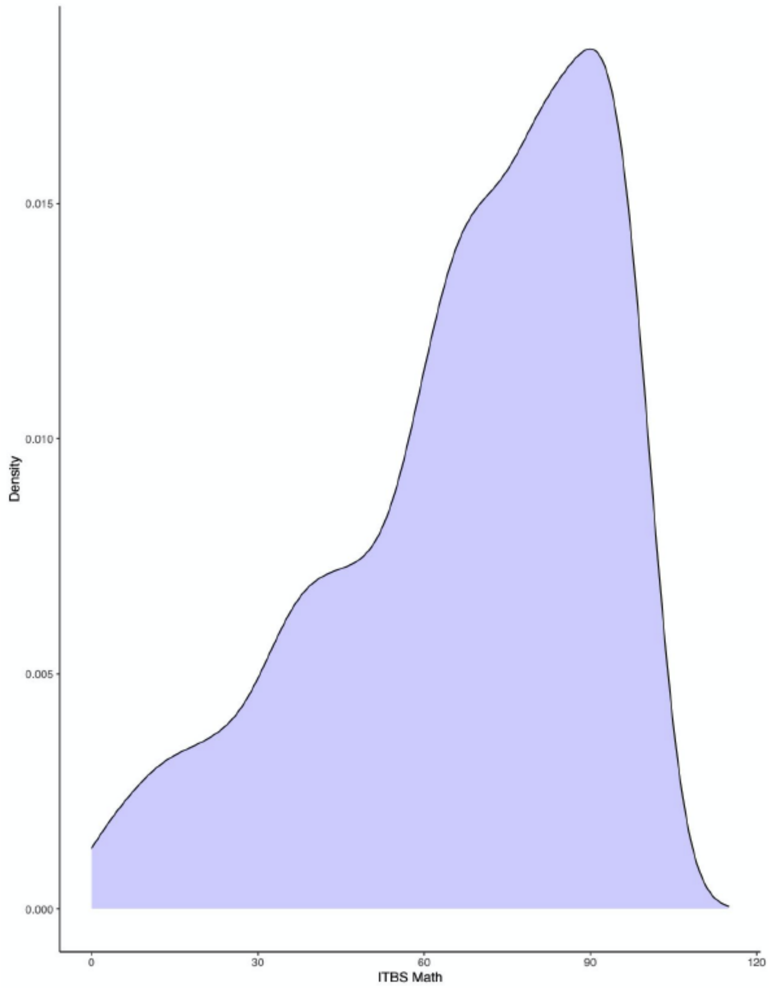
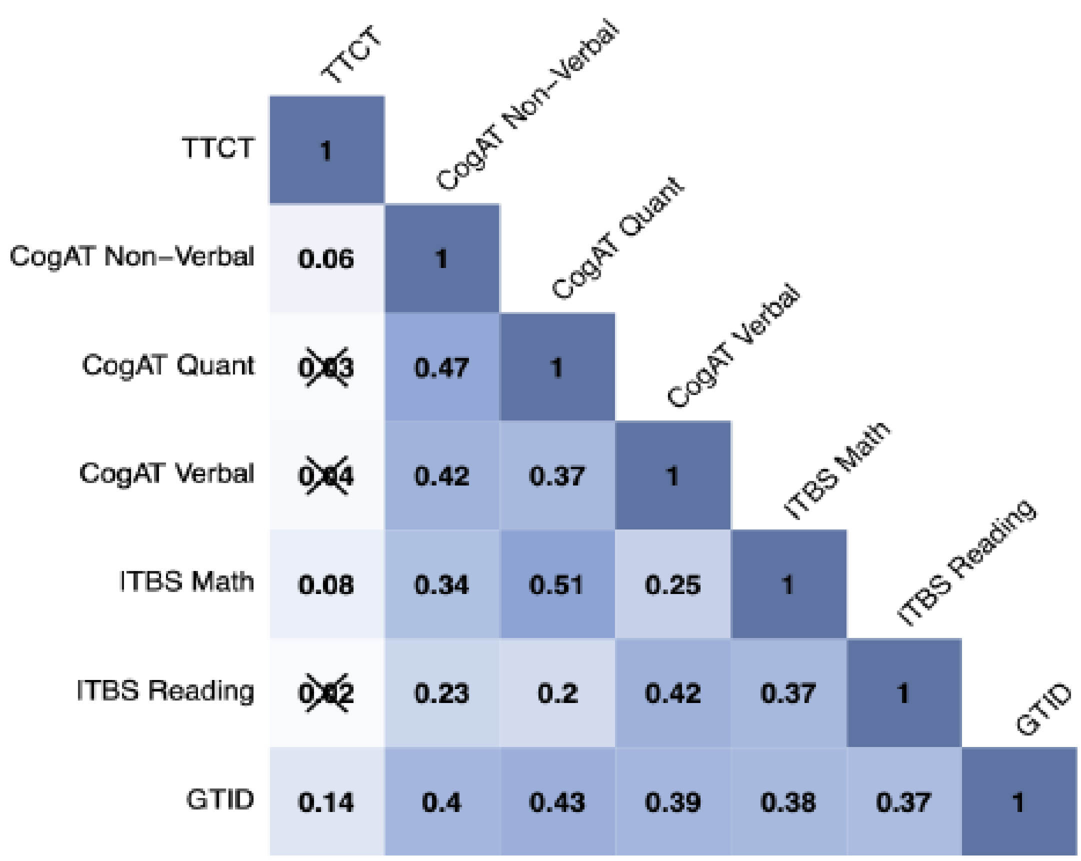


Figure 1.13

Correlation Matrix for Referrals



Note. Includes students referred after imputation. X indicates the correlation coefficient was $p > .05$.

Figure 1.14

Density Plot of Scores on the TTCT Form A or B for Universal Screened Students in First Grade

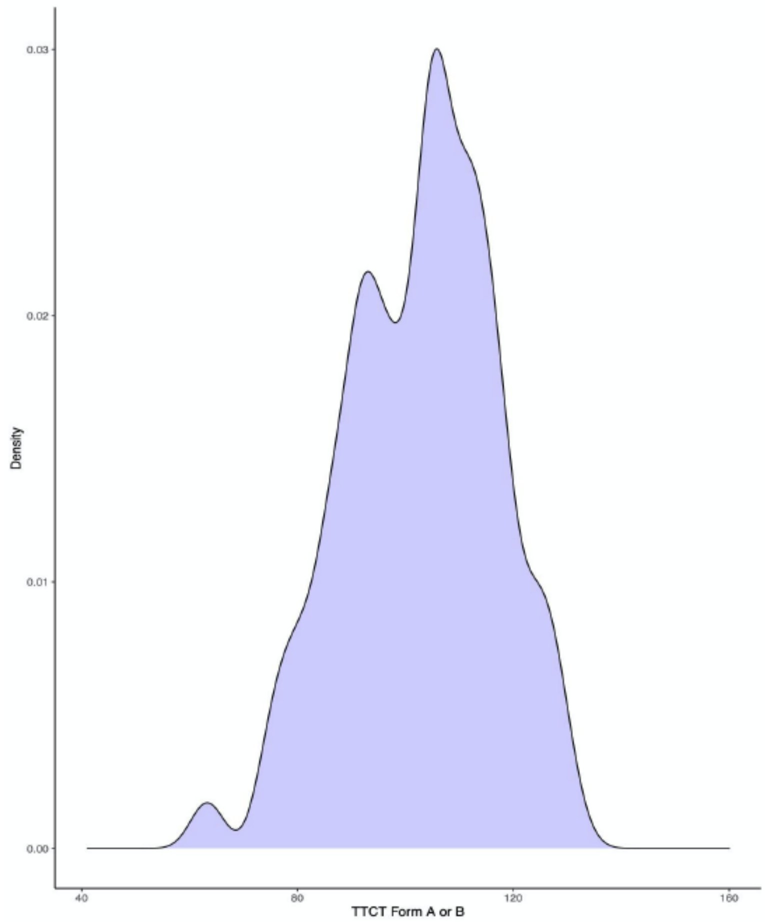


Figure 1.15

Density Curve Plot of Scores on the CogAT Non-Verbal for Universal Screened Students in First Grade

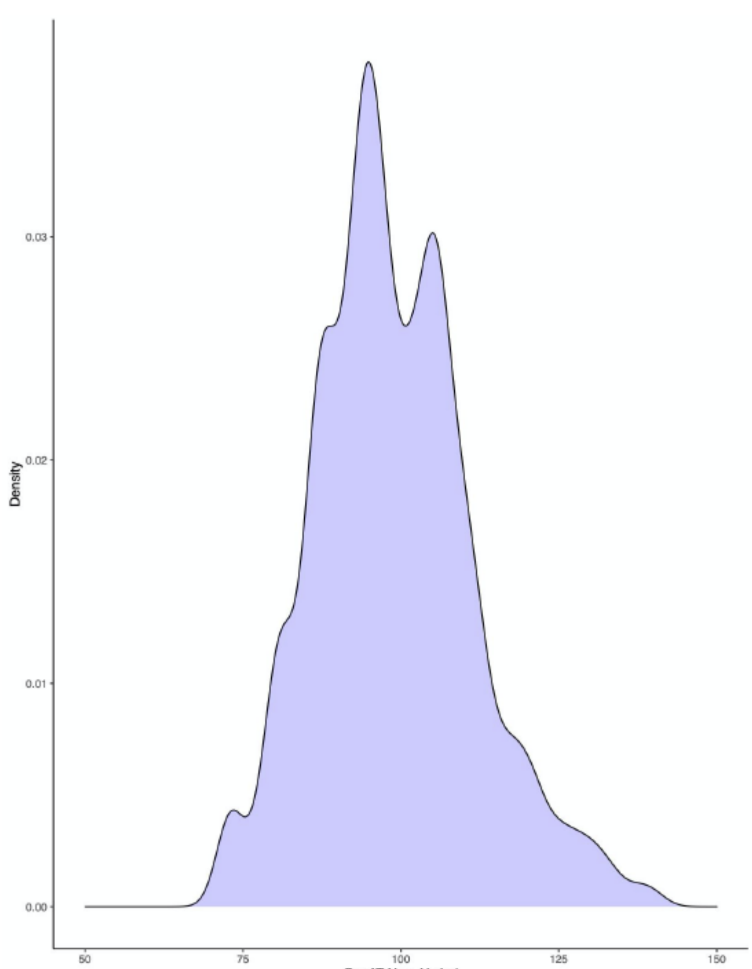


Figure 1.16

Density Curve Plot of Scores on the CogAT Quantitative for Universal Screened Students in First Grade

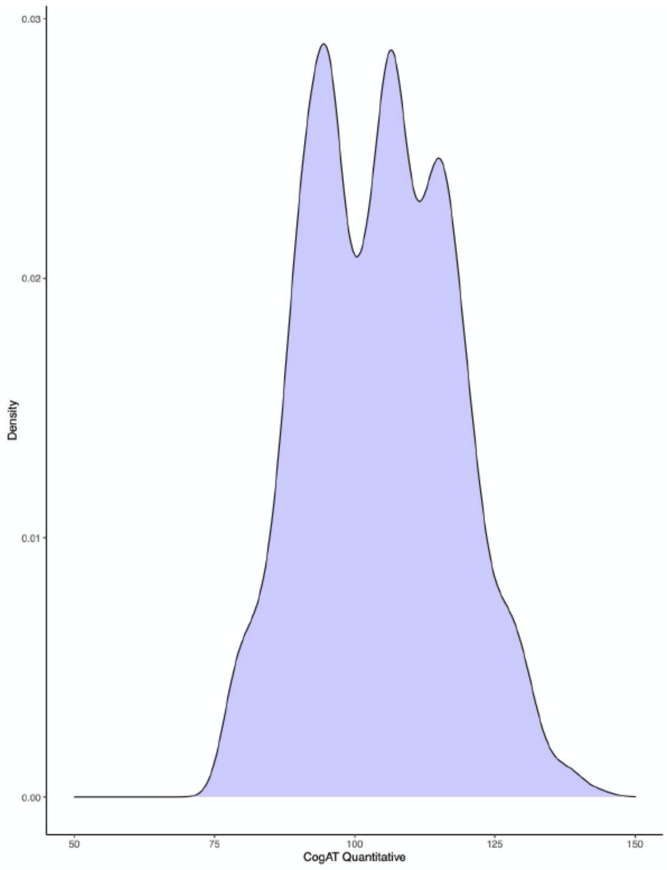


Figure 1.17

Density Curve Plot of Scores on the CogAT Verbal for Universal Screened Students in First Grade

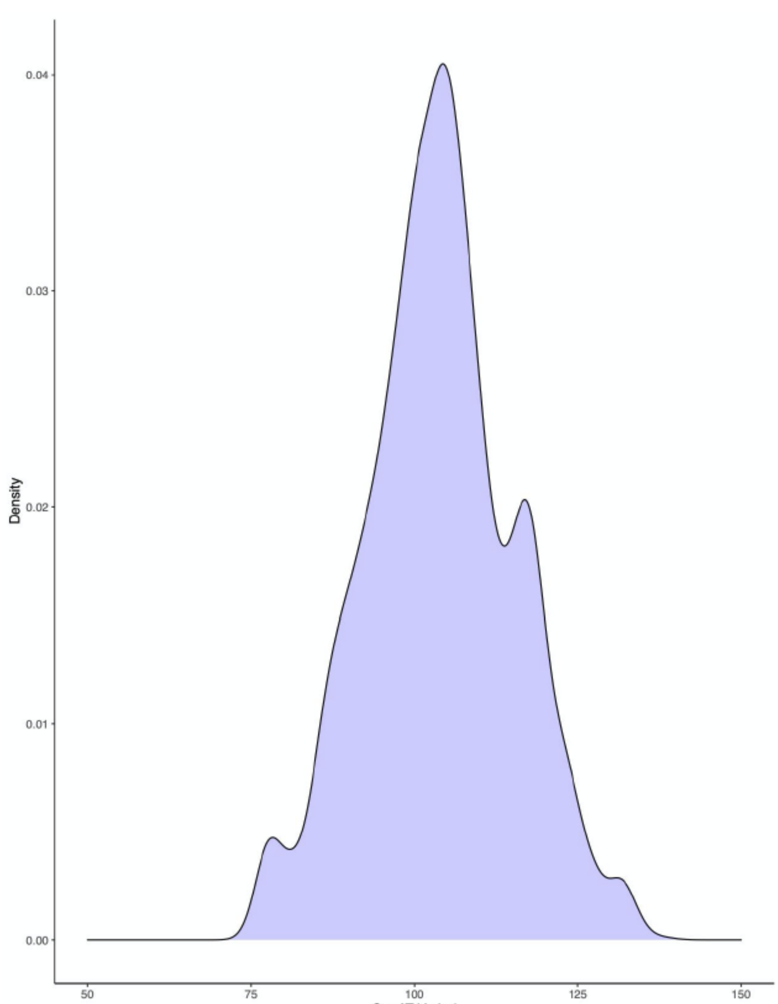


Figure 1.18

Density Curve Plot of Scores on the ITBS Reading for Universal Screened Students in First Grade

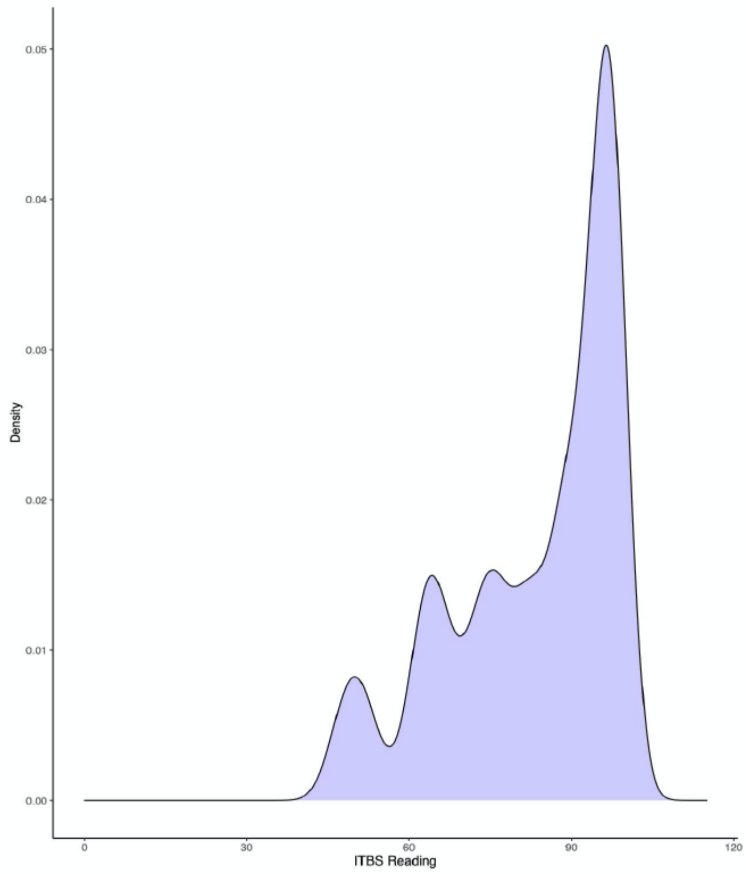


Figure 1.19

Density Curve Plot of Scores on the ITBS Math for Universal Screened Students in First Grade

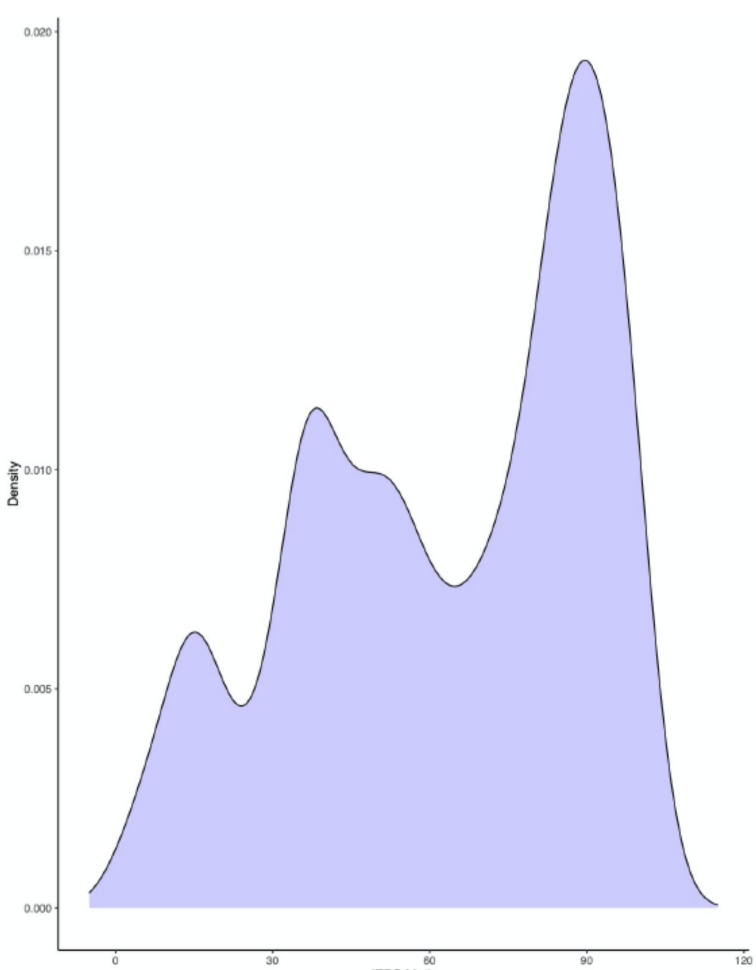


Figure 1.20

Density Curve Plot of Scores on the NNAT-2 for Universal Screened Students in First Grade

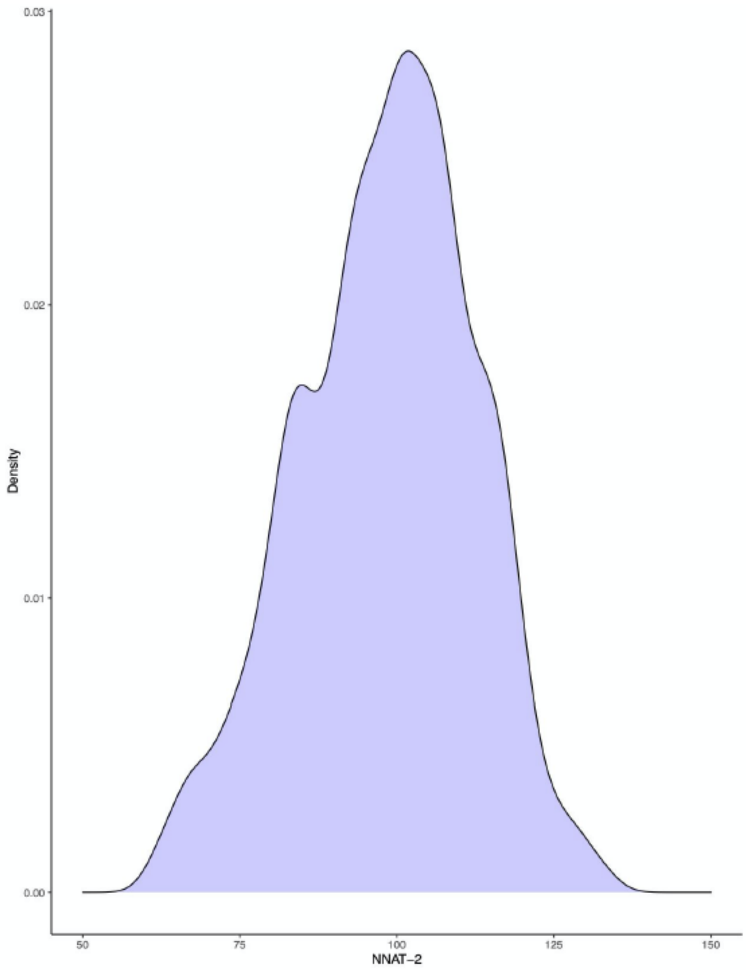
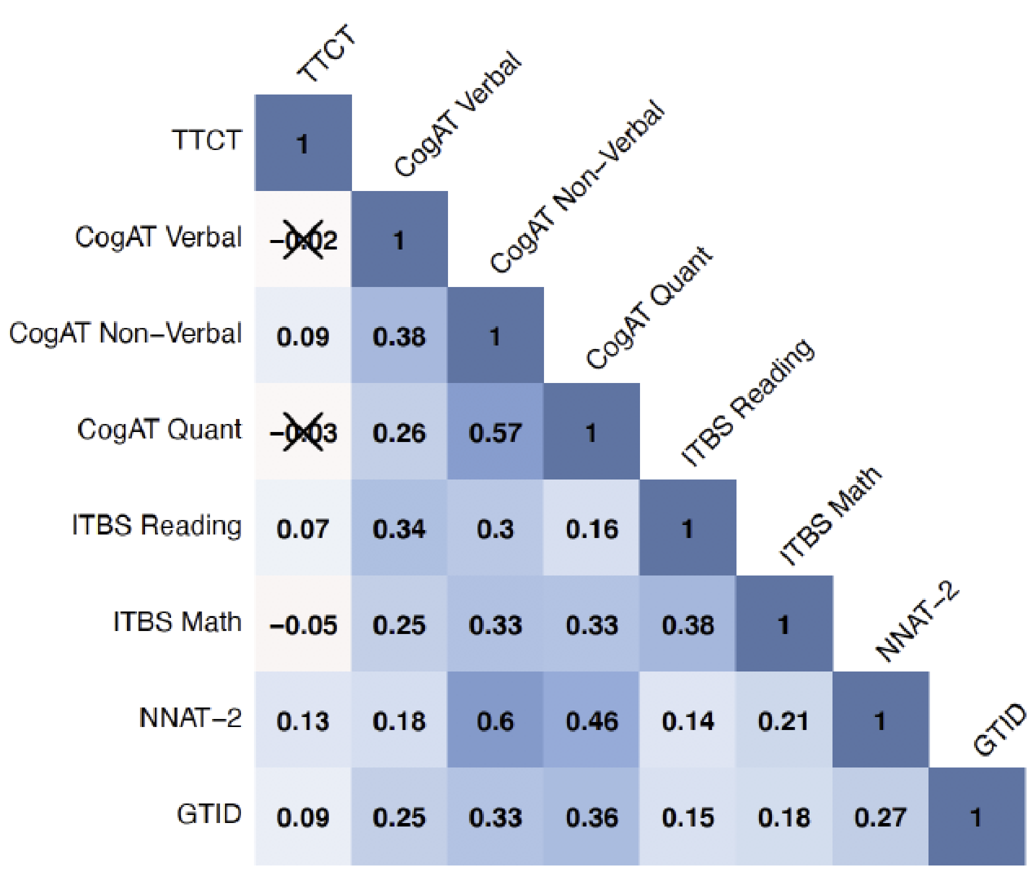


Figure 1.21

Correlation Matrix for First Grade Universal Screening



Note. Includes students referred after imputation. X indicates non-significant correlation coefficient.

Table 1.4

Multiple Regression for Referrals by Test and Race/Ethnicity

Variables	B	SE	β	r_s^2	95% CI	R^2	ΔR^2
TTCT Form A & B							
Constant	.06***	.04			[-.02, .14]	.016**	
Black/African American	-.13	.08	-.04	.047	[-.29, .04]		
Hispanic/Latinx	-.21	.07	-.09**	.392**	[-.36, -.07]		
Asian	.23	.13	.05	.296	[-.03, .49]		
Native American	-.45	.20	-.07	.227	[-.83, -.07]		
Other	.05	.18	.01	.023	[-.30, .40]		
CogAT7 Non-Verbal							
Constant	.08	.04			[-.00, .15]	.02***	.004
Black/African American	-.36	.08	-.13***	.703	[-.52, -.19]		
Hispanic/Latinx	-.11	.07	-.04	.013	[-.25, .04]		
Asian	.16	.13	.03	.149	[-.10, .42]		
Native American	-.35	.19	-.05	.087	[-.73, .03]		
Other	-.02	.18	-	.005	[-.36, .33]		
CogAT7 Quantitative							
Constant	.10*	.04			[.02, .17]	.034***	.018
Black/African American	-.31	.08	-.11*	.229	[-.47, -.14]		
Hispanic/Latinx	-.21	.07	-.09*	.110	[-.36, -.07]		
Asian	.36	.13	.08*	.338	[.10, .62]		
Native American	-.56	.19	-.08*	.148	[-.94, -.18]		
Other	-.45	.18	-.07*	.103	[-.79, -.10]		

(table continues)

Variables	B	SE	β	r_s^2	95% CI	R^2	ΔR^2
			CogAT7 Verbal				
Constant	.19***	.04			[.11, .26]	.056***	.04
Black/African American	-.20	.08	-.07	.001	[-.36, -.04]		
Hispanic/Latinx	-.59	.07	-.24***	.789	[-.74, -.45]		
Asian	-.25	.13	-.06	.004	[-.51, .00]		
Native American	-.37	.19	-.06	.014	[-.75, .003]		
Other	-.39	.17	-.06	.021	[-.73, -.05]		
			ITBS Reading				
Constant	.17***	.04			[.09, .24]	.038***	.022
Black/African American	-.44	.08	-.16***	.340	[-.60, -.27]		
Hispanic/Latinx	-.24	.07	-.10***	.043	[-.39, -.10]		
Asian	-.52	.13	-.12***	.182	[-.78, -.27]		
Native American	-.40	.19	-.06*	.035	[-.78, -.03]		
Other	-.49	.18	-.08**	.077	[-.83, -.14]		
			ITBS Math				
Constant	.06	.04			[-.02, .14]	.025***	.009
Black/African American	-.31	.08	-.11***	.436	[-.48, -.15]		
Hispanic/Latinx	-.06	.07	-.03	-	[-.21, .08]		
Asian	.35	.13	.08**	.358	[.09, .61]		
Native American	-.36	.19	-.05	.083	[-.74, .02]		
Other	-.40	.18	-.07	.131	[-.75, -.06]		

Note. CogAT7 = Cognitive Abilities Test. ITBS = Iowa Test of Basic Skills. TTCT = Torrance Test of Creative Thinking. Other category consists of two or more races and Native Hawaiian/Pacific Islander. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 1.5

Linear Regression for Referrals by English Learners and Test

Variables	B	SE	95% CI	R^2	ΔR^2
TTCT Form A & B					
Constant	.03	.03	[-.03, .10]	.007**	.007
English Learner	-.24**	.08	[-.40, -.08]		
CogAT Non-Verbal					
Constant	.01	.03	[-.05, .07]	.0007	-.0063
English Learner	-.08	.08	[-.24, .09]		
CogAT Quantitative					
Constant	.02	.03	[-.04, .08]	.002	-.005
English Learner	-.13	.08	[-.29, .03]		
CogAT Verbal					
Constant	.12	.03	[.06, .18]	.086***	.079
English Learner	-.83***	.08	[-.99, -.68]		
ITBS Reading					
Constant	.06	.03	[-.00, .12]	.019***	.012
English Learner	-.39***	.08	[-.55, -.23]		
ITBS Math					
Constant	-.01	.03	[-.08, .05]	.001	-.006
English Learner	.10	.08	[-.06, .26]		

Note. CogAT7 = Cognitive Abilities Test. ITBS = Iowa Test of Basic Skills. TTCT = Torrance Test of Creative Thinking. **p* < .05. ***p* < .01. ****p* < .001.

Table 1.6

Linear Regression for Referrals by Sex and Test

Variables	B	SE	95% CI	R^2	ΔR^2
TTCT Form A & B					
Constant	-.12**	.04	[-.20, -.05]	.016***	.016
Females	.26***	.06	[.14, .37]		
CogAT Non-Verbal					
Constant	.02	.04	[-.06, .10]	.0002	-.016
Females	-.03	.06	[-.15, .08]		
CogAT Quantitative					
Constant	.14**	.04	[.06, .21]	.019***	.003
Females	-.28***	.06	[-.39, -.17]		

(table continues)

Variables	B	SE	95% CI	R^2	ΔR^2
CogAT Verbal					
Constant	-.06	.04	[-.14, .02]	.004*	-.012
Females	.13*	.06	[.02, .24]		
ITBS Reading					
Constant	-.07	.04	[-.15, .01]	.005*	-.011
Females	.14*	.06	[.03, .26]		
ITBS Math					
Constant	.12**	.04	[.04, .19]	.014***	-.002
Females	-.24***	.06	[-.35, -.12]		

Note. CogAT7 = Cognitive Abilities Test. ITBS = Iowa Test of Basic Skills. TTCT = Torrance Test of Creative Thinking. **p* < .05. ***p* < .01. ****p* < .001.

Table 1.7

Linear Regression for Referrals by Free/Reduced Lunch and Test

Variables	B	SE	95% CI	R^2	ΔR^2
TTCT Form A & B					
Constant	.05	.04	[-.02, .12]	.005*	.005
Free/Reduced Lunch	-.15*	.06	[-.27, -.03]		
CogAT Non-Verbal					
Constant	.07	.04	[-.001, .14]	.009**	.004
Free/Reduced Lunch	-.20**	.06	[-.32, -.08]		
CogAT Quantitative					
Constant	.12**	.04	[.05, .18]	.025***	.02
Free/Reduced Lunch	-.34***	.06	[-.45, -.22]		
CogAT Verbal					
Constant	.17***	.03	[.10, .23]	.052***	.047
Free/Reduced Lunch	-.48***	.06	[-.60, -.37]		
ITBS Reading					
Constant	.11***	.04	[.04, .17]	.021***	.016
Free/Reduced Lunch	-.31***	.06	[-.43, -.19]		
ITBS Math					
Constant	.06	.04	[-.01, .13]	.006**	.001
Free/Reduced Lunch	-.16*	.06	[-.28, -.04]		

Note. CogAT7 = Cognitive Abilities Test. ITBS = Iowa Test of Basic Skills. TTCT = Torrance Test of Creative Thinking. **p* < .05. ***p* < .01. ****p* < .001.

Table 1.8

Multiple Regression for Students Universally Screened in First Grade by Test and Race/Ethnicity

Variables	B	SE	β	r_s^2	95% CI	R^2	ΔR^2
TTCT Form A & B							
Constant	.03	.03			[-.03, .10]	.026***	.026
Black/African American	.24***	.06	.09***	.561	[.12, .36]		
Hispanic/Latinx	-.20***	.05	-.09***	.404	[-.30, -.10]		
Asian	-.42***	.11	-.08***	.220	[-.65, -.20]		
Native American	-.12	.16	-.02	.005	[-.42, .19]		
Other	.20	.27	.02	.013	[-.32, .72]		
CogAT7 Non-Verbal							
Constant	.22***	.03			[.16, .28]	.062***	.036
Black/African American	-.47***	.06	-.18***	.213	[-.59, -.35]		
Hispanic/Latinx	-.47***	.05	-.21***	.428	[-.57, -.37]		
Asian	-.11	.11	-.02	.007	[-.33, .11]		
Native American	-.05	.15	-.01	.009	[-.35, .25]		
Other	.94**	.26	.08**	.147	[.43, 1.45]		
CogAT7 Quantitative							
Constant	0.18***	.03			[.12, .24]	.059***	.033
Black/African American	-.40***	.06	-0.15***	.172	[-.52, -.28]		
Hispanic/Latinx	-.35***	.05	-0.16***	.197	[-.45, -.25]		
Asian	.37***	.11	0.07***	.211	[.15, .59]		
Native American	-1.03***	.15	-0.14***	.250	[-1.33, -.73]		
Other	-.51	.26	-.04	.012	[-1.02, .00]		
CogAT7 Verbal							
Constant	.12***	.03			[.06, .18]	.061***	.035
Black/African American	-.05***	.06	-.02***	.019	[-.16, .07]		

(table continues)

Variables	B	SE	β	r_s^2	95% CI	R^2	ΔR^2
Hispanic/Latinx	-.41***	.05	-.19***	.559	[-.51, -.31]		
Asian	-.41**	.11	-.08**	.057	[-.63, -.19]		
Native American	.92***	.15	.13***	.365	[.62, 1.22]		
Other	.65	.26	.05	.066	[.14, 1.16]		
ITBS Reading							
Constant	-.05	.03			[-.11, .01]	.032***	.006
Black/African American	.19**	.06	.07**	.134	[.07, .31]		
Hispanic/Latinx	-.01	.05	.00	.043	[-.11, .09]		
Asian	.24	.11	.05	.045	[.01, .46]		
Native American	.84***	.16	.12***	.406	[.53, 1.14]		
Other	-1.30***	.27	-.11***	.387	[-1.82, -.78]		
ITBS Math							
Constant	.13**	.03			[.07, .19]	.037***	.011
Black/African American	-.32***	.06	-.12***	.201	[-.44, -.20]		
Hispanic/Latinx	-.31***	.05	-.14***	.374	[-.41, -.21]		
Asian	.07	.11	.01	.045	[-.15, .30]		
Native American	.50**	.15	.07**	.220	[.19, .80]		
Other	.61*	.26	.05*	.100	[.09, 1.13]		
NNAT-2							
Constant	.22***	.03			[.16, .28]	.075***	.049
Black/African American	-.66***	.06	-.25***	.539	[-.78, -.54]		
Hispanic/Latinx	-.37***	.05	-.17***	.130	[-.47, -.28]		
Asian	.30*	.11	.06*	.150	[.09, .52]		
Native American	-.40*	.15	-.06*	.009	[-.70, -.10]		
Other	-.25	.26	-.02	-	[-.76, .26]		

Note. CogAT7 = Cognitive Abilities Test. ITBS = Iowa Test of Basic Skills. TTCT = Torrance Test of Creative Thinking. Other category consists of two or more races and Native Hawaiian/Pacific Islander. **p* < .05. ***p* < .01. ****p* < .001.

Table 1.9

Linear Regression for Students Universally Screened in First Grade by English Learners and Test

Variables	B	SE	95% CI	R^2	ΔR^2
TTCT Form A & B					
Constant	.08**	.02	[.03, .13]	.033***	.033
English Learner	-.50***	.06	[-.62, -.38]		
CogAT Non-Verbal					
Constant	.06*	.02	[.01, .10]	.018***	-.015
English Learner	-.37***	.06	[-.48, -.25]		
CogAT Quantitative					
Constant	.04	.02	[-.01, .08]	.008***	-.025
English Learner	-.24***	.06	[-.36, -.12]		
CogAT Verbal					
Constant	.01	.02	[-.03, .06]	.01	-.023
English Learner	-.08	.06	[-.20, .03]		
ITBS Reading					
Constant	-.003	.02	[-.05, .04]	.00006	-.032
English Learner	.02	.06	[-.10, .14]		
ITBS Math					
Constant	-.07**	.02	[-.11, -.02]	.023***	-.01
English Learner	.41***	.06	[.30, .53]		
NNAT-2					
Constant	.04	.02	[-.01, .08]	.007***	-.026
English Learner	-.23***	.06	[-.35, -.11]		

Note. CogAT7 = Cognitive Abilities Test. ITBS = Iowa Test of Basic Skills. TTCT = Torrance Test of Creative Thinking. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 1.10

Simple Linear Regression for Students Universally Screened in First Grade by Sex and Test

Variables	B	SE	95% CI	R^2	ΔR^2
TTCT Form A & B					
Constant	-.11**	.03	[-.17, -.05]	.012***	.012
Female	.22***	.04	[.14, .31]		
CogAT Non-Verbal					
Constant	-.16***	.03	[-.22, -0.11]	.029***	.017
Female	.34***	.04	[.25, 0.42]		
CogAT Quantitative					
Constant	-	.03	[-.06, .06]	.00001	-.012
Female	.01	.04	[-.08, .09]		
CogAT Verbal					
Constant	-.28***	.03	[-.34, -.22]	.082***	.07
Female	.57***	.04	[.49, .65]		
ITBS Reading					
Constant	-.20***	.03	[-.26, -.14]	.042***	.03
Female	.41***	.04	[.33, .50]		
ITBS Math					
Constant	.03	.03	[-.03, .09]	.001	-.011
Female	-.07	.04	[-.15, .02]		
NNAT-2					
Constant	-.03	.03	[-.09, .03]	.001	-.011
Female	.06	.04	[-.02, .15]		

Note. CogAT7 = Cognitive Abilities Test. ITBS = Iowa Test of Basic Skills. TTCT = Torrance Test of Creative Thinking. *p < .05. **p < .01. ***p < .001.

Table 1.11

Simple Linear Regression for Students Universally Screened in First Grade by Free/Reduced Lunch Status and Test

Variables	B	SE	95% CI	R^2	ΔR^2
TTCT Form A & B					
Constant	-.02	.03	[-.07, .04]	.001	.001
Free/Reduced Lunch	.05	.05	[-.04, .14]		
CogAT Non-Verbal					
Constant	.07**	.03	[.02, .13]	.008***	.007
Free/Reduced Lunch	-.19***	.04	[-.28, -.10]		
CogAT Quantitative					
Constant	.17**	.03	[.11, .22]	.044***	.043
Free/Reduced Lunch	-.43***	.04	[-.51, -.34]		
CogAT Verbal					
Constant	.04**	.03	[-.01, .10]	.003	.002
Free/Reduced Lunch	-.11***	.05	[-.20, -.02]		
ITBS Reading					
Constant	-.09**	.03	[-.15, -.04]	.014***	.013
Free/Reduced Lunch	.24***	.04	[.15, .33]		
ITBS Math					
Constant	.02	.03	[-.03, .08]	.001	.000
Free/Reduced Lunch	-.06	.05	[-.15, .03]		
NNAT-2					
Constant	.20***	.03	[.15, .26]	.064***	.063
Free/Reduced Lunch	-.52***	.04	[-.61, -.44]		

Note. CogAT7 = Cognitive Abilities Test. ITBS = Iowa Test of Basic Skills. TTCT = Torrance Test of Creative Thinking. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 1.12

Hierarchical Regression Student Demographics and Identification Measures to the Torrance Test of Creative Thinking by Referrals

Variables	TTCT Form A					TTCT Form B					TTCT Form A or B				
	B	SE	β	95% CI	r_s^2	B	SE	β	95% CI	r_s^2	B	SE	β	95% CI	r_s^2
Constant	104.03***	.72		[102.62, 105.44]		102.52***	.72		[101.11, 103.92]		104.39***	.72		[102.97, 105.81]	
Black/African American	-1.44	1.23	-.04	[-3.85, .97]	.012	-1.59	1.23	-.04	[-4.00, .81]	.034	-1.83	1.23	-.05	[-4.25, .59]	.021
Hispanic/Latinx	-1.57	1.18	-.05	[-3.89, .74]	.194	-.83	1.18	-.02	[-3.14, 1.48]	.113	-1.68	1.19	-.05	[-4.01, .64]	.173
Asian	4.55	1.95	.07	[0.74, 8.37]	.154	4.40	1.94	.07	[.60, 8.21]	.133	4.27	1.95	.07	[.44, 8.10]	.131
Native American	-4.83	2.82	-.05	[-10.36, .70]	.100	2.20	2.82	.02	[-3.32, 7.72]	.009	-5.24	2.83	-.05	[-10.80, .31]	.101
Other	1.82	2.53	.02	[-3.13, 6.78]	.022	4.30	2.52	.05	[-.64, 9.25]	.084	1.13	2.54	.01	[-3.85, 6.10]	.010
Female	3.22**	.82	.11**	[1.61, 4.84]	.418	3.80***	.82	.13***	[2.19, 5.41]	.569	3.58**	.82	.12**	[1.97, 5.20]	.461
Free/Reduced Lunch	-.46	.96	-.02	[-2.35, 1.42]	.130	-.03	.96	-.001	[-1.91, 1.85]	.043	-.58	.96	-.02	[-2.47, 1.31]	.139
English Language Learner	-2.62	1.40	-.06	[-5.37, .13]	.224	-2.23	1.40	-.06	[-4.98, .51]	.104	-2.50	1.41	-.06	[-5.26, .26]	.200
R^2	.032***					.031***					.035***				
Constant	106.37***	5.44		[95.71, 117.02]		97.23***	5.41		[86.63, 107.84]		103.95***	5.45		[93.26, 114.64]	
Black/African American	-1.33	1.24	-.03	[-3.76, 1.11]	.009	-1.13	1.24	-.03	[-3.55, 1.29]	.023	-1.59	1.25	-.04	[-4.03, .85]	.016
Hispanic/Latino	-1.63	1.18	-.05	[-3.95, .69]	.148	-.72	1.18	-.02	[-3.03, 1.59]	.075	-1.74	1.19	-.05	[-4.07, .59]	.132
Asian	3.95	1.96	.06	[.10, 7.80]	.118	3.77	1.95	.06	[-.06, 7.60]	.088	3.62	1.97	.06	[-.24, 7.48]	.100
Native American	-4.62	2.82	-.05	[-10.15, .91]	.076	2.87	2.81	.03	[-2.63, 8.37]	.006	-4.90	2.83	-.05	[-10.45, .65]	.076
Other	1.77	2.54	.02	[-3.20, 6.74]	.017	4.97	2.52	.06	[.02, 9.91]	.056	1.13	2.55	.01	[-3.86, 6.12]	.008
Female	3.53**	.84	.12**	[1.89, 5.18]	.319	4.27***	.84	.15***	[2.63, 5.91]	.378	3.93***	.84	.14***	[2.28, 5.58]	.350
Free/Reduced Lunch	-.50	.97	-.02	[-2.40, 1.40]	.099	.25	.96	.01	[-1.63, 2.14]	.029	-.56	.97	-.02	[-2.46, 1.35]	.106
English Language Learner	-3.17	1.43	-.08	[-5.98, -.36]	.171	-2.60	1.43	-.06	[-5.39, .20]	.069	-3.08	1.44	-.08	[-5.90, -.27]	.152
CogAT Verbal	-.02	.04	-.02	[-.11, .07]	.024	-.01	.04	-.01	[-.10, .08]	.060	-.03	.04	-.02	[-.12, .06]	.032
CogAT Non-verbal	.04	.04	.03	[-.04, .12]	.041	.01	.04	.01	[-.08, .09]	.056	.06	.04	.05	[-.02, .14]	.078

(table continues)

Variables	TTCT Form A					TTCT Form B					TTCT Form A or B				
	B	SE	β	95% CI	r_s^2	B	SE	β	95% CI	r_s^2	B	SE	β	95% CI	r_s^2
CogAT Quantitative	-.06	.05	-.05	[-.15, .03]	.003	.004	.05	.003	[-.09, .09]	.060	-.05	.05	-.04	[-.14, .04]	.014
ITBS Reading	-.03	.02	-.05	[-.08, .01]	.002	-.002	.02	-.003	[-.05, .05]	.067	-.03	.02	-.04	[-.08, .02]	.010
ITBS Math	.07*	.02	.11*	[.03, .11]	.125	.07**	.02	.13**	[.03, .11]	.266	.07*	.02	.11*	[0.03, .11]	.145
R^2	.042***					.046***					.046***				
ΔR^2	.010					.016					.011				

Note. CogAT7 = Cognitive Abilities Test. ITBS = Iowa Test of Basic Skills. TTCT = Torrance Test of Creative Thinking. **p* < .05. ***p* < .01. ****p* < .001.

Table 1.13

Hierarchical Regression Student Demographics and Identification Measures to the Torrance Test of Creative Thinking by Universal Screening

Variables	TTCT Form A					TTCT Form B					TTCT Form A or B				
	B	SE	β	95% CI	r_s^2	B	SE	β	95% CI	r_s^2	B	SE	β	95% CI	r_s^2
Constant	101.23***	.53		[100.19, 102.26]		109.05***	.33		[108.41, 109.69]		101.38***	.54		[100.32, 102.44]	
Black/African American	.39	.85	.01	[-1.28, 2.05]	.092	0.93	.53	.04	[-0.10, 1.97]	.032	2.60	.87	0.07*	[.89, 4.30]	.238
Hispanic/Latinx	-1.19	.79	-.04	[-2.74, 0.37]	.145	-3.96	0.49	-.21***	[-4.93, -3.00]	.254	-1.05	.81	-.03	[-2.63, .54]	.171
Asian	-3.70	1.56	-.05	[-6.76, -0.64]	.095	-1.24	.97	-.03	[-3.14, 0.66]	.002	-3.74	1.60	-.05	[-6.87, -.60]	.093
Native American	-3.77	2.11	-.04	[-7.90, 0.37]	.036	1.44	1.31	.02	[-1.13, 4.01]	.021	-.20	2.16	-.002	[-4.43, 4.03]	.002
Other	2.61	3.58	.02	[-4.40, 9.63]	.007	7.31	2.22	.07**	[2.96, 11.66]	.050	3.16	3.66	.02	[-4.02, 10.34]	.006
Female	2.57***	.58	.09***	[1.43, 3.72]	.169	-4.28***	.36	-.24***	[-4.98, -3.57]	.510	3.06***	.60	.11***	[1.89, 4.23]	.189
Free/Reduced Lunch	2.83***	.69	.10**	[1.49, 4.18]	.006	-1.76***	.43	-.10**	[-2.59, -0.92]	.063	2.71**	.70	.09**	[1.34, 4.08]	.009
English Language Learner	-7.15***	.96	-.19***	[-9.04, -5.26]	.576	3.99***	.60	.17***	[2.81, 5.16]	.008	-7.24***	.99	-.19***	[-9.18, -5.31]	.507
R^2	.056					.117					.066				
Constant	90.57***	3.60		[83.51, 97.62]		72.64***	1.57		[69.57, 75.71]		98.86***	3.72		[91.56, 106.16]	
Black/African American	2.00	.85	.06	[.34, 3.66]	.042	3.85***	.37	.17***	[3.13, 4.57]	.006	3.67**	.88	.10**	[1.95, 5.38]	.142
Hispanic/Latino	-1.13	.80	-.04	[-2.69, .43]	.067	-1.81***	.35	-.09***	[-2.48, -1.13]	.050	-1.42	.82	-.05	[-3.04, .19]	.102

(table continues)

Variables	TTCT Form A					TTCT Form B					TTCT Form A or B				
	B	SE	β	95% CI	r_s^2	B	SE	β	95% CI	r_s^2	B	SE	β	95% CI	r_s^2
Asian	-5.19**	1.54	-.07**	[-8.21, -2.17]	.044	.35	.67	.01	[-0.96, 1.66]	-	-5.17**	1.59	-.07**	[-8.29, -2.05]	.056
Native American	-3.24	2.10	-.03	[-7.36, .89]	.016	1.83	.92	.03	[0.03, 3.62]	.004	.30	2.18	.003	[-3.97, 4.57]	.001
Other	3.81	3.56	.02	[-3.16, 10.78]	.003	-.97	1.55	-.01	[-4.00, 2.06]	.010	5.00	3.68	.03	[-2.21, 12.21]	.003
Female	2.63**	.61	.10**	[1.44, 3.83]	.078	-4.82***	.27	-.28***	[-5.34, -4.30]	.100	3.32***	.63	.12***	[2.08, 4.56]	.113
Free/Reduced Lunch	3.46***	.70	.12***	[2.10, 4.83]	.003	1.33***	.30	.07***	[0.74, 1.93]	.012	3.14**	.72	.11**	[1.73, 4.55]	.005
English Language Learner	-6.54***	.98	-.18***	[-8.46, -4.62]	.266	2.65***	.43	.11***	[1.82, 3.49]	.001	-6.57***	1.01	-.17***	[-8.56, -4.58]	.303
CogAT Verbal	-.13**	.03	-.10**	[-.19, -.07]	.002	.30	.01	.39	[0.28, 0.33]	.095	-.13**	.03	-.11**	[-.19, -.07]	.005
CogAT Non-verbal	.05	.03	.05	[-.01, .12]	.184	.01	.02	.01	[-0.02, 0.04]	.074	.01	.04	.01	[-.06, .08]	.069
CogAT Quantitative	-.10**	.03	-.09**	[-.15, -.04]	.001	-.11***	.01	-.16***	[-0.13, -0.09]	.021	-.10**	.03	-.09**	[-.16, -.04]	.010
ITBS Reading	.05	.02	.05	[.004, .09]	.050	-.28***	.01	-.48***	[-0.30, -0.26]	.159	.05	.02	.06	[.01, .10]	.040
ITBS Math	-.005	.01	-.01	[-.03, .02]	.002	.01	.01	.04	[.0008, 0.02]	.004	-.01	.01	-.03	[-.04, .01]	.021
	.25***	.03	0.25***	[.20, .30]	.332	.37***	.01	.59***	[0.35, 0.39]	.371	.22***	.03	.21***	[.16, .27]	.158
R^2	.120					.596					.110				
ΔR^2	.064					.479					.044				

Note. CogAT7 = Cognitive Abilities Test. ITBS = Iowa Test of Basic Skills. TTCT = Torrance Test of Creative Thinking. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 1.14

Representation Indices of Students Identified in the 2018-2019 School Year

Variables	Total Identified	Total Population	RI 2018-2019
Black/African American	10%	17%	.61
Hispanic/Latinx	18%	31%	.58
Asian	9%	3%	2.65
Native American	.90%	.6%	1.50
White	62%	47%	1.33
Native Hawaiian/Pacific Islander	.2%	.2%	1.00
Two or More Races	.2%	1.40%	.14
Free/Reduced Lunch	24%	46%	.53
English Learner	8%	15%	.54

Note. The identified column are proportions. Female and male total proportion of the population was not available in the Texas Performance Accountability Report, so total proportion was calculated based the total population of females (48.6%) and males (51.4%) for 2017-2018 reported in the Office of Civil Rights (OCR) database. 49% of males were identified, 45% of females were identified, where females are .93 as frequently represented and males are .95 represented. This is only the representation indices of students who were screened or referred in the 2018-2019 school year, not the entire student population identified.

Table 1.15

Combination Rules For All Students Tested for Gifted Services in 2018-2019

Variable	Hypothetical Combination Rules			Actual Identification	
	AVERAGE	OR	AND	Universal Screening	Referrals
Black/African American	12.00% (15)	11.32% (133)	-	7.81% (10)	10.62% (29)
Hispanic/Latino	7.56% (9)	24.17% (284)	-	19.53% (25)	16.84% (46)
Asian	6.72% (8)	6.21% (73)	-	9.38% (12)	9.16% (25)
Native American	-	2.04% (24)	-	-	1.46% (4)
White	73.11% (87)	55.23% (649)	100.00% (1)	63.28 (81)	61.17% (167)
Other[a]	-	1.02% (12)	-	-	.07% (2)
Female	44.54% (53)	49.02% (576)	100.00% (1)	46.88% (60)	44.32% (121)
Male	55.46% (66)	50.98% (599)	-	53.12% (68)	55.68 % (152)
Free/Reduced Lunch	15.13% (18)	28.34% (333)	-	15.63% (20)	27.84% (76)
English Language Learner	3.36% (4)	13.45% (158)	-	7.81% (10)	8.42% (23)
Total Identified[b]	3.65% (119)	36.03% (1175)	0.003% (1)	3.93% (128)	8.37% (273)

Note. For the AND and OR rules, the cut scores by the district were used. For the AVERAGE, the 95th percentile was used as the cut-off. a = category comprised of Native Hawaiian and Pacific Islanders; b = percentages were calculated with the total population that was tested for gifted services (n = 3261).

Figure 1.22

Generalized Linear Regression Plot for TTCT-Figural and the Probability of Being Identified for Gifted Services by Referrals

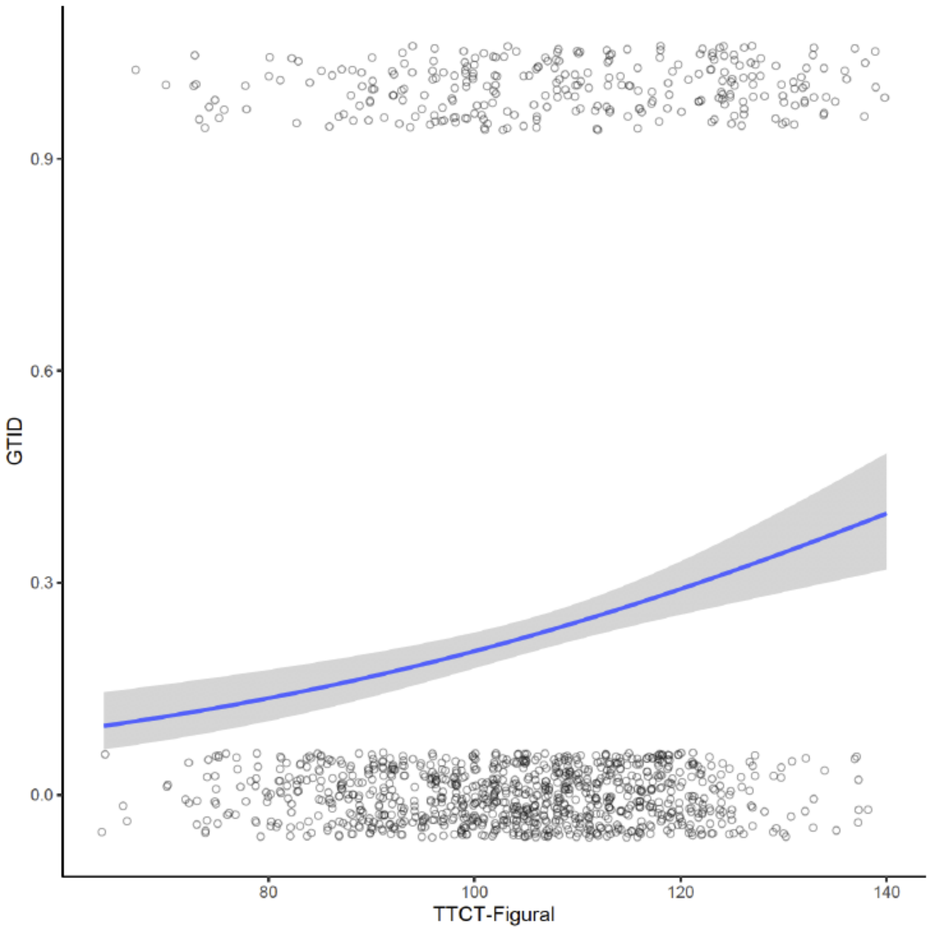


Figure 1.23

Generalized Linear Regression Plot for CogAT Non-Verbal and the Probability of Being Identified for Gifted Services by Referrals

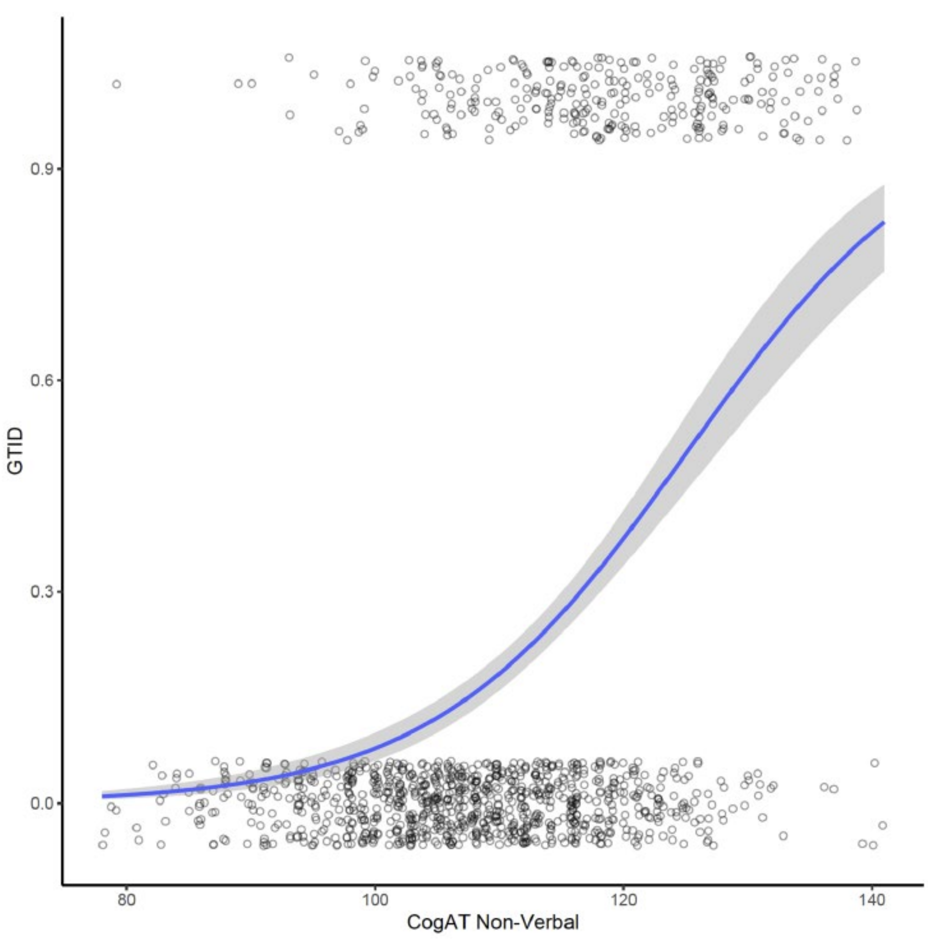


Figure 1.24

Generalized Linear Regression Plot for CogAT Quantitative and the Probability of Being Identified for Gifted Services by Referrals

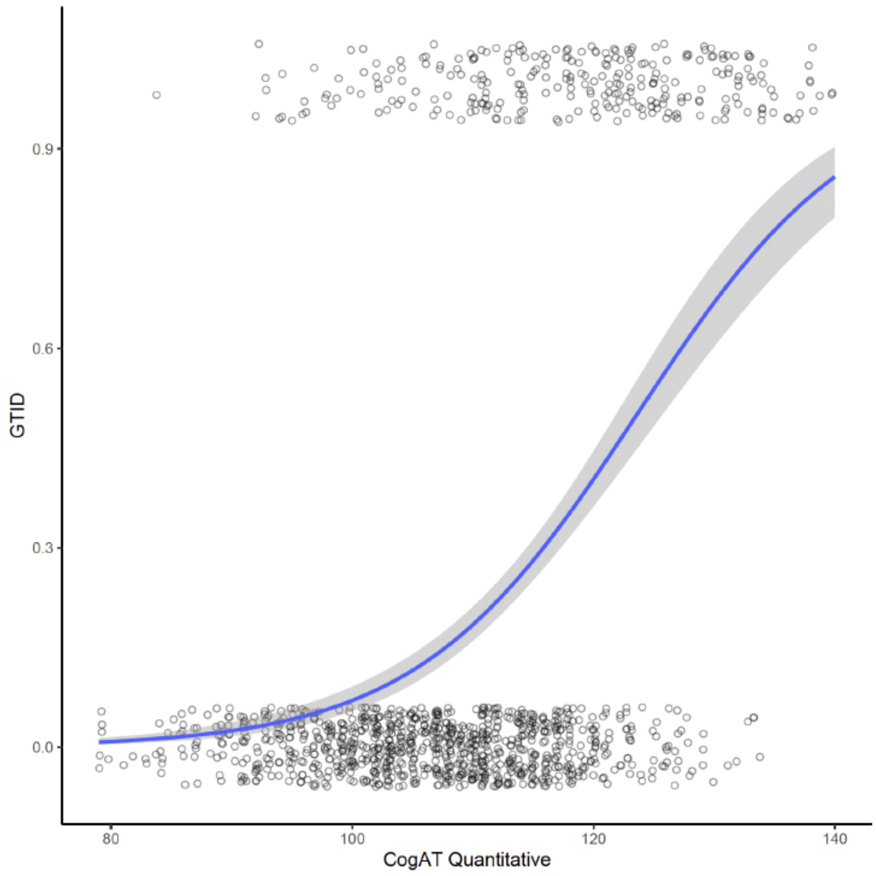


Figure 1.25

Generalized Linear Regression Plot for CogAT Verbal and the Probability of Being Identified for Gifted Services by Referrals

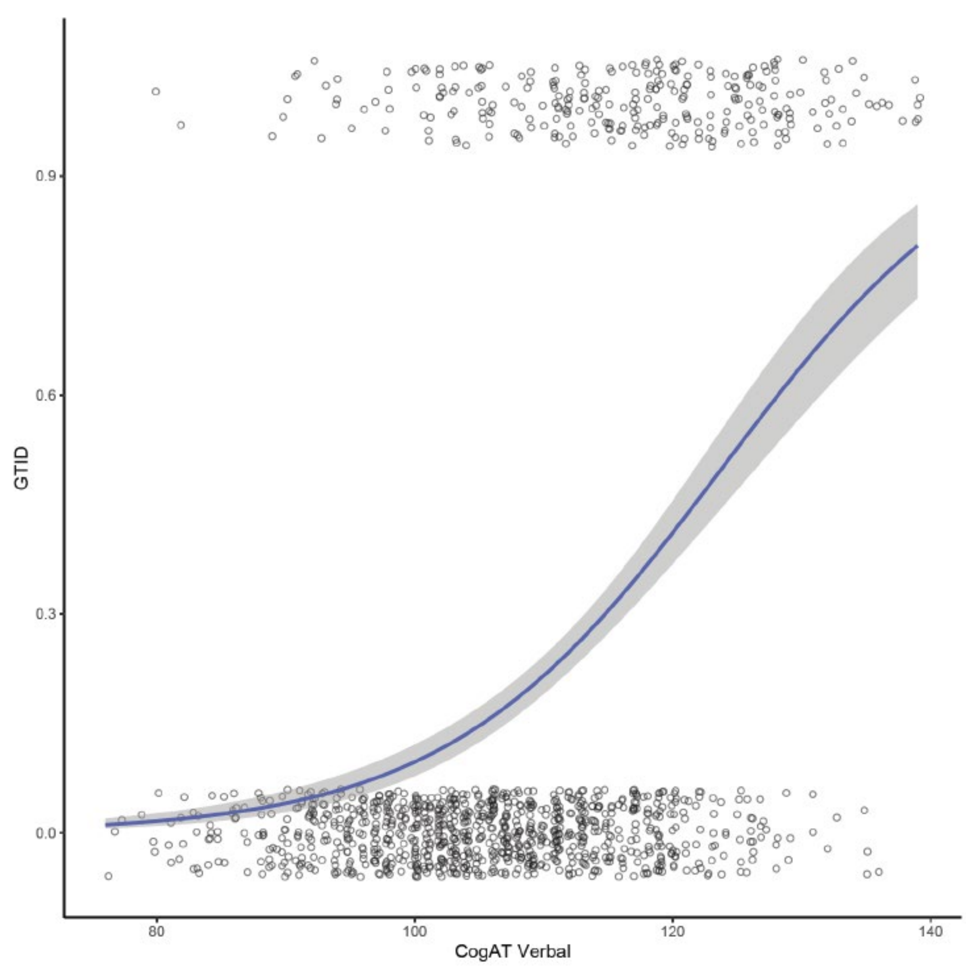


Figure 1.26

Generalized Linear Regression Plot for ITBS Reading and the Probability of Being Identified for Gifted Services by Referrals

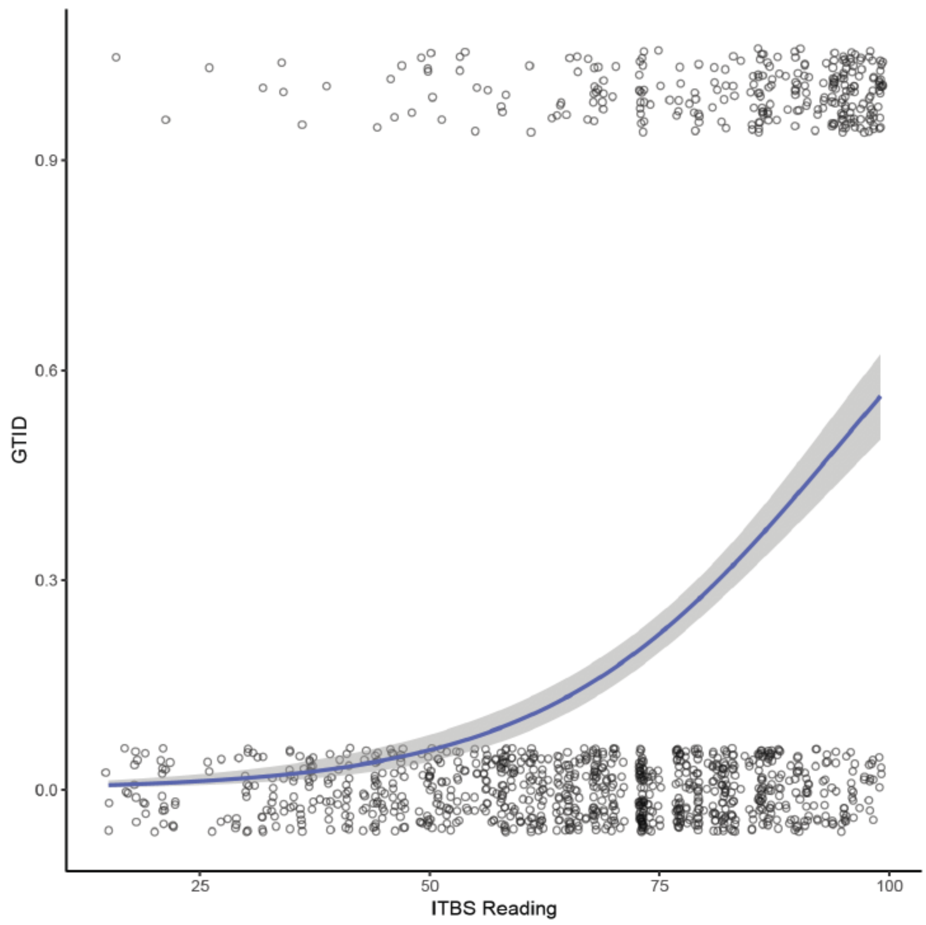


Figure 1.27

Generalized Linear Regression Plot for ITBS Math and the Probability of Being Identified for Gifted Services by Referrals

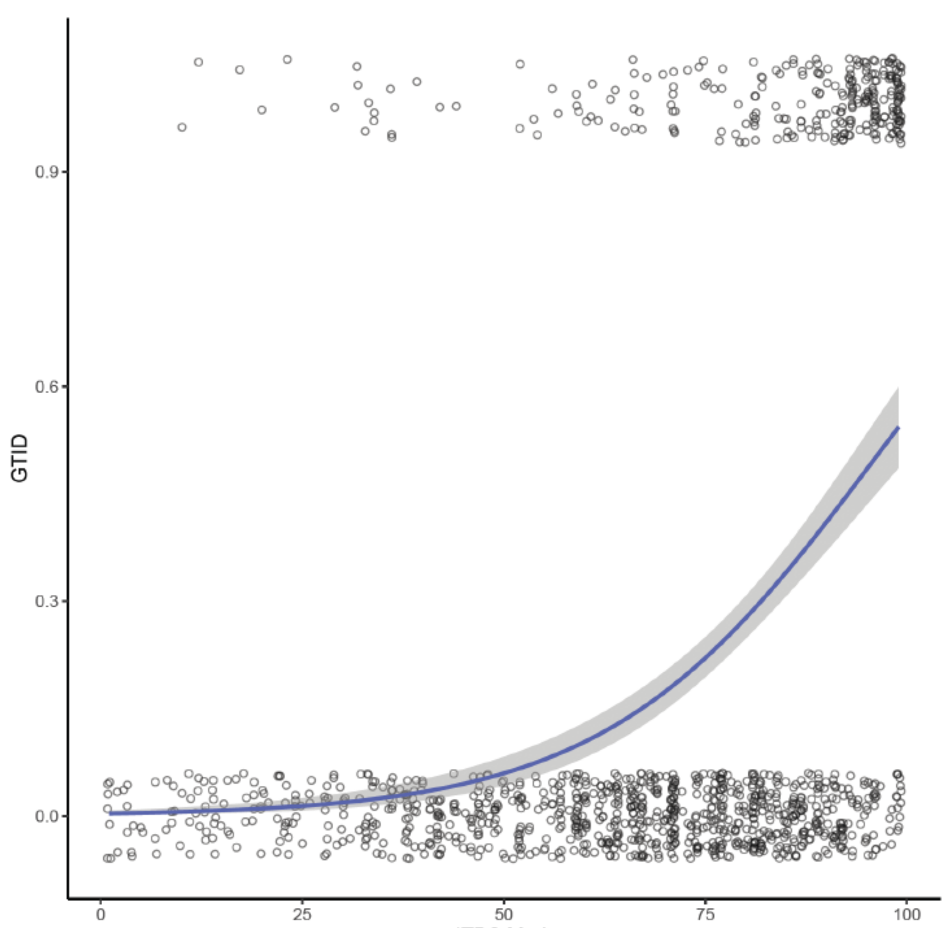


Table 1.16

Hierarchical Generalized Linear Regression for Referrals

Variable	Block 1				Block 2				Block 3				Block 4			
	b	SE(B)	Odds Ratio	95% CI	b	SE(B)	Odds Ratio	95% CI	b	SE(B)	Odds Ratio	95% CI	b	SE(B)	Odds Ratio	95% CI
Intercept	-.88***	1.12	.42	[1.20e-14, 2.88e-11]	-28.00**	7.24	6.91e-13	[1.20e-14, 2.88e-11]	-32.54***	7.38e-15***	9.97	[6.749e-17, 5.62e-13]	-29.81***	1.14e-13***	2.12	[1.48e-15, 6.15e-12]
Student Demographics																
Black/African American	-.54	.23	.58	[.37, .91]	-.02	.30	.98	[.55, 1.75]	.03	1.03	.31	[.56, 1.88]	.07	1.08	.31	[.59, 1.97]
Hispanic/Latinx	-.17	.20	.84	[.57, 1.25]	.19	.28	1.21	[.70, 2.09]	.26	1.30	.29	[.74, 2.29]	.27	1.31	.29	[0.74, 2.32]
Asian	.97***	.29	2.63***	[1.48, 4.66]	1.51***	.42	4.55***	[2.01, 10.28]	1.48	4.40***	.42	[1.92, 10.07]	1.35	3.86***	.45	[1.58, 9.41]
Native American	-.45	.56	.64	[.21, 1.90]	.002	.75	1.00	[.23, 4.39]	.26	1.30	.76	[.29, 5.77]	.43	1.54	.83	[.30, 7.85]
Other	-1.73	.74	.18	[.04, 0.75]	-.81	.85	.44	[.08, 2.35]	-.86	.42	.87	[.08, 2.33]	-5.06	.01	2.33	[.00003, .61]
Female	-.25	.14	.78	[.59, 1.03]	-.03	.19	.97	[.66, 1.42]	-.18	.83	.20	[.56, 1.24]	-.19	.83	.21	[.55, 1.24]
Free/Reduced Lunch	-.10	.17	.90	[.65, 1.26]	.41	.23	1.50	[.96, 2.37]	.43	1.53	.24	[.96, 2.44]	.37	1.45	.25	[.90, 2.36]
English Learner	-.85***	.27	.43**	[.25, 0.73]	-.70	.38	.50	[.24, 1.04]	-.63	.53	.38	[.25, 1.12]	-.57	.56	.39	[.26, 1.21]
Cognitive Ability																
CogAT Non-Verbal					.06***	.01	1.07***	[1.04, 1.09]	.06***	1.07***	.01	[1.04, 1.09]	.06***	1.07***	.01	[1.05, 1.09]
CogAT Quantitative					.07***	.01	1.07***	[1.04, 1.09]	.07***	1.07***	.01	[1.05, 1.09]	.07***	1.07***	.01	[1.05, 1.10]
CogAT Verbal					.05***	.01	1.05***	[1.03, 1.07]	.05***	1.05***	.01	[1.03, 1.07]	.05***	1.06***	.01	[1.03, 1.08]
Academic Achievement																
ITBS Reading					.05***	.01	1.06***	[1.04, 1.07]	.06***	1.06***	.01	[1.04, 1.08]	.06***	1.06***	.01	[1.05, 1.08]
ITBS Math					.03***	.01	1.03***	[1.02, 1.05]	.03***	1.03***	.01	[1.02, 1.04]	.03***	1.03***	.01	[1.02, 1.04]
Creativity																
TTCT-Figural									.03***	1.03***	.01	[1.02, 1.05]	.03	1.03	.01	[1.00, 1.05]

(table continues)

Variable	Block 1				Block 2				Block 3				Block 4			
	b	SE(B)	Odds Ratio	95% CI	b	SE(B)	Odds Ratio	95% CI	b	SE(B)	Odds Ratio	95% CI	b	SE(B)	Odds Ratio	95% CI
Interaction Effects																
TTCT*Sex														1.00	.01	[.97, 1.03]
TTCT*English Learner														.99	.03	[.93, 1.04]
TTCT*Free/ Reduced Lunch														1.02	.02	[.99, 1.06]
TTCT*Black/ African American														.98	.02	[.94, 1.03]
TTCT*Hispanic/ Latinx														1.01	.02	[.98, 1.05]
TTCT*Asian														1.03	.03	[.98, 1.09]
TTCT*Native American														1.03	.06	[.92, 1.15]
TTCT*Other														1.36	.11	[1.09, 1.70]
Tjur R^2	.039				.457				.476				.483			

Note. CogAT7 = Cognitive Abilities Test. ITBS = Iowa Test of Basic Skills. TTCT = Torrance Test of Creative Thinking. NNAT-2 = Naglieri Nonverbal Ability Test 2nd Edition. SE = Standard Error. CI = Confidence Interval. SE(B) and CI are related to the odds ratio. For Block 4, all continuous predictors were mean-centered. **p* < .05. ***p* <.01.****p*< .001.

Table 1.17

Hierarchical Generalized Linear Regression for Universal Screening

Variable	Block 1				Block 2				Block 3				Block 4			
	b	Odds Ratio	SE(B)	95% CI	b	Odds Ratio	SE(B)	95% CI	b	Odds Ratio	SE(B)	95% CI	b	Odds Ratio	SE(B)	95% CI
Intercept	-2.26***	0.10***	1.16	[.08, .13]	-35.76***	2.95e-16***	14.77	[1.14e-18, 4.44e-14]	-41.33**	1.12e-18	22.97	[1.73e-21 3.84e-16]	-4.99***	.007	1.44	[.003, .013]
Student Demographics																
Black/African American	-.81	.44	.35	[.22, .87]	-.35	.70	.44	[.29, 1.68]	-.38	.69	.45	[.29, 1.65]	-.59	.56	.54	[.19, 1.61]
Hispanic/Latinx	-.24	.79	.25	[.48, 1.30]	.57	1.77	.33	[.92, 3.38]	.69	1.99	.34	[1.02, 3.88]	.76	2.14	.35	[1.08, 4.26]
Asian	.84	2.33	.35	[1.18, 4.60]	1.03	2.79	.48	[1.10, 7.09]	1.38	3.98*	.50	[1.50, 10.53]	1.29	3.63	.51	[1.34, 9.83]
Female	-.07	.93	.19	[.65, 1.33]	-.73	0.48*	.24	[.30, .78]	-.91	.40**	.25	[.24, .66]	-.93	0.39**	.26	[.23, .66]

(table continues)

Variable	Block 1				Block 2				Block 3				Block 4			
	b	Odds Ratio	SE(B)	95% CI	b	Odds Ratio	SE(B)	95% CI	b	Odds Ratio	SE(B)	95% CI	b	Odds Ratio	SE(B)	95% CI
Free/Reduced Lunch	-1.03***	.36***	.27	[.21, .61]	-.49	.61	.33	[.32, 1.17]	-.54	.58	.34	[.30, 1.12]	-.57	.56	.36	[.28, 1.13]
English Learner	-.36	.70	.38	[.33, 1.48]	-.53	.59	.46	[.24, 1.45]	-.37	.69	.47	[.27, 1.73]	-.43	.65	.47	[.26, 1.64]
Cognitive Ability																
NNAT-2					.02	1.06***	.01	[1.03, 1.08]	.05	1.05***	.01	[1.03, 1.08]	.06	1.06***	.01	[1.03, 1.08]
CogAT Non-Verbal					.02	1.02	.01	[1.00, 1.04]	.02	1.02	.01	[1.00, 1.05]	.02	1.02	.01	[1.00, 1.05]
CogAT Quantitative					.10	1.11***	.01	[1.08, 1.14]	.10	1.11***	.01	[1.08, 1.14]	.11	1.11***	.01	[1.08, 1.14]
CogAT Verbal					.08	1.09***	.01	[1.06, 1.11]	.09	1.09***	.01	[1.06, 1.12]	.09	1.09***	.01	[1.07, 1.12]
Academic Achievement																
ITBS Reading					.05	1.05**	.01	[1.02, 1.07]	.05	1.05**	.01	[1.02, 1.07]	.05**	1.05**	.01	[1.02, 1.07]
ITBS Math					.01	1.01	.01	[1.00, 1.02]	.01	1.01	.01	[1.00, 1.02]	.01	1.01	.01	[1.00, 1.02]
Creativity																
TTCT-Figural									.04	1.05***	.01	[1.03, 1.06]	.04	1.04	.02	[1.01, 1.08]
Interaction Effects																
TTCT*Sex														1.01	.02	[.97, 1.04]
TTCT*English Learner														1.04	.04	[.97, 1.11]
TTCT*Free/ Reduced Lunch														1.01	.02	[.97, 1.06]
TTCT*Black/ African American														1.03	.04	[.96, 1.10]
TTCT*Hispanic/ Latinx														.98	.02	[.93, 1.03]
TTCT*Asian														.95	.03	[.89, 1.01]
Tjur R^2	.023				.395				.417				.420			

Note. CogAT7 = Cognitive Abilities Test. ITBS = Iowa Test of Basic Skills. TTCT = Torrance Test of Creative Thinking. NNAT-2 = Naglieri Nonverbal Ability Test 2nd Edition. SE = Standard Error. CI = Confidence Interval. SE(B) and CI are related to the odds ratio. For Block 4, all continuous predictors were mean-centered. **p* < .05. ***p* <.01.****p*< .001.

Figure 1.28

Generalized Linear Regression Plot for TTCT-Figural and the Probability of Being Identified for Gifted Services by Students Universally Screened in First Grade

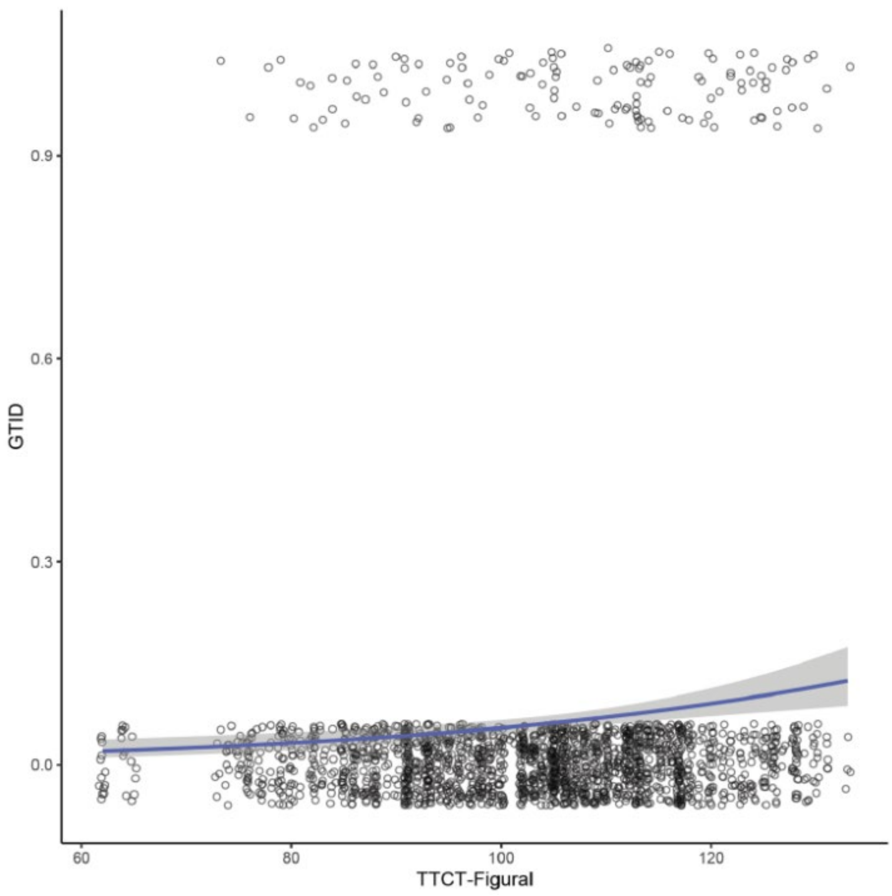


Figure 1.29

Generalized Linear Regression Plot for CogAT Non-Verbal and the Probability of Being Identified for Gifted Services by Students Universally Screened in First Grade

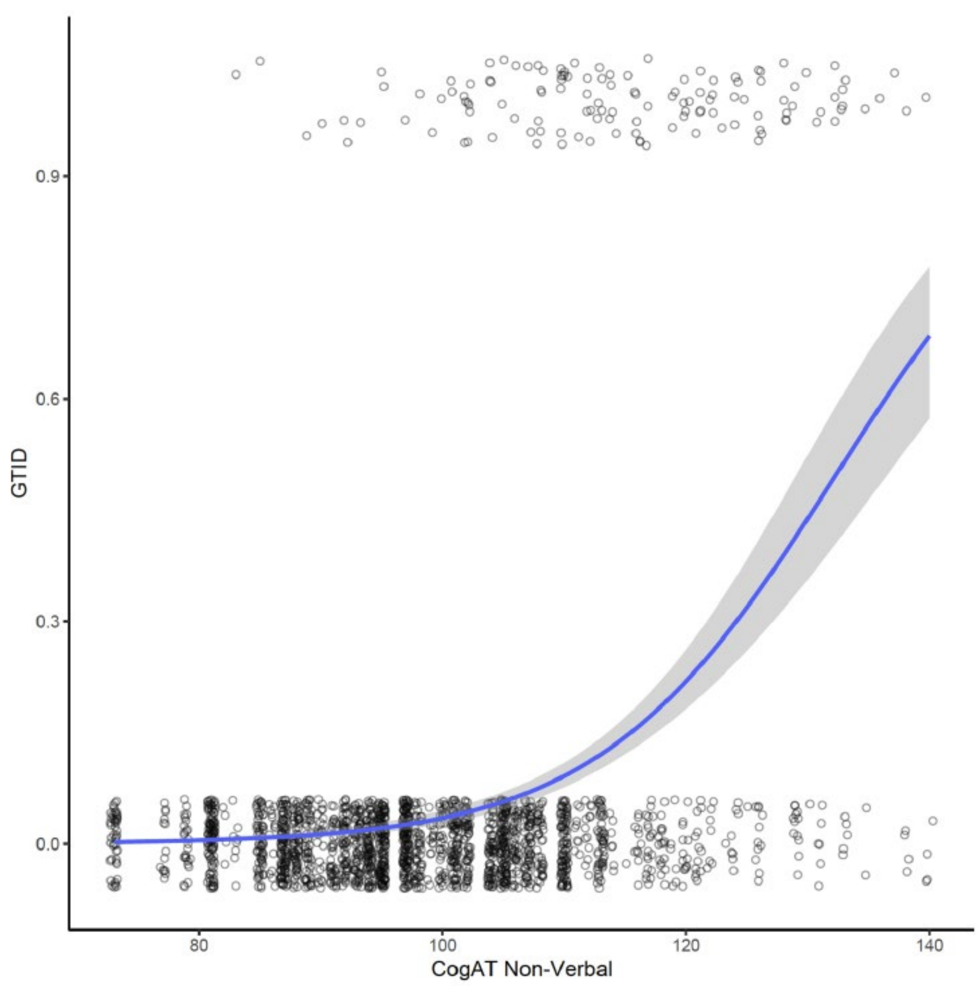


Figure 1.30

Generalized Linear Regression Plot for CogAT Quantitative and the Probability of Being Identified for Gifted Services by Students Universally Screened in First Grade

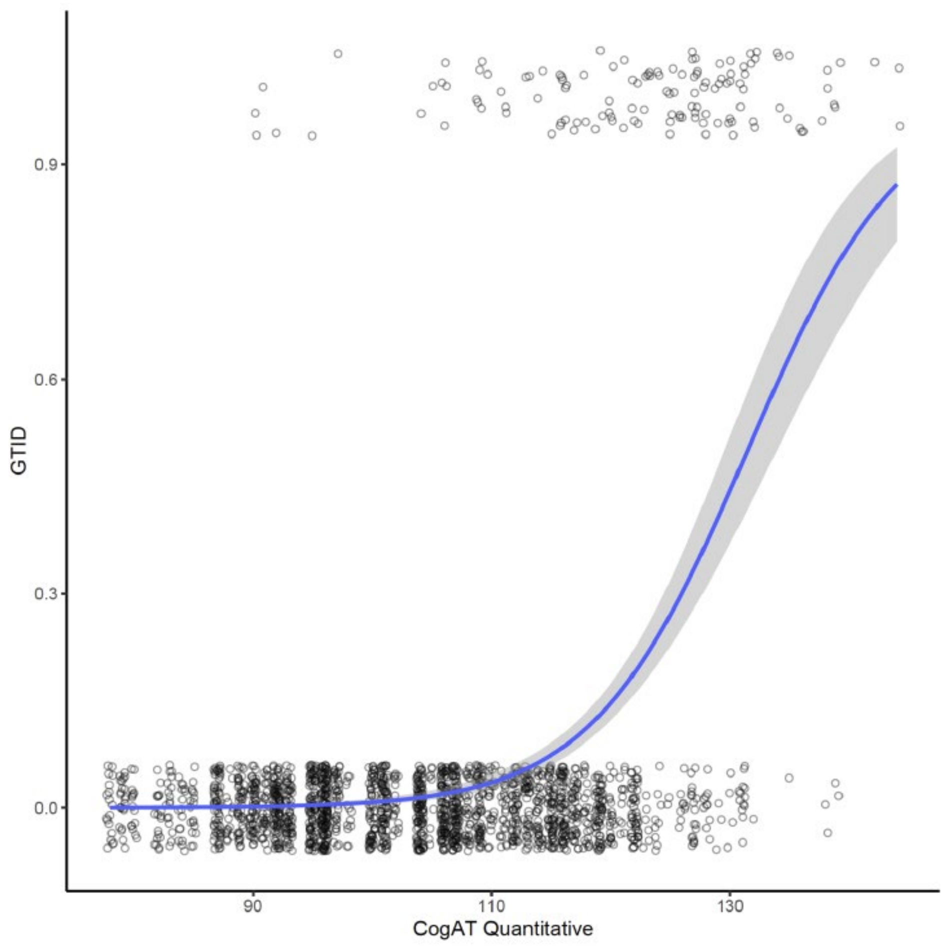


Figure 1.31

Generalized Linear Regression Plot for CogAT Verbal and the Probability of Being Identified for Gifted Services by Students Universally Screened in First Grade

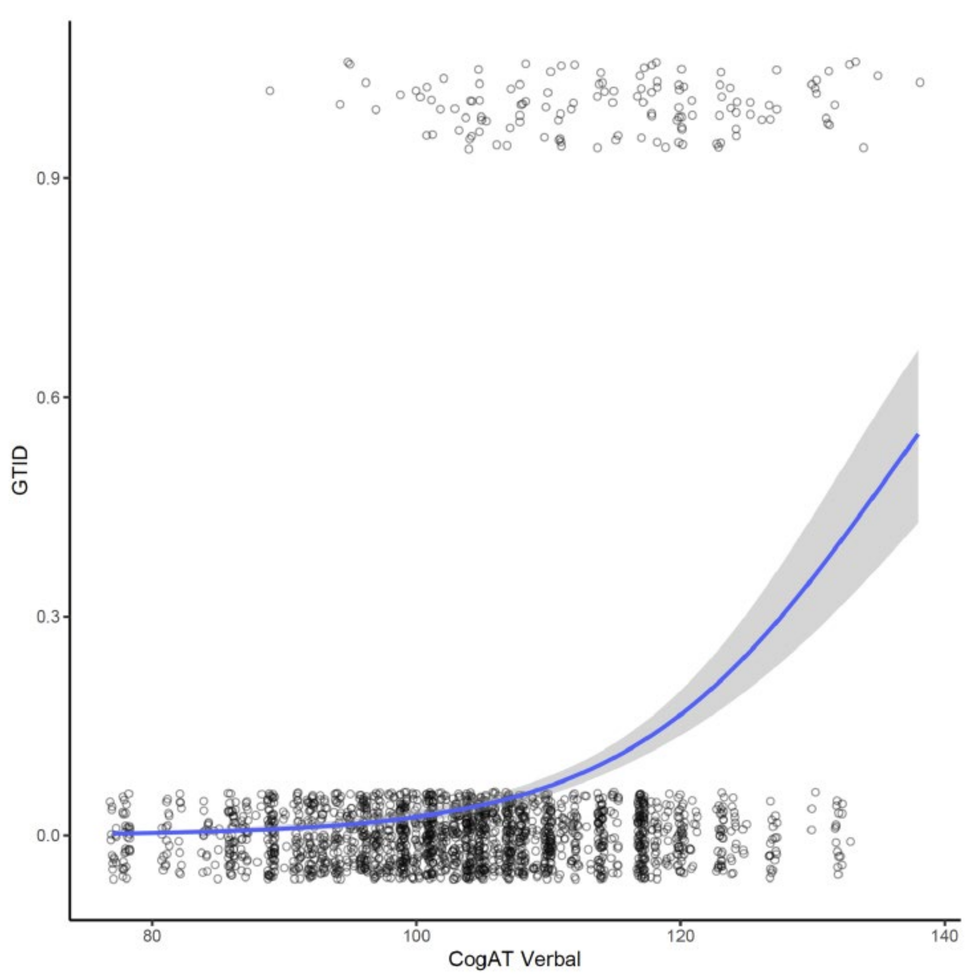


Figure 1.32

Generalized Linear Regression Plot for ITBS Reading and the Probability of Being Identified for Gifted Services by Students Universally Screened in First Grade

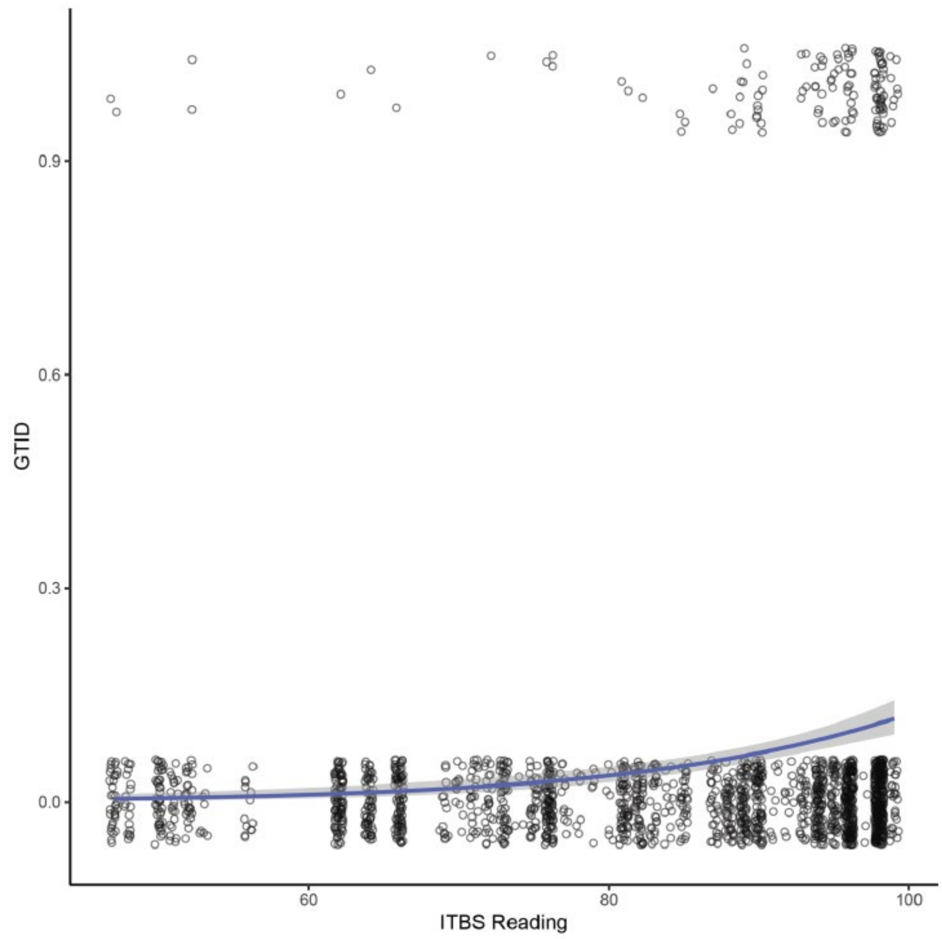


Figure 1.33

Generalized Linear Regression Plot for ITBS Math and the Probability of Being Identified for Gifted Services by Students Universally Screened in First Grade

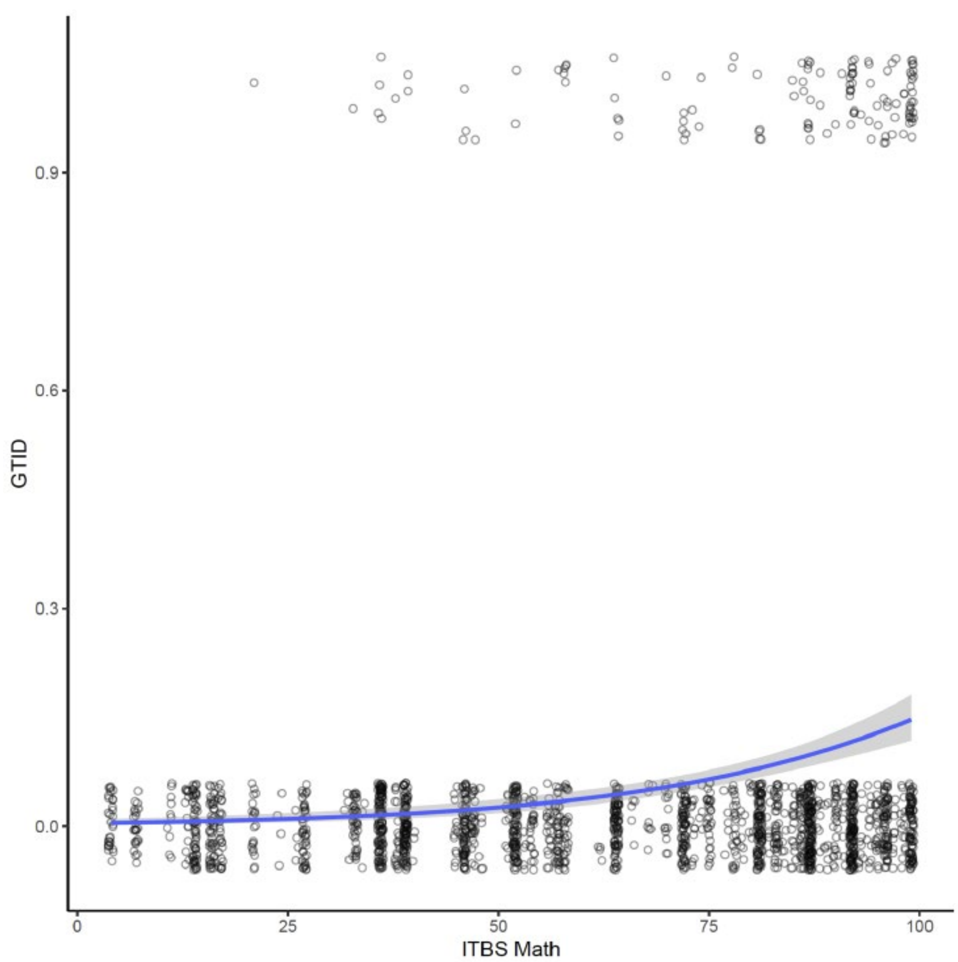


Figure 1.34

Generalized Linear Regression Plot for NNAT-2 and the Probability of Being Identified for Gifted Services by Students Universally Screened in First Grade

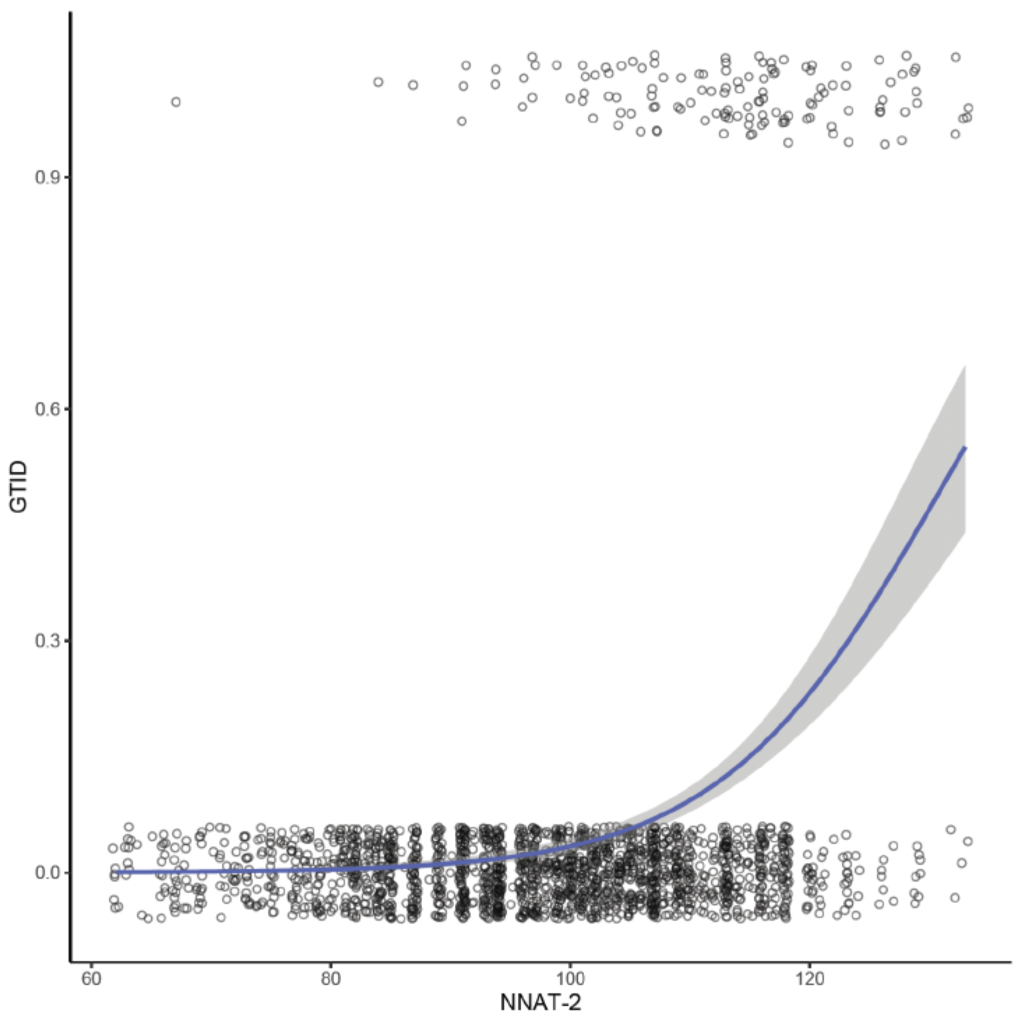


Table 1.18

Multilevel Generalized Linear Regression for Referrals

Fixed Effect	b	Final Model		
		Odds Ratio	SE (B)	95% CI
Intercept	-2.70***	.07***	.23	[.04, .11]
Student Demographics				
Black/African American	.03	1.03	.31	[.56, 1.90]
Hispanic/Latinx	.26	1.30	.29	[.73, 2.29]
Asian	1.44**	4.23**	.43	[1.83, 9.79]
Native American	.23	1.26	.76	[.28, 5.61]
Other	-.80	.45	.87	[.08, 2.48]
Female	-.18	.84	.20	[.56, 1.24]
Free/Reduced Lunch	.42	1.52	.24	[.94, 2.44]
English Learner	-.67	.51	.38	[.24, 1.09]
Average Cognitive Ability				
CogAT7 Non-Verbal	.06***	1.06***	.01	[1.04, 1.09]
CogAT7 Quantitative	.07***	1.07***	.01	[1.05, 1.10]
CogAT7 Verbal	.05***	1.05***	.01	[1.03, 1.08]
Average Academic Achievement				
ITBS Reading	.06***	1.06***	.01	[1.05, 1.08]
ITBS Math	.03***	1.03***	.01	[1.02, 1.04]
Average Creativity				
TTCT-Figural	.03***	1.03***	.01	[1.02, 1.05]
Random Effect τ00		.09		
Marginal R^2		.720		
Conditional R^2		.728		
AIC		741.68		
BIC		823.00		

Note. CogAT7 = Cognitive Abilities Test. ITBS = Iowa Test of Basic Skills. TTCT = Torrance Test of Creative Thinking. NNAT-2 = Naglieri Nonverbal Ability Test 2nd Edition *p < .05. **p <.01. ***p< .001.

Table 1.19

Multilevel Generalized Linear Regression for Students Universal Screened in First Grade

Fixed Effect	b	Final Model		
		Odds Ratio	SE (B)	95% CI
Intercept	-5.76***	.01***	.38	[.001, .01]
Student Demographics				
Black/African American	-.56	.75	.46	[.30, 1.85]
Hispanic/Latinx	.59	2.09	.35	[1.05, 4.16]
Asian	1.33*	4.42*	.51	[1.63, 12.00]
Female	-1.06***	.39***	.26	[.24, 0.65]
Free/Reduced Lunch	-.55	.56	.35	[.28, 1.12]
English Learner	-.38	.66	.48	[.26, 1.70]
Average Cognitive Ability				
NNAT-2	.06***	1.06***	.01	[1.03, 1.08]
CogAT7 Non-Verbal	.03	1.02	.01	[1.00, 1.05]
CogAT7 Quantitative	.12***	1.11***	.01	[1.08, 1.14]
CogAT7 Verbal	.09***	1.09***	.01	[1.07, 1.12]
Average Academic Achievement				
ITBS Reading	.05***	1.05**	.01	[1.02, 1.07]
ITBS Math	.02	1.01	.01	[1.00, 1.02]
Average Creativity				
TTCT-Figural	.05***	1.04***	.01	[1.02, 1.06]
Random Effect τ00		.26		
Marginal R^2		.721		
Conditional R^2		.742		
AIC		520.13		
BIC		604.7		

CHAPTER 2

THE SEARCH FOR CREATIVITY AND TALENT IN DIVERSE POPULATIONS: A SYSTEMATIC REVIEW

2.1 Abstract

Despite decades of research, creativity remains an elusive construct within the search for domain-specific talent in diverse student populations. The purpose of this article was to conduct a systematic review of the empirical literature on how creative measures related to equitable access in gifted programs. Thematic analysis was used to analyze the reviewed articles and the following three overarching themes emerged: (a) overcoming identification barriers through inclusive practices for developing exceptional talent, (b) conceptions of creativity used in identification processes, (c) research focused on validity and reliability of creative measures.

2.2 Introduction

Culturally, linguistically, and economically diverse (CLED) students and twice-exceptional students (e.g., students identified as both gifted and having a disability; Rimm et al., 2018) have historically been underrepresented and underserved within gifted programs across the country. This is largely attributable to marginalization, discrimination, and systemic inequities found in education that have long impacted gifted and advanced programs (Ford et al., 2008; Mun et al., 2020a; Worrell & Dixson, 2018). As the United States population continues to become increasingly diverse over time (Mun et al., 2020a; U. S. Department of Education, 2020), the United States educational systems need further evaluation and refinement to best meet the needs of students. This continued underrepresentation in gifted education is, in part, due to an enduring history of achievement gaps, coupled with deficit perspectives (i.e., negative, stereotypical views) and a desire for conforming behaviors in classrooms (Erwin & Worrell, 2012; Rimm et al., 2018; Subotnik et al., 2011; Torrance, 1971, 2004).

Although gifted programs have been historically fraught with issues of inequitable access, there has been considerable traction in mitigating disparities to advanced educational opportunities. Decades ago, more broadened conceptions of giftedness emerged and have impacted how we identify students for gifted programs (Frasier, 1997, Renzulli, 1978). A significant contribution to changing conceptions of giftedness across the United States was the creation of the Marland Report (1972), when the federal government publicly issued a federal definition with a broadened operationalization of giftedness that states “gifted and talented children are those identified...by virtue of outstanding abilities, are capable of high performance” and “have demonstrated achievement and/or potential ability” (p. 2) in an array of different areas either solely or simultaneously. These areas include general intellectual ability, specific academic aptitude, creative or productive thinking, leadership ability, visual and performing arts aptitude, and psychomotor ability.

This multidimensional view of giftedness provided grounds for the usage of multiple criteria within the identification process to identify various abilities within school programs (Jolly & Robins, 2018). Many school systems in the United States utilize multiple criteria to determine entry into gifted programming (e.g., *State of the States*; Rinn et al., 2020). This is also due to multiple criteria systems being encouraged to mitigate underrepresentation within gifted education with the inclusion of multiple data points to better exemplify the strengths of students in the identification process (Erwin & Worrell, 2012;). Lohman (2009) describes how aptitude and achievement are not the only considerations in identification processes, as “motivation, interest, and creativity” (p. 983) also are important criteria to include. Criteria used in identification practices have evolved from traditional conceptions focused solely on scores of cognitive ability (e.g., Lohman, 2012; Cognitive Ability Test [CogAT]) or achievement, and have evolved to become multidimensional with the inclusion of non-cognitive components (e.g., creativity, motivation, leadership). Within multiple criteria systems, alternative measures are commonly incorporated to assess non-cognitive components (e.g., creativity, learning, leadership, motivation, interest) in the form of checklists/rating scales, portfolios, observations, performance-based assessments, or norm-referenced non-cognitive tests (VanTassel-Baska, 2008).

The mere usage of alternative measures is not going to solve issues of equity, as judgments for how to weigh and combine alternative criteria with ability and achievement are still ambiguous within identification processes. In addition to the usage of multiple criteria systems, a central focus in gifted scholarship has been on identification and equitable access, whereby the literature has centered on the usage of non-verbal cognitive measures (Carman et al., 2020; Lohman & Grammell, 2012; Naglieri & Ford, 2003; Peters & Engarrand, 2016), universal screening (Card & Guiliano, 2016; Lakin, 2016), nomination practices (McBee et al., 2016), and how the combination of scores (Lakin, 2018; McBee et al., 2014) impacts access to gifted programs. Since cognitive ability measures are commonly used in gifted identification procedures in the form of universal screening (e.g., Carman et al., 2020), there remains unanswered questions on specific demographic differences in alternative measures that persist beyond the universal screening process. For instance, measures of creativity are often used as alternative assessments for entry into gifted programs (Lee et al., 2020; Peters et al., 2020; Van-Tassel Baska, 2008).

2.2.1 Conceptual Framework

2.2.1.1 The Search for Creativity and Talent

Within the field of gifted education, paradigmatic shifts have influenced how we identify students for gifted services (Lo & Porath, 2017). Dai and Chen (2014) suggest that identification has moved from static, monolithic conceptions of giftedness (i.e., IQ-based, cognitive ability measures) to a more dynamic, broader conception of giftedness (i.e., domain-specific, creative potential). Plucker and Callahan (2014) emphasized that we have made significant developments in empirical research regarding "identification, talent development, and creativity" (p. 395) and there is a growing body of literature on how to identify historically underrepresented students.

Individuals who espouse the paradigmatic perspective of talent development suggest identification for gifted services based on the creation of a talent pool in various domains of strength. Scholars also believe potential can be developed within specialized programs (inside or outside of school environments) to generate future creative contributions for society (Dai & Chen, 2014; Treffinger & Feldhusen, 1996; Subotnik et al., 2019). Talent development frameworks have been offered as a possible solution to mitigate barriers to access and nurture emerging talent toward "exceptional levels" (Siegle et al., 2016, p. 115). Dai and Chen (2014) emphasize how developing talent requires:

> Negotiating the priorities of excellence, equity, and diversity. Excellence (selectivity, productivity), equity, and diversity are three priorities that need to be balanced. The trade-off between maximizing participation and quality control, between agendas for diverse talents to be identified and expressed and agendas with a sharp focus, has to be thought through in programming and identification. (p. 155)

Despite decades of research with models of talent development, creativity remains an elusive construct within the search for domain-specific talent in diverse populations. Starko (2018) advises that cultivating creativity is "an essential part of advanced talent development" (p. 231), thus it is imperative to understand the role of creativity within the identification process. In numerous theories of giftedness and intelligence, creativity is a facet (Miller, 2012). This is seen in theoretical models such as the three-ring model of giftedness (Renzulli, 1978), integrative model of talent development (IMTD; Gagné, 2017), and triarchic theory of intelligence (Sternberg, 1985). Thus, assessments of creativity are often included in the search for talent.

Multiple scholars in the field have touted high levels of creative performance and productivity as the outcome of interest within talent development (Dai, 2018; Olszewski-Kubilius et al., 2017; Reis & Peters, 2020; Renzulli, 2012; Subontik et al., 2011; Tannenbaum, 2003; Treffinger, 2004). Specifically, Subotnik et al.'s (2011) talent development megamodel, specified creative performance and productivity as outcomes related to development of eminence in a talent domain; Theories of creativity and expertise are intertwined within this model. Particularly, the creative person learns the creative process within a domain to creatively perform or produce (Rhodes, 1961; Subotnik et al., 2011; Tannenbaum, 1983).

Domain-specificity. Scholars within the field of creativity and gifted education have credited the need for domain-specific (e.g., content and task) measures of creative potential to help identify talent (Baer, 2016; Kaufman et al., 2012; Plucker, 2011). Treffinger and Feldhusen (1996) pointed out that the identification of creativity within a child according to their domain specific ability could help in designing educational programs to enhance and nurture these abilities. The role of domain-specific ability or talent is reiterated throughout conceptual frameworks including Tannenbaum's (1983, 2003) sea star model and Gagné's (2017) integrative model of talent development that recognize individuals contain intellectual abilities and non-intellectual abilities and the importance of the interaction of personality, environment, and chance in the development of talent.

2.2.1.2 Creativity as Criteria

When using multiple criteria, Torrance (2004) advocated:

> Creativity should almost always be one of the criteria, though not the sole criterion. In general, when creativity indicators are used, students who might otherwise be missed, should be included rather than to exclude anyone. (p. 84-85)

If creativity should be criteria to include, this requires administrators to specify instruments and procedures in the identification process to identify creative potential. Hunsaker and Callahan (1995) posit the usage of a single creativity measure within gifted identification practices does not properly reflect the range of possible creative behaviors that children can exhibit; Thus, urging more than one creativity assessments to be used in the process. However, there are numerous measures of creativity to choose from when making decisions for identification. In addition, decisions are further confounded by persistent controversy surrounding construct and criterion validity of creativity assessments (e.g., Lemons, 2011; Kaufman et al., 2008). Determination of what creativity assessments to use for gifted identification is also a practical concern for administrators.

2.2.1.3 Measures of Creativity

Creativity measures can be norm-referenced or criterion-referenced assessments (Crocker & Algina, 1986), as well as performance-based or non-performance based (e.g., Acar et al., 2016). Norm-referenced assessments require comparing performance to a specific group, whereas criterion-referenced assessments measure performance to specific objectives or criteria with the aim to assess mastery without comparing to others (Crocker & Algina, 1986). For example, both types of assessments are seen in the figural form of Torrance Test of Creative Thinking which uses both norm-referenced (e.g., subscales that measure fluency, originality, elaboration, abstractness of titles, and resistance to premature closure) and criterion-referenced elements (i.e., Creative Strengths Checklist). Performance-based assessments result in a score on specific tasks performed by the student (Acar et al., 2016; VanTassel-Baska et al., 2007).

Performance-based measures of creative expression have been encouraged (Kaufman et al., 2012), especially assessments that consider expert opinions of performance (e.g., Consensual Assessment Technique; Amabile, 1996). On the other hand, non-performance methods utilize internal and external "judgments, instead of test scores" (Acar et al., 2016, p. 82) from teachers, parents, or other students. For example, non-performance approaches use "teacher rating scales, teacher and parent nominations, self-ratings, and peer-ratings" (p. 82) and characterize many alternative assessments used within gifted identification practices.

2.2.2 The Current Study

Through the paradigmatic lens of talent development, equity and excellence are notable priorities as well as outcomes of domain-specific creative performance and production. Luria et al. (2016) suggests that gifted programs should place more emphasis on the use of creative measures within identification processes to increase diversity in gifted programs. The inclusion of measures of creativity have long been a component of many gifted identification processes (Torrance, 1974, 2004; Treffinger, 2004), however creativity is often a secondary concern and not of central focus in identification research as another solution to mitigate demographic disparities in gifted education (Baldwin, 2002). From a psychometric standpoint, how we assess students for developing their talent should be linked to the specified goals and outcomes of talent development programs (Callahan et al., 1995; Kaufman et al., 2008; Lee et al., 2020). If the purposes of talent development include identifying student strengths in particular domains, developing creative potential into creative productivity, and lessening demographic disparities in identification and achievement, then it seems logical to incorporate identification methods that assess creativity. However, there remains uncertainty on how the role of creativity assessments relate to equitably identifying talent for gifted programs. Underrepresentation of CLED students remains an issue within gifted education. Talent development models and programs have largely focused on criterion validity of intellectual ability and achievement, yet creativity has not had a larger focus in relation to equitably identifying domain-specific talent. The purpose of this article is to provide a systematic review of the empirical literature on the role of creativity used in gifted identification processes and the relation to equitable access in gifted programs. The following research questions guided this systematic review of the literature:

1. What is the role of creativity in gifted identification processes in relation to equitable access in gifted programs?
2. What is the status of research on assessments that measure creativity in relation to equitable access in gifted identification?
3. What recommendations are given in the literature about creativity assessments regarding inclusive gifted identification practices?

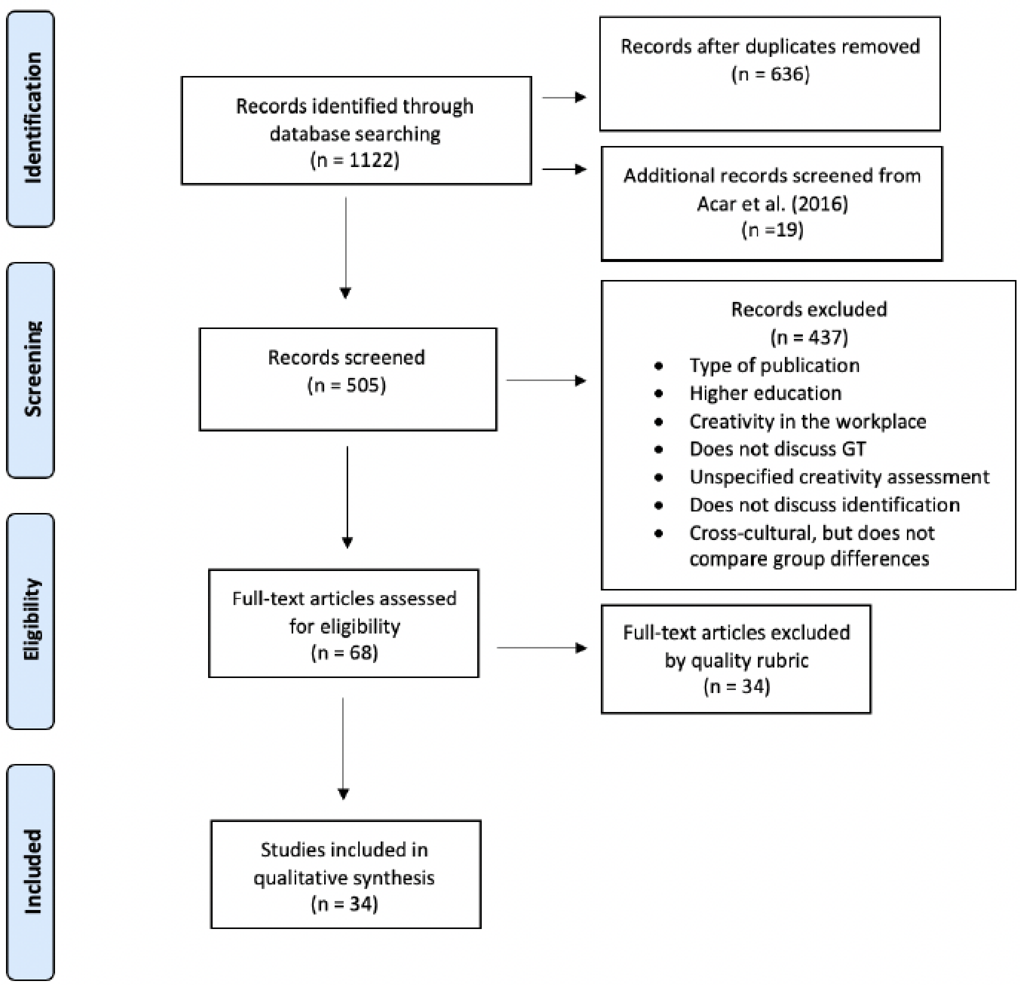
2.3 Methods

2.3.1 Search Parameters

The search for literature on creativity within gifted identification processes and equitable access within gifted programs included the domains of education, psychology, and creativity. The electronic databases searched within this review were as follows: Academic Search Complete, Educational Resources Information Center (ERIC), Education Source, Professional Development Collection, Psychology and Behavioral Sciences Collection, APA PsycArticles, and APA PsycINFO. Each search was limited to empirical, peer-reviewed journal articles and published in English. The time frame was not restricted in this search. The oldest articles found were from 1959, meaning the articles were either the earliest found in the database or the included journals only went back to 1959. I used advanced searches within each database and searched the terms “creative ability” or “creative potential” or “creativity” within all texts in the database, and searched the terms “gifted”,”talent”, “identification,” or “assessment” within abstracts; this was to yield more broad search returns that incorporated creativity anywhere in the text of the manuscript. I used an open-source reference management software called Zotero (Zotero, 2020) to export bibliographic information (e.g., citations, abstracts, journals) from each database to be scanned within the next initial screening phase. Also, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist was used to guide analysis (Liberati et al., 2009; Moher et al., 2009; see Figure 2.1)). These searches generated 1112 articles, and 626 were found to be duplicates (see Table 2.1).

Figure 2.1

PRISMA Search Protocol



Note. PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

2.3.2 Inclusion Criteria

Once duplicates were removed, the remainder of articles were screened by title and abstract to ensure initial requirements for inclusion (*n* = 486). The meta-analysis by Acar et al. (2016) was used as an additional resource to screen articles based on their use of instruments that have measures of creativity (*n* = 19). Therefore, a total of 505 articles were screened for initial inclusion. Articles were included if they discussed creativity (i.e., divergent thinking, problem-solving), identification, K-12 student participants or experiences, and were empirical (e.g., quantitative, qualitative, mixed methods; see Table 2.2). Also, articles were included if they analyzed group differences (i.e., gender, race/ethnicity, socioeconomic status, rurality, English language learners, special education, or learning disabilities) or discussed experiences of underserved populations (i.e., Black/African American, Hispanic, Native Americans) within gifted programs. Cross-sectional studies were included within this initial phase of screening to determine if they examined group differences in the quality screening phase. After initial screening, 437 articles were excluded, and 68 articles were kept to continue for quality screening.

2.3.3 Quality Screening

To assess the quality of the included articles, the Mixed Methods Appraisal Tool (MMAT) was used to determine final inclusion (Hong et al., 2018). The MMAT is a critical appraisal tool that is used for empirical studies and uses guiding questions to determine criteria for methodological quality. The first appraisal consists of two general screening questions on if the clarity of the research questions and the second question asks if the collected data connect to those questions. If the answer is “no” or “can’t tell”, then those articles may not be able to answer the further questions based on methods used in the article. If articles did not include research questions, a purpose statement or hypotheses sufficed for this review. The remainder of criteria appraised was based on the category of the study design (i.e., qualitative, quantitative non-randomized, quantitative descriptive, mixed methods). At this stage, cross-sectional studies were eliminated if they did not include comparison of subgroups (i.e., gender, race/ethnicity, socioeconomic status, disability, or special education status) or a historically underrepresented group. Quantitative articles were excluded if there was no clear purpose or research questions, used inappropriate statistical analyses, or had incomplete descriptions of results or missing discussions; qualitative articles were excluded if they did not explicitly state how they qualitatively analyzed the data or if they did not include methods to establish trustworthiness (i.e., credibility, transferability, dependability, confirmability; Lincoln & Guba, 1985). After quality screening, 34 articles remained for further data analysis. See Table 2.3 for included studies from the quality rubric.

2.3.4 Data Analysis

Included articles were qualitatively examined using descriptive and interpretive thematic analysis. Descriptive analyses aim "to summarize and describe patterned meaning", whereas interpretive analyses go beyond basic description to interpret "deeper meanings" found within data (Clarke et al., 2015, p. 226). Six phases of thematic analysis were used in this analysis that include: familiarization, coding, searching for themes, reviewing themes, defining themes, and writing the report (Braun & Clarke, 2006). Within this study, familiarization occurred through reading each article and making initial notes and observations. Then, the first phase of coding included the creation of attribute codes to describe specific aspects of each study (e.g., demographic information, instruments used; Richards, 2015) and descriptive codes to identify topics within each article that relate to the research questions (e.g., equitable access, inclusive practices, conceptions of creativity). Descriptive information for each article (e.g., purpose, participants, methods, major findings, conceptions of creativity, instruments used), as well as article excerpts and data memos assisted in thematic organization. In the second phase of coding, pattern coding of each article condensed attributes and descriptive codes into a smaller number of categories. In order to search for themes, these categories were analyzed to determine broader themes and subthemes derived from the data (Ryan & Bernard, 2003). Themes and subthemes were reviewed and broadly defined within the context of the research questions (Clarke et al., 2017). See Table 2.4 for coding structure, subthemes, and themes.

2.4 Results

Articles included within this systematic review ranged in publication dates from 1970 to August 2020. The studies included were conducted in Australia, Brazil, Canada, China, Germany, Jordan, Lebanon, Puerto Rico, Romania, Spain, South Korea, Turkey, and the United States; the majority of studies were conducted within the United States (n = 24). Table 2.5 summarizes the characteristics from each study.

Three overarching themes emerged from the reviewed articles: (a) overcoming identification barriers through inclusive practices for developing exceptional talent, (b) conceptions of creativity used in identification processes, (c) research focused on validity and reliability of creative measures. Details of themes and subthemes are discussed below.

2.4.1 Overcoming Identification Barriers Through Inclusive Practices for Developing Talent

A major focus was overcoming identification barriers through inclusive practices. Articles discussed specific barriers to talent development, how to incorporate inclusive identification practices, and broadening conceptions of giftedness to identify and develop exceptional talent.

2.4.1.1 Barriers to Talent Development

Underrepresentation in gifted programs has been a perpetual issue within educational systems (Anderson, 2020; Harris et al., 2009; Maker, 2020). Romanoff et al. (2009) pointed out that these educational inequities are present in kindergarten and "the limited presence of several minority groups among high achieving students cuts across class lines" (p. 158). Demographic subgroups at risk of underrepresentation and being underserved within gifted programs mentioned within articles included: racial/ethnic backgrounds (i.e., Black, Hispanic, Native American), students on free/reduced lunch, multilingual students, students served within special education, students with learning disability status, and females.

Specific obstacles mentioned within identification practices related to standardized test bias (Chambers, 1980; Romanoff et al., 2009), restrictive procedures (e.g., two-stage systems; Peters & Pereira, 2017), high cut-off scores (Nakano et al., 2016), monetary costs (Rosado et al., 2008), time-consuming assessments (Mann, 2009; Rosado et al., 2008), and problems related to "lack of program match for such students once they were identified" (VanTassel-Baska et al., 2007, p. 8). Other barriers included nomination bias and/or deficit thinking (Harris et al., 2009; Oreck et al., 2003; Zhabonva et al., 2015), students masking their abilities and underachieving (Cunningham et al., 1998; Maker, 2020), and lack of communication between parents, general education teachers, and special education or ELL teachers (Harris et al., 2009).

Chambers (1980) echoes sentiments shared by current scholars that commonly used standardized measures (e.g., IQ tests) and rigid cut-off scores within the top 3-5% "severely penalizes persons with culturally different backgrounds" (p. 123). Further, deficit-oriented perspectives of ability can impact nomination practices and create unconducive learning environments (Anderson, 2020; Diener et al., 2014). Multiple authors suggest the need for re-evaluation of existing instruments to determine if they are valid and reliable measures used in identification systems as they could exacerbate underrepresentation if flawed (Peters & Pereira, 2017; Pfeiffer et al., 2007).

2.4.1.2 Inclusive Identification Practices

Integrating more inclusive (rather than exclusive) identification practices is one-way schools can overcome barriers to talent development (Subhi, 1997; VanTassel-Baska et al., 2007). Nakano et al. (2016) suggested to integrate a more inclusive “comprehensive process” to capture the “multidimensionality and complexity of giftedness and talent” (p. 628). In essence, identification processes should utilize multiple criteria to provide an opportunity for students from diverse backgrounds and abilities to be identified (Maker, 2020; Romanoff et al., 2009; Subhi, 1997). Harty et al. (2001) advocates for how gifted identification needs multiple methods to better locate students within the general population, specifically for minority and economically disadvantaged students. Beyond measures of intellectual or academic ability, multiple sources of evidence should identify creativity, leadership, motivation, and artistic talent, amongst other areas of talent (Pfeiffer et al., 2007; Pfeiffer & Jarosewich, 2007).

Identification processes usually begin with a screening stage using either a cognitive ability measure and/or nomination from parents, teachers, peers (Cunningham et al., 1998; Labâr & Frumos, 2013; Nakano et al., 2016; Peters & Pereira, 2017). McBee et al. (2016) described three criteria for “successful two-stage gifted identification systems (those that utilize a screening/nomination and confirmation phase): (a) a strong correlation between phases or measures, (b) high instrument reliability, and (c) inclusive cut scores on the nomination phase” (as cited in Peters & Pereira, 2017, p. 102).

2.4.1.3 Broadened Conceptions of Giftedness for Identifying and Developing Exceptional Talent

The majority of articles propagated an inclusive conception of giftedness, whereby a authors suggest using a broadened, multidimensional view for the identification and development of exceptional domain-specific talent (e.g., learning, motivation, leadership, creativity, artistic ability, performing arts; Maker, 2020; Maker & Zimmerman, 2020; Peters & Pereira, 2017; Sarouphim & Maker, 2010). Karadağ et al. (2016) describe the broadening of "giftedness" as the "ability to solve problems, incentive, skill, creativity, leadership, etc." (p. 8) and how this evolution of an inclusive definition of giftedness has continued to be supported by scholars in the field of gifted education. More specifically, the Marland Report (1972) was cited in six articles within this review (Cunningham et al., 1998; Harris et al., Labâr & Frumos, 2013; Kim et al., 2009; Oreck et al., 2003; Zhbanova et al., 2015), of which four articles clearly articulate the multidimensional characteristics of creativity and leadership (Kim et al., 2009; Labâr & Frumos, 2013; Oreck et al., 2003; Zhbanova et al., 2015). Furthermore, Harris et al. (2009) advocates that "identification procedures ought to concentrate on a broader conception of giftedness that includes nontraditional approaches that consider culture"(p. 371). For instance, the Munich Model of Giftedness and Talent (MMGT) and the classification of giftedness used in the U.S. Department of Education Report, *National Excellence: A Case for Developing America's Talent*, are the conceptual foundation for the Gifted Rating Scales (GRS; Pfeiffer & Jarosewich, 2003) that was utilized in seven articles (Karadağ et al., 2016; Labâr & Frumos, 2013; Li et al., 2009; Peters & Pereira, 2017; Pfeiffer & Jaroeswich, 2007; Pfeiffer et al., 2007; Rosado et al., 2008), which outlines giftedness as a reciprocal interaction amongst individual factors, personality, and environmental characteristics of which influence the active learning process that impacts achievement (Heller et al., 2005).

Strength-based. The usage of terminology to denote talent as strengths was common within articles (e.g., Discovering Intellectual Strength and Capabilities while allowing for Varied Ethnic Responses [DISCOVER]; Maker, 2001; Sarouphim, 2000, Wu et al., 2019). Rather than focusing on deficit views of ability (or what they were lacking), authors suggested educators focus on strengths of diverse students (Maker, 2020; Sarouphim, 2009). Using observation tools, such as the Frasier's Traits, Aptitudes, and Behavior's guide (TABS; Frasier et al., 1995), Anderson (2020) suggested educators should engage in identifying "strengths" within specific domains and provide students with feedback. Reid et al. (2000) also suggested using problem-solving assessments to provide a way to see "students' strengths utilized in open-ended, hands-on problem-solving as opposed to restating information in paper-and-pencil testing" (p. 7). There are other assessments that explicitly focus on searching for strengths. This is also evident within the criterion-referenced portion of the Torrance Test of Creative Thinking (Torrance, 2017), the Creative Strength Checklist, used to identify specific creative strengths (e.g., humor, storytelling articulateness, expressiveness of titles, emotional expressiveness; LaFrance, 1995).

2.4.2 Conceptions of Creativity Used in Gifted Identification Processes

Creativity as a characteristic of giftedness was echoed within all included studies. Creative ability is an explicit component within several models of giftedness such as Gagné's (2017) integrative model of talent development, Renzulli's (1978) three-ring conception of giftedness, and Sternberg's (2003) wisdom, intelligence, and creativity (WIC) model. Thus, the inclusion of measures of creative potential is described as a necessary part of the identification process, even despite controversy over agreed upon construct validity and reliability.

Conceptions of creativity related to domain-general or domain-specific creativity. For instance, domain-specificity is discussed as part of the theoretical foundation of the DISCOVER, a performance-based assessment (Maker, 2001; Maker, 2020; Maker & Zimmerman, 2020; Sarouphim, 2000, 2004, 2009; Sarouphim & Maker, 2010; Wu et al., 2019). Likewise, conceptions were also related to Rhodes' (1961) 4P Model of Creativity (i.e., person, process, product, press; Kim et al., 2009; Mann, 2009), whereby children could be identified based on a product or performance, creative processes (e.g., divergent thinking, problem finding or solving), creative behaviors or personality characteristics. As for creative press, students could be identified based on the type of environment that encouraged creative engagement within the program or curriculum.

2.4.2.1 Type of Creative Assessment

Authors included the use of alternative assessments that measured creativity or some aspect of creativity, such as the use of authentic performance-based assessments (e.g., Sarouphim, 2000, 2004, 2009; Romanoff et al., 2009; VanTassel-Baska et al., 2007; Zhabonva et al., 2015), portfolios (Harris et al., 2009), and peer nominations (e.g., Cunningham et al., 1998; Zhabonva et al., 2015). Authentic performance-based assessments were often domain-specific (e.g., Science, Technology, Engineering, Mathematics [STEM]; Maker, 2020; Wu et al., 2019). Other alternative assessments incorporated creative measures that assessed divergent thinking (Chambers, 1980; LaFrance, 1995), problem-solving (Reid et al., 2000; Romanoff et al., 2009; Zhbanova et al., 2015), creative behaviors or components associated with the creative personality (e.g., Anderson, 2020; Cunningham et al., 1998; Pfeiffer et al., 2007). Rating scales/checklists largely identified creative behaviors or personality (e.g., risk-taking, humor, non-conformity) for peers, teachers, and parents to assess within the identification process (Cunningham et al., 1998; Peters & Pereira, 2017; Pfeiffer et al., 2007; Pfeiffer & Jarosewich, 2007), whereas observations and performance assessments focused on problem-finding or solving (Anderson, 2020; Maker, 2020; Romanoff et al., 2009; VanTassel-Baska et al., 2007; Wu et al., 2019). See Table 2.6 for a list of the creativity assessments used within the articles and their associated conception of creativity.

2.4.2.2 Domain-General versus Domain-Specific Creativity in Gifted Identification

Creativity was conceived as either domain-general or as domain-specific (i.e., content or task) and provided implications for these distinctions in relation to gifted identification. Wu et al. (2019) argued that developing exceptional talent requires the "development of domain-specific, integrated knowledge structure but also the development of domain-general, creative problem-solving abilities" (p. 479).

When considering domain-generality and specificity within creativity, domain-general approaches mainly focus upon either creative behaviors or personality characteristics (Kim et al., 2009; Pfeiffer et al., 2007), but also include general problem-solving abilities (Romanoff et al., 2009). For example, Kim et al. (2009) administered a self-questionnaire on implicit theories of giftedness including items like "I have more ideas than my peers", "I am imaginative", or "I have original thinking" (p. 106). This was also evident within items associated with creativity on multiple forms of the Gifted Rating Scales and the Scales for Rating the Behavioral Characteristics of Superior Students (SRBCSS; Renzulli & Hartman, 1971). Domain-generality broadly defined and generalized creativity within multiple content areas (Karadağ et al., 2016; Li et al., 2009; Peters & Pereira, 2017; Pfeiffer et al., 2007; Pfeiffer & Jarosewich, 2007; Rosado et al., 2008), whereas domain-specific approaches were more pinpointed to specific content or tasks (e.g., Diener et al., 2014; Maker, 2020; Maker & Zimmerman, 2020; Mann, 2009; Rostan et al., 2009; Sarouphim, 2000, 2004, 2009; Sarouphim & Maker, 2010; Zhbanova et al., 2015).

The majority of performance-based assessments within this review were more targeted for specific tasks within a domain. For example, Diener et al. (2014) assessed visual-spatial creativity in children with Autism Spectrum Disorder by administering a Modified Creativity Assessment Packet for experts to give ratings for 3D modeling projects. The DISCOVER (Maker, 2001) assessment is a performance-based assessment based on Gardner's (1983) theory of multiple intelligences and Maker's definition of giftedness that focuses on creative problem-solving abilities, defined as "the ability to solve the most complex problems in the most efficient, effective, or economical ways" (Maker, 1993, p. 71; as cited in Sarouphim, 2009, p. 278). The inclusion of problem-solving tasks within the DISCOVER assessment was used to examine "students' creative processes and products in domains (e.g., logical-mathematical, oral linguistic, spatial artistic, writing, spatial analytical; Sarouphim, 2009). This was further supported by other studies using the DISCOVER assessment or modified versions of the DISCOVER assessment that incorporate domain specific problem-solving tasks (Maker, 2020; Maker & Zimmerman, 2020; Sarouphim, 2000, 2009; Sarouphim & Maker, 2010; Wu et al., 2019).

2.4.2.3 Identifying Creativity as a Product or Performance

Included studies posited the usage of a number of performance-based instruments to identify diverse student populations as they provide an alternative perspective to assess student creative products or performances (Maker, 2020; Maker & Zimmerman, 2020; VanTassel-Baska et al., 2007; Wu et al., 2019). Performance tasks came in the form of drawing tasks (Rostan, 2005), imaginative writing tasks (Dewing, 1970), creation of 3D models (Diener et al., 2014), development of portfolios (Harris et al., 2009), general problem-solving tasks (Reid et al., 2000; Romanoff et al., 2009; VanTassel-Baska et al., 2007), and problem-solving tasks within specific domains (Maker, 2020; Maker & Zimmerman, 2020). Throughout this review, authors continue to advocate for usage of authentic performance-based assessment to more equitably identify students from different racial/ethnic backgrounds, genders, socioeconomic statuses, students with disabilities, and those who are multilingual for specific domains (Oreck et al., 2003; Romanoff et al., 2009; Sarouphim, 2000, 2009; Sarouphim & Maker, 2010; VanTassel-Baska et al., 2007).

There were numerous performance-based assessments whereby raters had to judge (give a rating to) the final products of students. For instance, the DISCOVER assessment (Maker, 2001) consists of five performance-based activities influenced by MI intelligence theory (e.g., spatial, spatial/logical-mathematical, logical-mathematical, oral linguistic, written linguistic; Sarouphim, 2000, 2004, 2009; Sarouphim & Maker, 2010), where participants engage in a series of tasks that progress from structured to unstructured (Sarouphim & Maker, 2010). To assess the products and processes associated with artistic talent, Rostan (2005) used Amabile's (1996) componential models of the creative process that emphasizes the "interaction of domain-relevant skills (e.g., knowledge, technical skills, and talents relevant to the domain), task motivation (i.e., intrinsic and extrinsic motivation to engage in a task), and creativity-relevant skills (e.g., problem-finding)" (p. 239) and used the Consensual Assessment Technique (Amabile, 1996) to assess drawing tasks (life drawings, imagination drawings). Similarly, Diener et al. (2014) used a performance assessment of visual-spatial ability using the design program, SketchUp, where students created 3D models that were later assessed by Google Experts. Within the performing arts, Oreck et al. (2003) used a performance assessment called the Talent Assessment Process in Dance, Music, and Theater (D/M/T TAP). The D/M/T TAP was "designed to assess systematically the artistic talents of all students...and to provide empirical data for designation of students as gifted and talented" (p. 71). Oreck et al. found that students identified through the D/M/T TAP, "accurately represented the demographics of the schools, including students in self-contained special education and bilingual classrooms" (p. 81).

Performance-based assessments are often criticized "for their high-cost, time-consuming procedures; domain underrepresentation; and, mostly, their lack of psychometrically sound qualities" (Sarouphim, 2000, p. 2). Despite these obstacles, authors propose the versatility and benefit of using performance-based assessments to more adequately assess what students can do outside of traditional paper-and-pencil test procedures. Authors suggest more studies are needed to determine the reliability of scoring with multiple raters, as well as the validity of the performance-based assessments across various student demographics (Diener et al., 2014; Maker, 2020).

2.4.2.4 Identifying Creative Processes: Divergent Thinking & Problem-finding/Problem-solving

Many of the articles conceptualized divergent thinking (Chambers, 1980; LaFrance, 1995; Subhi, 1997) and problem-solving (Anderson, 2020; Maker, 2020; Maker & Zimmerman, 2020; Romanoff et al., 2009; Zhbanova et al., 2015) as synonymous with creativity and utilized measures associated with each to identify creative potential.

Divergent thinking. Divergent thinking was synonymous with creativity within articles describing its role within gifted identification. Freund and Holling (2008) utilized the Berlin Intelligence Structure Model (Jâgar, 1984), a hierarchical model that synthesizes intelligence and creativity tasks similar to Guilford's Structure of Intellect Model (Guilford, 1967). As one operational factor of the model, creativity is defined as "fluid, flexible, and original production of ideas, requiring diverse information, wealth of imagination, and ability to see many different sides, variations, reasons, and possibilities in problem-oriented-not purely imaginative-solutions" (p. 311). Using the Battery for Giftedness Assessment, Nakano et al. (2016) explored "if assessing different domains adds additional information in predicting areas of giftedness" (p. 634) and found artistically gifted students scored higher on figural divergent thinking tasks than academically gifted students. Additionally, they found a strong relationship between divergent production of metaphors with fluid reasoning. The Battery for Giftedness assesses fluid intelligence, metaphor production, figural fluency, and divergent thinking figural task quality, and differentiates between two domains of giftedness, including academic and productive-creative or artistic giftedness (Nakano et al., 2016).

Other studies included the Torrance Test of Creative Thinking (LaFrance, 1995; Subhi, 1997) and the Wallach-Kogan Tests of Creativity (Chamber, 1980) to assess divergent thinking abilities. The figural form of the Torrance Test of Creative Thinking (Torrance, 2017) measures individuals on five subscales, these are fluency, originality, elaboration, abstractness of titles, resistance to premature closure, and includes a criterion referenced creative strengths checklist. Similarly, the Wallach-Kogan Test of Creativity (Wallach & Kogan, 1965) measures individuals on four subscales, originality, fluency, flexibility, and elaboration. The popularity of assessments such as the Torrance Test of Creative Thinking is well-documented (Kaufman et al., 2008); however, only a few studies were found to incorporate measures of group differences beyond age-level and grade to be included within this review.

Creative problem-finding and problem-solving. Many studies described the importance of creative problem-finding and problem-solving within the creative process. Csikszentmihalyi (1975) advocated for how problem-finding is an important component of creativity (as cited in Mann, 2009). Rostan (2005) conveyed that:

> expertise acquired through mastery of domain-relevant skills, problem-finding in varied contexts, interactive dialogue with a professional geared to the individuals child development, and the luxury of time, resources and support in the execution of purposeful work can prepare young artists for novel skill applications. (p. 258)

The importance of problem-finding or problem-solving in creativity assessments used in gifted identification was reiterated through a number of studies (Romanoff et al., 2009; Sarouphim &

Maker, 2010). Creative problem-solving was described in 19 articles. Maker (2020) and other authors suggest that creativity and intelligence are intertwined. This is particularly evident within measures that incorporate problem-solving as a component of fluid reasoning (Freund & Holling, 2008; Nakano et al., 2016) or intellectual ability (Pfeiffer et al., 2007; Pfeiffer & Jarosewich, 2007; Rosado et al., 2008). More broadly, authors describe the needs to assess creative-problem solving skills that are vital to the creative process (Rostan, 2005) and longer-term investments in developing domain-specific talent (Maker, 2020). Specifically, Maker (2020) propagates how creative problem-solving is "essential for innovation in STEM" (p. 175)

Van-Tassel Baska et al. (2007) and Romanoff et al. (2009) express the benefit of using performance assessments that focus on creative problem-solving to more equitable identify diverse student groups. Several studies used performance-based assessments that were based on problem-solving tasks within specific domains (Maker, 2020; Maker & Zimmerman, 2020; Sarouphim, 2000, 2004, 2009; Wu et al., 2020). For instance, Romanoff et al. (2009) evaluated a gifted program, where second grade teachers made referrals for testing for gifted services based on high achievement in reading and math in combination with work samples from "creative, hands-on, open-ended pre-assessment lessons in linguistics, logical-mathematical, and spatial problem-solving" activities (Romanoff et al., 2009, p. 160). After referral, students were formally tested with a Problem-Solving Assessment. Within their longitudinal study, they found evidence that students placed in gifted services based on the Problem Solving Assessment scored significantly higher on their "end-of-grade tests in reading and math" (p. 170) than those not selected. Conversely, Freund and Holling (2008) found reasoning ability (operationalized to include problem-solving) and creativity (i.e., divergent thinking) to be inversely related to grade point average; In other words, the higher the reasoning ability and creativity scores, the lower the grade point average, indicating further evaluation needed within classrooms on the relationship between creativity and other criterions related to achievement. As another solution, Reid et al. (2000) evaluated an alternative screening procedure that utilized a problem-solving assessment and found the procedure yielded a more diverse population of students identified for gifted services.

2.4.2.5 Identifying Creative People: Creative Behaviors or Personality Characteristics

Creative behaviors or creative personality characteristics were used within rating scales or checklists to describe students (Cunningham et al., 1998; Elliot et al., 1986; Harty et al., 1984; Karadağ et al., 2016; Labâr & Frumos, 2013; Li et al., 2009; Mann, 2009; Peters & Pereira, 2017; Pfeiffer et al., 2007; Pfeiffer & Jarosewich, 2007; Rosado et al., 2008; Subhi, 1997; Zhbanova et al., 2015). The studies included the Scales for Rating Behavioral Characteristics of Superior Students (SRBCSS; Renzulli & Hartman, 1971), the Scales for Identifying Gifted Students (Ryser & McConnell, 2004), the Gifted Rating Scales (Pfeiffer & Jarosewich, 2003), Frasier's Traits, Aptitudes, and Behaviors (Frasier, 1997), Udall's peer referral form (Udall, 1987), and a Sociogram survey based on the Alpha Project Peer Nomination Simulation (Renzulli et al., 1981). These ratings scales/checklists asked teachers, parents, or peers questions such as if students have "an adventurous spirit or a willingness to take risks" (e.g., SRBCSS; Renzulli et al., 2002), or if their peers know "what girl or boy is really good at making up games" (Cunningham et al., 1998, p. 200). Specifically, these assessments are designed to understand the person and the perspectives of others around them to triangulate gifted behaviors or characteristics that incorporate creativity.

2.4.2.6 The Creative Press: Environments that Encourage Creative Potential

Beyond the person, process, or product, creativity can flourish in specific environments.

Western societies tend to place considerable emphasis on how environments affect creativity, comparable to other parts of the world that emphasize moral reasons to engage in creativity (e.g., Korea; Kim et al., 2009). Maker and Zimmerman (2020) suggest knowing the context (i.e., the factors in the environment) is essential to designing and interpreting measures of developing expertise" (p. 259). Anderson (2020) stresses the importance of learning environments for Black girls, especially the need for "spaces of affirmation, anti-racist policies" (p. 97) and how adaptations to the curriculum can bolster talent development of CLED students. Freund and Holling (2008) discuss the need for evaluation of specific classroom environments beyond German schools, and specifically the need to evaluate teacher attitudes toward creativity and use of creative pedagogy, and how that can impact scholastic achievement. Moreover, Diener et al (2014) discusses the need for creativity measures for authentic learning environments, specific to real-life contexts, especially for students with Autism Spectrum Disorder. Using Renzulli's Enrichment Triad Model (Renzulli, 2012), Zhbanova et al. (2015) show how Type II enrichment activities aimed at skill development could help identify leadership and creative skills within Black/African American students.

Student engagement within STEM programs can help to enhance their interest in future careers and develop creative problem-solving skills related to their domains of interest. Wu et al. (2019) reported that exceptionally talented students engaged in the Keep Engaging Youth in Science summer internship program, were inspired and motivated by their engagement in problem-solving activities. Engagement within digital environments was discussed, such as using SketchUp to develop 3D models (Diener et al., 2014).

2.4.3 Research Focused on Validity of Creative Measures

2.4.3.1 Content, Construct, and Criterion Validity of Creative Measures

There was an abundance of validation studies of commonly used measurements in gifted identification processes. Of the article included, they each discussed validity in terms of content validity (Karadağ et al., 2016; Oreck et al., 2003; Subhi, 1997), construct validity (e.g., convergent, discriminant; Cunningham et al., 1998; Karadağ et al., 2016; Labâr & Frumos, 2013; Li et al., 2009; Peters & Periera, 2017; Pfeiffer et al., 2007; Sarouphim, 2007), and criterion validity (e.g., predictive, concurrent; Chambers, 1980; Freund & Holling, 2008; Harty et al., 2001; Karadağ et al., 2016; Labâr & Frumos, 2013; Pfeiffer et al., 2007; Sarouphim, 2009). The main concern of the included studies was determining how rating scales and measures of creativity related to other assessments and how equitable they were in identifying different subgroups (e.g., age, grade level, gender) and particularly underserved populations in gifted education.

Most commonly, researchers studied construct and criterion validity of creative measures (e.g., Harty et al., 1984; Peters & Pereira, 2017; Pfeiffer & Jarosewich, 2007; Sarouphim, 2000). Three measures were frequently used across the reviewed studies: The Gifted Rating Scales (i.e., School form or Preschool/Kindergarten form; Pfeiffer & Jarosowich, 2003), SRBCSS (Renzulli & Hartman, 1971), and the DISCOVER assessment (Maker, 2001). Pfeiffer and Jarosewich (2007) studied the construct validity and criterion validity of the Gifted Rating Scales-School Form. The authors found evidence of a six factor solution (i.e., Intellectual, Academic, Creativity, Artistic, Leadership, Motivation) and criterion-related validity for identifying traditionally underrepresented students using the Gifted Rating Scales. Similarly, Pfieffer et al. (2007), Karadağ et al. (2016), and Li et al. (2009) ran confirmatory factor analyses to assess the factor structure of the Gifted Rating Scales (School Form or Preschool/Kindergarten Form; Pfeiffer & Jarosewich, 2003) and each reported strong evidence of construct validity, but differences in criterion validity were apparent (e.g., creativity subscale showed a discrepancy in race/ethnicity scores; Pfeiffer et al., 2007). Peters and Pereira (2017) conducted a replication and extension to Pfeffier and Jarosewich's (2007) study and found poor fit and variance across groups of race/ethnicity and income, suggesting additional revisions to the scale are needed.

Elliot et al. (1986) assessed the predictive validity of the SRBCSS (Renzulli & Hartman, 1971) to future scholastic achievement (IQ or achievement tests) based on Anglo mid-high socioeconomic status, Anglo low socioeconomic status, and Hispanic low socioeconomic status and found minimal evidence. Harty et al. (2001) examined the content and criterion-related validity when using SRBCSS (Renzulli & Hartman, 1971) within a battery of other assessments and found evidence that dimensions of learning and creativity were sensitive and discriminated across three treatment groups (special-gifted-school classes, self-contained-regular classes, and a randomized control group). Likewise, with an Romanian sample, Labâr and Frumos (2013) studied 7th and 10th graders who participated in national and regional academic olympics compared to students who had not participated (or were in a local phase of academic competitions) on scores on the SRBCSS (Renzulli & Hartman, 1971). They found evidence of construct validity (i.e., convergent and discriminant validity) on measures of the SRBCSS when comparing to a Needs Assessment Questionnaire that measures achievement, affiliation, autonomy, and dominance. Additionally, they found evidence of criterion validity (i.e., predictive validity) in that Olympic students scored significantly higher on dimensions of the SRBCSS than non-Olympic students.

There were several studies that analyzed the DISCOVER (e.g., Sarouphim, 2000, 2004, 2009; Sarouphim & Maker, 2010). The main concern of the studies were issues of criterion validity, rather than construct validity. Sarouphim (2000) claims to have assessed the construct validity of the DISCOVER (Maker, 2001) in relation to the theory of multiple intelligences (see Gardner, 1983), and reported Spearman Rho correlations instead of more advanced statistical analyses. Sarouphim (2004), Sarouphim (2009), and Sarouphim and Maker (2010) extended the argument by assessing the criterion validity of the DISCOVER in relation to multiple intelligences, gender, and ethnic differences. Overall, in all DISCOVER studies included, they found evidence of increased diversity and that the percentage of students identified as gifted was higher than students identified through traditional standardized tests.

2.4.3.2 Group Differences on Creative Measures

2.4.3.2.1 Race/Ethnicity and Socioeconomic Status

Using problem-solving assessment tasks, studies found that students from disadvantaged backgrounds were more likely to be identified for gifted services (Maker, 2020; Romanoff et al., 2009; Sarouphim, 2000; VanTassel-Baska et al., 2007). Using the DISCOVER (Maker, 2001), Sarouphim (2004) found no significant main effect for ethnicity, nor an interaction effect between gender and ethnicity. Sarouphim and Maker (2010) found a "high percentage of South Pacific/Pacific Islanders identified (37%)" (p. 52) using the DISCOVER that needs to be further evaluated; Although there were other ethnic differences found in identifying South Pacific/Pacific Islanders, all other ethnic groups were well-represented, and more minority students were identified (3% more).

Similarly, using the SRBCSS (Renzulli & Hartman, 1971), Harty et al. (1984) posited that the subscales of learning and creativity should be used within formal identification processes (with appropriate weights) since it was found to be "one of the better discriminators of giftedness and one of the least discriminatory among economically disadvantaged and/or minority students" (p. 341). This was partially supported by Elliot et al. (1986), whose findings suggest there may be predictive value for Hispanic students using the SRBCSS on the creativity scale. Specifically, the authors found "scores on the creativity scale of the SRBCSS accounted for 54% of the variance in the performance of Hispanic third- and fourth-grade students on the SAT Reading and Comprehension scale" (p. 32); however, this was a small sample (*n* = 11).

Pfeiffer and Jarosewich (2007) did not find any significant differences for race/ethnicity on any of the GRS-S scales with the standardization sample. However, they did find "slightly higher mean scores from Asian American and White students when compared to African American and Hispanic students" (p. 47). This was echoed with Pfeiffer et al. (2007) when using the Gifted Rating Scales-Preschool form and found Caucasian/White students yielded higher scores on the creativity subscale than Hispanic and African American students. Even in the face of these differences, the authors still suggest their overall findings indicate that both the Gifted Rating Scales-School and Preschool form would be useful tools within gifted identification processes. Similarly, Li et al. (2009) found good fit for the Gifted Rating Scales across cultural groups (i.e., China, Puerto Rico, South Korea, Turkey, United States). However, Li et al. did find evidence of item-level variance on the creativity, motivation, academic, and artistic subscales, meaning item loadings for these scales were inconsistent across all locations. Additionally, Peters and Pereira (2017) analyzed the Scales for Identifying Gifted Students (Ryser & McConnell, 2004), Gifted Rating Scales (Pfeiffer & Jarosewich, 2003), and the HOPE Scale (Gentry et al., 2015) and found evidence of poor fit using multigroup confirmatory factor analysis. Specifically, the Scales for Identifying Gifted Students and Gifted Rating Scales did not have similar fit across groups; although, the Gifted Rating Scales showed better fit for underrepresented students. In regard to socioeconomic status, specifically with the Scales for Identifying Gifted Students and Gifted Rating Scales, low-income students fit the "instruments' models better than high-income students" (p. 115). Unfortunately, each instrument showed problematic fit across all groups and further evaluation is warranted.

2.4.3.2.2 Gender/Sex

Included studies conceptualized gender as a binary conception or biological sex (i.e., male and female). Karadağ et al. (2016) found that "subscale scores for the intellectual ability, academic ability and creativity subscales did not differ across genders, where significant differences were obtained in the artistic ability and motivation subscales on behalf of girls" on the Turkish version of the Preschool/Kindergarten form of the Gifted Rating Scales (GRS-P; p. 14). Likewise, Labâr and Frumos (2013) studied a Romanian sample and found no gender differences on any of the dimensions of the SRBCSS. Subhi (1997) reports no gender/sex differences between all scores used within a computerized identification procedure (i.e., peer nomination, Raven's Progressive Matrices, Mathematical Skills Assessment, Torrance Test of Creative Thinking Verbal & Figual, SRBCSS). Likewise, Dewing (1970) found no sex differences were found within creative ability or performance.

Conversely, Pfeiffer and Jarosewich (2007) did find modest gender differences on all six subscales of Gifted Rating Scale-School form, with girls scoring higher than boys. There were mixed evidence found in studies using the DISCOVER assessment (Maker, 2001); Sarouphim and Maker (2010) and Sarouphim (2009) found no significant gender differences on the DISCOVER, however boys scored higher than girls; caution should be taken because of the low sample size (*n* = 36; Sarouphim, 2009).

2.4.3.2.3 Twice-Exceptional Students

LaFrance (1995) found that twice exceptional children (i.e., children who are gifted and learning disabled within this study) scored higher than other students on emotional expressiveness and were similar to gifted students in terms of a number of areas on the creative strengths checklist, specifically on storytelling articulateness, expressiveness of titles, synthesizing circles, and synthesizing incomplete figures. Diener et al. (2014) developed a creativity measure that could identify strengths of youth with Autism Spectrum Disorder. Using the Creativity Assessment Packet, researchers compared their scores on student products (i.e., 3D models) from (i.e., subscales on fluency, flexibility, originality, and elaboration; Williams, 1980) with ratings of Google Experts. They found high correlations on all subscales, however scores of originality from the Google Experts and the researchers were not significantly related.

2.5 Discussion

The purpose of this study was to examine the existing empirical literature on creativity measures in the gifted identification process and the relation to equitable access. Overcoming obstacles within identification systems to identify domain-specific talent was a central theme that permeated throughout many articles. Through the paradigmatic lens of talent development, domain-specific creativity measures within the gifted identification process were often attributed as providing greater access for students of different abilities and backgrounds. However, there were mixed findings on criterion validity in relation equitable access to gifted programs. Of all creativity assessments used, performance-based assessments using problem-based tasks showed promising performance for use within gifted identification (Diener et al., 2014; Maker, 2020; Maker & Zimmerman, 2020; Rostan, 2005; VanTassel-Baska et al., 2007) and connected to longer term achievement (e.g., Romanoff et al., 2009). The status of empirical research on creative measures largely centered on the validity of assessments. Recommendations for usage of creativity-based assessments and inclusive practices are discussed below.

2.5.1 Overcoming Obstacles in Identification Systems

In regard to the influence of the Marland Report over time, creativity is a single characteristic that has been integrated into several theories of giftedness and has encouraged a wider conception of what to assess in identification systems. When contemplating the criteria in an identification system, there are multiple factors to consider. This is inclusive of how to identify creativity and what creativity assessments are used in regard to their sensitivity of identifying students and, more importantly, the decisions of their usage within gifted identification processes, and how these decisions help diversify the student populations within gifted programs.

Critical evaluation of our current gifted identification systems is necessary if we want to provide more equitable services for our diverse student populations. In order for leaders in gifted education to pursue equity and excellence in gifted programs, this requires a balance of providing equitable access to gifted programs and ensure the quality that students are receiving within the services of which they were identified. This necessitates a concerted effort with stakeholders to identify the goals of gifted programs, how they identify students for specified services, and how to measure student success. For instance, Treffinger (2004) suggests that "creativity assessment data should be used diagnostically, to guide instructional planning, rather than merely as criteria for selecting students" (p. 89); Thus, there should be a match in identification assessments to the services provided. Creativity should not just be assessed in the identification process but should be integrated into gifted programs through classroom environments conducive to creative productivity (Lamb, 2020; Lee et al., 2021; Nakano et al., 2016). A closer examination of multiple criteria systems is needed to better understand this alignment issue, as well as how students are nominated (McBee et al., 2014; Lee et al., 2020) and how their creativity scores are combined for decisions on placement in gifted programs (Lakin, 2018; McBee et al., 2016).

2.5.1.1 Using Domain-Specific Creativity to Identify Talent

Scholars in gifted education have propagated how gifted programs should find methods to better locate students with domain-specific talent (Olszewski-Kubilius et al., 2017; Subotnik et al., 2011). This also aligns with creativity literature that suggests a greater focus on domain-specific creativity (Baer, 2016; Diener et al., 2014; Rostan et al., 2009; Maker, 2020). Common conceptions include viewing creativity as a domain-general phenomenon that can be applied to a wide range of topics, whereas other research has suggested evidence of domain-specificity (Baer, 2016).

Many scholars believe tasks within tests, like the Torrance Test of Creative Thinking, measure an aspect of creativity (i.e., divergent thinking) and more broadly how creativity is domain-general (e.g., creative thinking skills). Others have suggested the usage of more domain-specific methods in the assessment of creativity (e.g., portfolios, performance-based assessments, Consensual Assessment Technique; Amabile, 1996; Baer, 2016; Diener et al., 2014; Rostan, 2005; Romanoff et al., 2009; VanTassel-Baska et al., 2007). Dai (2010) suggests that individuals lie on a continuum of domain experiences that range from "universal to unique" (p. 89), whereby some experiences are universally developed (e.g., logic, reasoning) and moves to more unique specializations (patent lawyer, athletes, phlebotomy).

For students to be globally competitive on the job market, schools need to identify talent in an array of domains (e.g., STEM disciplines) and provide opportunities to develop talent through effective programs. Barab and Plucker (2002) suggest talents to be developed by "the interaction of the individual, environment, and sociocultural content" (as cited in Plucker &

Callahan, 2014, p. 392) and how gifted programs should help students solve authentic problems to develop their "full potential as creative, real-world problem solvers" (p. 392). For example, in Finland, this is seen within systematic educational reforms that have focused on developing a culture of "diversity, trust, respect within Finnish society" (Darling-Hammond, 2010; p. 168) and refined their national curriculum to be more locally developed with a focus on "science, technology, and innovation, leading to an emphasis on teaching students how to think creatively and manage their own learning"(p. 169). Specifically, the Finnish school system integrated creative problem-solving and cross-curricular projects to connect to authentic learning experiences for career development (rather than accountability testing, as seen in the United States; Hodges, 2018). Similarly, from this review of the literature, the majority of authors suggest to utilize domain-specific measures that incorporate more context-specific problem-solving, authentic performance-based assessments within the identification process whenever possible to more equitably identify talent.

2.5.1.2 What Creativity Measures Should Be Used?

There is no clear-cut answer for which creativity measure is the correct one to use. Whatever measure is chosen should align with the goals of the specified program. Moreover, this coincides with the state versus trait argument of gifted identification. This is dependent on if the program goals align best with identifying traits of the creative personality or specific states involved in the creative process and creative production. Overall, the development of learning environments (i.e., programs, curriculum, classrooms) to engage in creative processes (i.e., divergent thinking, problem-solving) for the creation of domain-specific products were reiterated throughout many of the included studies (Harris et al., 2009; Maker, 2020; Maker & Zimmerman, 2020; Romanoff et al., 2009; Sarouphim, 2000, 2004, 2009; VanTassel-Baska et al., 2007; Wu et al., 2019; Zhbanova et al., 2015).

Scholars have suggested that performance-based assessments show evidence for domain-specific creativity, whereas other assessments like divergent thinking have shown evidence for domain-general creativity (e.g., fluency, originality, elaboration, flexibility); although, keep in mind, Plucker (2011) found evidence that verbal divergent thinking tests as more predictive of creative achievement. Similarly, context-dependent creativity was a driving force for performance-based assessments. In other words, creativity assessments that allowed demonstration of creative skills in specific areas of expertise (i.e., math, science, technology) were strongly suggested within this review of the literature (Diener et al., 2014; Maker, 2020; Rostan, 2005; Zhbanova et al., 2015). Performance-based assessments that utilize problem-solving tasks within specific domains showed promise in more equitably identifying CLED students (Maker, 2020; Maker & Zimmerman, 2020; Romanoff et al., 2009; VanTassel-Baska et al., 2007).

2.5.2 Validity of Creativity Measures and Equitable Opportunities

Validity of measures of creativity continue to be a concern for identification systems, especially in relation to how they encourage equitable access for diverse populations to opportunities in gifted programs. Although, authentic domain-specific performance-based assessments were overtly encouraged within this review, caution should be used when using them because there is limited information over their construct validity, as well as the costly and time-consuming nature of implementation. However, studies report promising results for equitable access when using performance-based assessments and there should be further evaluation in the future.

Better understanding the predictive validity of creative measures can help us understand longer term outcomes for the talent development process. Divergent thinking has been studied in terms of predictive validity; specifically, Plucker (2011) found verbal divergent thinking to be strongly related with creative achievement (as cited in Nakano et al., 2016). Although divergent thinking assessments have been the most popularly used assessments of creativity within gifted education (Hunsaker & Callahan, 1995), none of the articles included in this review discussed the relation of divergent thinking tests to equitable access beyond gender/sex, grade, or age level. Articles excluded that used divergent thinking tests did include cross-sectional validation studies of the Torrance Test of Creative Thinking and Wallach-Kogan assessments, but they did not include a measure of group differences beyond age or grade-level. Specifically, more research is needed to determine group differences of performance on divergent thinking assessments used in gifted identification processes (e.g., race/ethnicity, socioeconomic status, English language learners, specific diagnoses of twice exceptionality).

In regard to behavioral rating scales, only two studies used multigroup confirmatory factor analysis and measurement invariance (Li et al., 2009; Peters & Pereira, 2017), showing a greater need for studies to focus on measurement invariance or differential item functioning specific to group differences (i.e., race/ethnicity, socioeconomic status, twice-exceptionality) for creativity measures used in gifted identification (Peters, 2016).

A notable observation through this analysis was that many of the validation studies were conducted by creators of the instruments. More replications and validation studies should be conducted by researchers unaffiliated with the original creators of the instruments to provide additional evidence (Smith et al., 2017).

2.5.3 Recommendations for Creativity Assessments and Inclusive Gifted Identification

There is no silver bullet for identifying talent (Callahan et al., 2005). Similar sentiments are echoed by Makel and Johnsen (2020) who caution administrators and researchers that there is no one set criteria that will solve problems with identification. However, providing alternative pathways to identification can widen consideration. Alternative assessments offer another avenue for students to showcase their strengths, specifically their potential for creative productivity or performance (VanTassel-Baska et al., 2007). Even with alternative routes in place, issues of underrepresentation remain. Multiple criteria systems need further critical evaluation of how criteria are being implemented, particularly with the usage of whatever creativity assessment within the identification process. For instance, Torrance (1974) cautioned the usage of composite scores in decision-making processes (or a single score that exemplifies creativity), as they do not represent an individuals' creative strengths, even so, composite scores are often used in gifted identification (Kim, 2006).

Continued demographic disproportionality requires systematic exploration of multiple facets beyond which measurements are used, such as monetary costs of tests, funding disparities, policy evaluation, teacher preparation, professional development, program goals and outcomes, other aspects that contribute to the identification process (Hodges et al., 2021; Peters, 2021).

No matter what creativity assessment that is used within gifted identification, the definition of creativity assigned by the district and the creativity assessment should align (Callahan et al., 1995). Likewise, suggestions posed by Gubbins et al. (2021), Lee et al. (2020), and Peters et al., (2020) advocate that the goals of gifted programs should match their selection process and how those students are served within the classroom; this includes goals that relate to creativity (Baer, 2013).

There needs to be purposeful alignment between goals of the program, what is used in the selection process, and the actual expectations of the program. For example, if creativity is a goal, then a creativity assessment should be used to assess creativity that will match with the program offered (Peters et al., 2020). School districts that have more knowledge of how these decisions in their multiple criteria systems are possibly impacting their program diversity can re-adjust expectations and policies to incorporate more equitable practices and have more populations that are reflective of their student demographics (i.e., proportionality). In future research, qualitative and mixed methods analyses are needed to assess how performance-based creativity assessments are being judged and incorporated into a gifted identification process (i.e., how committees are deciding). This includes if the consensual assessment technique is being utilized in evaluation from 'expert evaluators' and how they impact multiple criteria systems. Based on findings from this review, specific recommendations for creativity assessments and inclusive gifted identification practices are as follows:

- Align program goals of developing creativity with assessments that measure creativity, especially if creative production or performance are designated program outcomes.
- Use the most reliable and valid measures for creativity
- Continue to use multiple criteria to provide multiple pathways to be identified, but exercise caution in how scores are combined.
- Educate teachers and parents about how to find creative potential and adopt strength-based perspectives when identifying talent.
- Periodically evaluate identification methods to determine demographic representation.

Although these are the best practices, systems need to further evaluate what practices they currently use for identification and evaluate program diversity. To date, there has not been a study of how using a creativity assessment to universally consider all students for entry into programs. Often, creativity assessments usually take a secondary role within the identification process and are used in combination with cognitive ability and achievement tests. This warrants further research of how creativity assessments could be used for universal consideration and how processes differ based on the creativity assessments used within the process (i.e., portfolios, planned experiences; Mun et al., 2020b).

2.5.4 Limitations and Implications for Future Research

There are limitations that are noteworthy that should be addressed in future research regarding creativity and equitable access in gifted education. Specifically, the keywords and combinations chosen within this study could have limited relevant articles for inclusion. Likewise, the exclusion criteria that eliminated any articles that did not measure group differences or study an underrepresented group in gifted education could have limited the scope of articles. Although there was an abundance of validity studies outside of the United States, many of these cross-cultural studies were removed from this systematic review due to not measuring a group difference. Future research should continue to examine the validity of creativity assessments, but researchers interested in reviewing the literature should consider evaluating cross-cultural studies of creativity assessments and their relation to gifted identification. Additional meta-analytic techniques could be used to further understand validity, publication bias, and other moderating factors (e.g., group differences) across studies that use creativity assessments in relation to equitable access. More studies should focus on the relationship between special education, students with documented 504s for learning disabilities, creativity, and giftedness. Also, future studies should consider wider conceptions of gender in relation to creativity beyond binary conceptions of gender or biological sex.

Another limitation was that many studies included used a cross-sectional design. Future studies on creativity and equitable access should consider implementing other designs, such as longitudinal designs, experimental designs, exploratory designs within qualitative research, or mixed methods. Although there were several validation studies, only two studies explicitly report using either multigroup confirmatory factor analysis or measurement invariance with specified groups (Li et al., 2009; Peters & Pereira, 2017). There was limited information on the criterion validity of the Torrance Test of Creative Thinking with various demographic subgroups beyond age, grade-level, and gender. Additionally, the content and construct validity of task-oriented performance-based assessments and problem-solving assessments were limited within the included studies. Beyond the use of creativity-based alternative assessments in gifted identification processes, there were little to no recommendations of how to combine assessments that use creativity within the process (Lakin, 2018; McBee et al., 2014). Few authors addressed how assessments were combined within the identification process and how that related to program diversity (Peters & Pereira, 2017; VanTassel-Baska et al., 2007). Rather, the focus was upon how the usage of specific creative measures could provide greater equity and inclusivity and the identification process was not specified. Future research should address the usage of assessments that incorporate creativity in the identification process, and more specifically evaluate the relationship of creativity measures and combination rules used within the gifted identification process.

2.5.5 Conclusion

Systemic inequities have plagued gifted identification systems, largely due to achievement and excellence gaps that exacerbate underrepresentation of CLED students (Anderson, 2020; Harris et al., 2009; Romanoff et al., 2009; Worrell & Dixson, 2018). To increase more equitable access to gifted programs, expanded notions of giftedness and intelligence have helped to change federal and state policies in broadening conceptions of giftedness, and influenced the adoption of multiple criteria (Callahan & Plucker, 2014; Frasier, 1997; Marland Report, 1972;), specifically the use of creativity assessments, within the gifted identification process. Within this review, there were multiple studies with domain-specific, context-driven creativity measures and performance-based assessments that propagate more equitable access. However, no matter the creativity assessment used, there needs to be alignment between the definition of creativity assigned by the district, the designated program outcomes, and the creativity assessment used (Callahan et al., 1995; Lee et al., 2020). This knowledge has practical significance in that it can help school districts make more evidence-driven decisions and understand the research of creative assessments in the identification of talent.

2.6 References

*Studies included in the systematic review

Acar, S., Sen, S., & Cayirdag, N. (2016). Consistency of the performance and nonperformance methods in gifted identification: A multilevel meta-analytic review. *Gifted Child Quarterly, 60*(2), 81-101. https://doi.org/10.1177/0016986216634438

Anderson, B. N. (2020). “See me, See us”: Understanding the intersections and continued marginalization of adolescent gifted Black girls in U.S. Classrooms. *Gifted Child Today*, *43*(2), 86–100. *https://doi.org/10.1177/1076217519898216*

Amabile, T. M. (1996). *Creativity in context.* Westview Press.

*Anderson, B. N. (2020). “See me, see us”: Understanding the intersections and continued marginalization of adolescent gifted Black girls in U.S. classrooms. *Gifted Child Today*, *43*(2), 86–100. https://doi.org/10.1177/1076217519898216

Baer J. (2013) Teaching for creativity: Domains and divergent thinking, intrinsic motivation, and evaluation. In M. Gregerson, J. Kaufman, & H. Snyder (Eds.), *Teaching Creatively and Teaching Creativity*. Springer. https://doi.org/10.1007/978-1-4614-5185-3_13

Baer, J. (2016). *Domain specificity of creativity.* Elsevier.

Baldwin, A. Y. (2002). Culturally diverse students who are gifted. *Exceptionality, 10*(2), 139–147. https://doi.org/10.1207/S15327035EX1002_7

Barab, S. A., & Plucker, J. A. (2002). Smart people or smart contexts? Cognition, ability, and talent development in an age of situated approaches to knowing and learning. *Educational Psychologist, 37*(3), 165-182. *https://doi.org/10.1207/S15326985EP3703_3*

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Callahan, C. M., Tomlinson, C. A., Hunsaker, S. L., Bland, L. C., & Moon, T. (1995). *Instruments and evaluation designs used in gifted programs.* University of Connecticut, National Research Center on the Gifted and Talented.

Callahan, C. M. (2005). Identifying gifted students from underrepresented populations. *Theory Into Practice, 44*(2), 98-104. https://doi.org/10.1207/s15430421tip4402_4

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 of the United States of America, 113,* 13678-13683. https://doi.org/10.1073/pnas.160504311

Carman, C. A., Walther, C. A. P., & Bartsch, R. A. (2020). Differences in Using the Cognitive Abilities Test (CogAT) 7 Nonverbal Battery Versus the Naglieri Nonverbal Ability Test (NNAT) 2 to Identify the Gifted/Talented. *Gifted Child Quarterly, 64*(3), 171–191. https://doi.org/10.1177/0016986220921164

*Chambers, J. A., Barron, F., & Sprecher, J. W. (1980). Identifying gifted Mexican-American students. *Gifted Child Quarterly, 24*(3), 123-128. https://doi.org/10.1177/001698628002400306

Clarke, V., Braun, V., & Hayfield, N. (2015). Thematic analysis. In J. A. Smith (Ed.), *Qualitative psychology: A practical guide to research methods* (pp. 222-248). Sage Publications.

Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart, and Winston.

*Cunningham, C. M., Callahan, C. M., & Plucker, J. A. (1998). Identifying Hispanic students of outstanding talent: Psychometric integrity of a peer nomination form. *Exceptional Children*, *64*(2), 197–209. https://doi.org/10.1177/001440299806400204

Dai, D. Y. (2010). *The nature and nurture of giftedness: A new framework for understanding gifted education.* Teachers College Press.

Dai, D. Y., & Chen, F. (2014). *Paradigms of gifted education: A guide to theory-based practice-focused research.* Prufrock Press.

Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future.* Teachers College Press.

*Dewing, K. (1970). The reliability and validity of selected tests of creative thinking in a sample of seventh-grade West Australian children. *British Journal of Educational Psychology, 40*(1), 35-42. https://doi.org/10.1111/j.2044-8279.1970.tb02096.x

*Diener, M. L., Wright, C. A., Smith, K. N., & Wright, S. D. (2014). Assessing visual-spatial creativity in youth on the autism spectrum. *Creativity Research Journal, 26*(3), 328-337. *https://doi.org/10.1080/10400419.2014.929421*

*Elliott, S. N., Argulewicz, E. N., & Turco, T. L. (1986) Predictive validity of the scales for rating the behavioral characteristics of superior students for gifted children from three sociocultural groups, *The Journal of Experimental Education, 55*(1), 27-32, https://doi.org/10.1080/00220973.1986.10806431

Erwin, J. O., & Worrell, F. C. (2012). Assessment practices and the underrepresentation of minority students in gifted and talented education. *Journal of Psychoeducational Assessment, 30*(1), 74-87. *https://doi.org/10.1177/0743282911428197*

Ford, D. Y., & Grantham, T. C. (2003). Providing access for culturally diverse gifted students: From deficit to dynamic thinking. *Theory Into Practice, 42*(3), 217-225. *https://doi.org/10.1207/s15430421tip4203_8*

Ford, D. Y., Grantham, T. C., & Whiting, G. W. (2008). Culturally and linguistically diverse students in gifted education: Recruitment and retention issues. *Exceptional Children, 74*(3)*,* 289-306. *https://doi.org/10.1177/001440290807400302*

Frasier, M. M., García, J. H., & Passow, A. H. (1995). *A review of assessment issues in gifted education and their implications for identifying gifted minority students* (RM95204). Storrs: The National Research Center on the Gifted and Talented, University of Connecticut. Retrieved from http://nrcgt.uconn.edu/research-based_resources/frasgarc/

Frasier, M. M. (1997). Multiple criteria: The mandate and the challenge. *Roeper Review, 20*(2), A4-A6. https://doi.org/10.1080/02783199709553868

*Freund, P. A., & Holling, H. (2008). Creativity in the classroom: A multilevel analysis investigating the impact of creativity and reasoning ability on GPA. *Creativity Research Journal, 20*(3), 309–318. *https://doi.org/10.1080/10400410802278776*

Gagné, F. (2017). The integrative model of talent development (IMTD): From theory to educational applications. In J. A. Plucker, Rinn, A. N., & Makel, M. C. (Eds.),*From Giftedness to Gifted Education: Reflecting Theory in Practice* (pp. 149-182). Prufrock Press.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* Basic Books.

Gentry, M., Peters, S. J., Pereira, N. P., McIntosh, J., & Fugate, C. M. (2015). *HOPE Teacher Rating Scale: Involving teachers in equitable identification of gifted and talented students in K-12* [Measurement instrument]. Prufrock Press.

Getzels, J. W., & Csikszentmihalyi, M. (1976). The creative vision: A longitudinal study of problem finding in art. John Wiley & Sons.

Gubbins, E. J., Siegle, D., Ottone-Cross, K., McCoach, D. B., Langley, S. D., Callahan, C. M., Brodersen, A. V., & Caughey, M. (2021). Identifying and Serving Gifted and Talented Students: Are Identification and Services Connected? *Gifted Child Quarterly*, *65*(2), 115–131. *https://doi.org/10.1177/0016986220988308*

Guilford, J. P. (1967). Creativity: Yesterday, today, and tomorrow. *The Journal of Creative Behavior, 1*(1), 3–14. https://doi.org/10.1002/j.2162-6057.1967.tb00002.x

*Harris, B., Plucker, J. A., Rapp, K. E., & Martínez, R. S. (2009). Identifying gifted and talented English language learners: A case study. *Journal for the Education of the Gifted*, *32*(3), 368–393. https://doi.org/10.4219/jeg-2009-858

*Harty, H., Adkins, D. M., & Sherwood, R. D. (1984). Predictability of giftedness identification indices for two recognized approaches to elementary school gifted education. *Journal of Educational Research*, *77*(6), 337–342. https://doi.org/10.1080/00220671.1984.10885552

Heller, K., Perleth, C., & Lim, T. (2005). The Munich Model of Giftedness designed to identify and promote gifted students. In R. Sternberg & J. Davidson (Eds.), *Conceptions of Giftedness* (pp. 147-170). Cambridge University Press. https://doi.org/10.1017/CBO9780511610455.010

Hodges, J. (2018). Assessing the influence of No Child Left Behind on gifted education funding in Texas: A descriptive study. *Journal of Advanced Academics*, *29*(4), 321-342. https://doi.org/10.1177/1932202X18779343

Hodges, J., Crutchfield, K. C., & Lee, L. E. (2021). Equitable access, locale, funding, and the alignment to performance ratings in gifted programs. *Journal for the Education of the Gifted*, *44*(1), 3–30. https://doi.org/10.1177/0162353220978306

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M. P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M. C., & Vedel, I. (2018). Mixed methods appraisal tool (MMAT). Registration of Copyright (#1148552), Canadian Intellectual Property Office, Industry Canada.

Hunsaker, S. L., & Callahan, C. M. (1995). Creativity and giftedness: Published instrument uses and abuses. *Gifted Child Quarterly*, *39*(2), 110-114. *https://doi.org/10.1177/001698629503900207*

Jâger, A. O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven [Structural research on intelligence: Competing models, new developments, perspectives]. Psychologische Rundschau, 35, 21–35.

Jolly, J. L., & Robins, J. H. (2018). History of gifted education. In J. L. Roberts, T. F. Inman, & J. H. Robins (Eds.), *Introduction to Gifted Education* (pp. 17-30). Prufrock Press.

*Karadağ, F., Karabey, B. & Pfeiffer, S. (2016). Identifying gifted preschoolers in Turkey: The reliability and validity of the Turkish-translated version of the GRS-

Preschool/Kindergarten form. *Journal of Education and Training Studies*, *4*(10), 8–16. http://dx.doi.org/10.11114/jets.v4i10.1686

Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment.* John Wiley & Sons.

Kaufman, J. C., Plucker, J. A., & Russell, C. M. (2012). Identifying and assessing creativity as a component of giftedness. *Journal of Psychoeducational Assessment, 30*(1), 60-73. *https://doi.org/10.1177/0734282911428196*

Kim, K. H. (2006). Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT). *Creativity Research Journal, 18*(1), 3–14. *https://doi.org/10.1207/s15326934crj1801_2*

*Kim, K. H., Shim, J., & Hull, M. (2009). Korean concepts of giftedness and the self-perceived characteristics students selected for gifted programs. *Psychology of Aesthetics, Creativity, and the Arts, 3*(2), 104-111. https://doi.org/10.1037/a0013324

*Labâr, A. V., & Frumos, F. V. (2013). Adaptation and validation of the Scales for Rating the Behavioral Characteristics of Superior Students (SRBCSS) for self-assessment on a Romanian sample. *Scientific Annals of the UAIC Iasi: Educational Sciences Series, 17*(1), 103–116.

*LaFrance, E. B. (1995). Creative thinking differences in three groups of exceptional children as expressed through completion of figural forms. *Roeper Review*, *17*(4), 248. https://doi.org/10.1080/02783199509553675

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

Lakin, J. (2018). Making the cut in gifted selection: Score combination rules and their impact on program diversity. *Gifted Child Quarterly, 62*(2), 210-219. https://doi.org/10.1177/0016986217752099

Lamb, K. N. (2020). Cultivating creativity in the classroom. In J. L. Jolly & J. H. Robins (Eds.), *Methods & Materials for Teaching the Gifted* (pp. 279-297). Prufrock Press.

Lee, L. E., Meyer, M. S., & Crutchfield, K. C. (2021). Gifted classroom environments and the creative process: A systematic review. *Journal of the Education of the Gifted*.

Lee., L. E., Ottwein, J. K., & Peters, S. J. (2020). Eight universal truths of identifying students for advanced academic interventions. In J. L. Jolly & J. H. Robins (Eds.), *Methods & Materials for Teaching the Gifted* (pp. 61-79). Prufrock Press.

Lemons, G. (2011). Diverse perspectives of creativity testing: Controversial issues when used for inclusion into gifted programs. *Journal for the Education of the Gifted, 34*(5), 742-772. http://dx.doi.org/10.1177/0162353211417221

*Li, H., Lee, D., Pfeiffer, S. I., Kamata, A., Kumtepe, A. T., & Rosado, J. (2009). Measurement invariance of the Gifted Rating Scales—School Form across five cultural groups. *School Psychology Quarterly*, *24*(3), 186–198. https://doi.org/10.1037/a0017382

Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic inquiry.* Sage Publications.

Lo, C. O., & Porath, M. (2017). Paradigm shifts in gifted education: An examination vis-a-vis its historical situatedness and pedagogical sensibilities. *Gifted Child Quarterly, 61*(4), 343-360. https://doi.org/10.1177/0016986217722840

Lohman, D. F. (2009). Identifying academically talented students: Some general principles, two specific procedures. In L. V. Shavinina (Ed.), *International handbook on giftedness* (pp. 971–997). Springer.

Lohman, D. F. (2012). *Cognitive Abilities Test Form 7: Research and development guide*. Riverside Assessments, LLC.

Lohman, D. F., & Gambrell, J. (2012). Use of nonverbal measures in gifted identification. *Journal of Psychoeducational Assessment*, *30*(1), 25-44. https://doi.org/10.1177/0734282911428194

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P. A., … Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Journal of Clinical Epidemiology, 62*(10), e1-34. https://doi.org/10.1016/j.jclinepi.2009.06.006

Luria, S. R., O'Brien, R. L., & Kaufman, J. C. (2016). Creativity in gifted identification: Increasing accuracy and diversity. *Annals of the New York Academy of Sciences, 1377*(1), 44-52. *https://doi.org/10.1111/nyas.13136*

Makel, M. C., & Johnsen, S. K. (2020). Conceptions of giftedness. In J. A. Plucker & C. M. Callahan (Eds.), *Critical Issues and Practices in Gifted Education* (3rd ed., pp. 113-122). Prufrock Press.

Maker, C. J. (2001). DISCOVER: Assessing and developing problem solving. *Gifted Education International*, *15*(3), 232–251. https://doi.org/10.1177/026142940101500303

*Maker, C. J. (2020). Identifying exceptional talent in science, technology, engineering, and mathematics: Increasing diversity and assessing creative problem-solving. *Journal of Advanced Academics, 31*(3), 161-210. https://doi.org/10.1177/1932202X20918203

*Maker, C. J., & Zimmerman, R. H. (2020). Concept maps as assessments of expertise: Understanding of the complexity and interrelationships of concepts in science. *Journal of Advanced Academics, 31*(3), 254-297. https://doi.org/10.1177/1932202X20921770

*Mann, E. L. (2009). The search for mathematical creativity: Identifying creative potential in middle school students. *Creativity Research Journal, 21*(4), 338-348. https://doi.org/10.1080/10400410903297402

Marland, S. P. (1972). Education of the gifted and talented: Report to the Congress of the United States by the U.S. Commissioner of Education and background papers submitted to the U.S. Office of Education. 2 vols (Government Documents, Y4.L 11/2: G36). Government Printing Office.

McBee, M. T., Peters, S. J., & Miller, E. M. (2016). The impact of the nomination stage on gifted program identification: A comprehensive psychometric analysis. *Gifted Child Quarterly*, *60*(4), 258–278. https://doi.org/10.1177/0016986216656256

McBee, M. T., Peters, S. J., & Waterman, C. (2014). Combining scores in multiple-criteria assessment systems: The impact of combination rule. *Gifted Child Quarterly*, *58*(1), 69–89. https://doi.org/10.1177/0016986213513794

Miller, A. L. (2012). Conceptualizations of creativity: Comparing theories and models of giftedness. *Roeper Review, 34*(2), 94-103. *https://doi.org/10.1080/02783193.2012.660683*

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med 6*(7), e1000097. https://doi.org/10.1371/journal.pmed1000097

Mun, R. U., Ezzani, M., & Lee, L. E. (2020a). Culturally relevant leadership in gifted education: A systematic review. *Journal for the Education of the Gifted*, 1-34. Advance online publication. https://doi.org/10.1177/0162353220912009

Mun, R. U., Ezzani, M., Lee, L. E., & Ottwein, J. K. (2020b). Building systemic capacity to improve identification and services in gifted education: A case study of one district. Gifted Child Quarterly. Advance online publication. https://doi.org/10.1177/0016986220967376

Naglieri, J. A., & Ford, D. Y. (2003). Addressing underrepresentation of gifted minority children Using the Naglieri nonverbal ability test (NNAT). *Gifted Child Quarterly*, *47*(2), 155–160. https://doi.org/10.1177/001698620304700206

*Nakano, T. C., Primi, R., Ribeiro, W. J., & Almeida, L. S. (2016). Multidimensional assessment of giftedness: Criterion validity of battery of intelligence and creativity measures in predicting arts and academic talents. *Anales de Psicología, 32*(3), 628-637. http://dx.doi.org/10.6018/analesps.32.3.259391

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

Olszewski-Kubilius, P., Subotnik, R. F., & Worrell, F. C. (2017). The role of domains in the conceptualization of talent. *Roeper Review, 39*(1), 59-69. https://doi.org/10.1080/02783193.2017.1247310

*Oreck, B. A., Owen, S. V., & Baum, S. M. (2003). Validity, reliability, and equity issues in an observational talent assessment process in the performing arts. *Journal for the Education of the Gifted*, *27*(1), 62–94. *https://doi.org/10.1177/016235320302700104*

Peters, S. J. (2021). The challenges of achieving equity within public school gifted and talented programs. *Gifted Child Quarterly*. *https://doi.org/10.1177/00169862211002535*

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.org/10.1177/0016986216643165

Peters, S. J., Ottwein, J. K., Lee, L. E., & Matthews, M. S. (2020). Identification. In J. A. Plucker & C. M. Callahan (Eds.), *Critical Issues and Practices in Gifted Education* (3rd ed., pp. 261-272). Prufrock Press.

*Peters, S. J., & Pereira, N. (2017). A replication of the internal validity structure of three major teaching rating scales. *Journal of Advanced Academics*, *28*(2), 101–119. *https://doi.org/10.1177/1932202X17701940*

Pfeiffer, S. I., & Jarosewich, T. (2003). *Gifted Rating Scales.* Pearson Education.

*Pfeiffer, S. I., & Jarosewich, T. (2007). The gifted rating scales-school form: An analysis of the standardization sample based on age, gender, race, and diagnostic efficiency. *Gifted Child Quarterly, 51*(1), 39-50. https://doi.org/10.1177/0016986206296658

*Pfeiffer, S. I., Petscher, Y., & Jarosewich, T. (2007). Sharpening identification tools. *Roeper Review*, *29*(3), 206–211. https://doi.org/10.1080/02783190709554410

Plucker, J. A., & Callahan, C. M. (2014). Research on giftedness and gifted education: Status of the field and considerations for the future. *Exceptional Children, 80*(4), 390-406. https://doi.org/10.1177/0014402914527244

Plucker, J. A. (2011). Reanalysis of student responses to creativity checklists: Evidence of content generality. *Journal of Creative Behavior*, 33(2), 136–137. *https://doi.org/10.1002/j.2162-6057.1999.tb01042.x*

*Reid, C., Romanoff, B., & Algozzine, R. (2000). An evaluation of alternative screening procedures. *Journal of the Education of the Gifted, 23*(4), 378-396. *https://doi.org/10.1177/016235320002300404*

Reis, S. M., & Peters, P. M. (2020). Research on the schoolwide enrichment model: Four decades of insights, innovation, and evolution. *Gifted Education International*. Advance online publication. https://doi.org/10.1177/0261429420963987

Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan, 60*(3), 180-184, 261. *https://doi.org/10.1177/003172171109200821*

Renzulli, J. S., Reis, S. M., & Smith, L. H. (1981). The revolving door identification model. Creative Learning Press.

Renzulli, J. S. (2012). Reexamining the role of gifted education and talent development for the 21st century: A four-part theoretical approach. *Gifted Child Quarterly, 56*(3), 150-159. https://doi.org/10.1177/0016986212444901

Renzulli, J. S., & Hartman, R. K. (1971). Scale for rating behavioral characteristics of superior students. *Exceptional Children, 38*(3), 243-248. *https://doi.org/10.1177/001440297103800309*

Renzulli, J. S., Smith, L. H., White, A. J., Callahan, C. M., Hartman, R. K., Westberg, K. L. (2002). Scales for rating the behavioral characteristics of superior students – revised edition. Creative Learning Press.

Rhodes, M. (1961). An analysis of creativity. *Phi Delta Kappan, 42*(7), 305-310. Retrieved from http://www.jstor.org/stable/20342603

Richards, L. (2015). *Handling qualitative data: A practical guide* (3rd ed.). Sage Publications.

Rimm, S. B., Siegle, D., & Davis, G. A. (2018). Education of the gifted and talented (7th ed.). Pearson.

Rinn, A. N., Mun, R. U., & Hodges, J. (2020). *2018-2019 State of the states in gifted education*. National Association of Gifted Children and the Council of State Directors of Programs for the Gifted. https://www.nagc.org/2018-2019-state-states-gifted-education

*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*

*Rosado, J. I., Pfeiffer, S. I., & Petscher, Y. (2008). The reliability and validity of a Spanish translated version of the gifted rating scales. *Gifted and Talented International, 23*(1), 105-114. https://doi.org/10.1080/15332276.2008.11673517

*Rostan, S. M. (2005). Educational intervention and the development of young art students' talent and creativity. *Journal of Creative Behavior, 39*(4), 237-261. https://doi.org/10.1002/j.2162-6057.2005.tb01260.x

Ryan, G. W., & Bernard, H. R. (2003). Techniques to identify themes. *Field Methods, 15*(1), 85- 109. https://doi.org/10.1177/1525822X02239569

Ryser, G. R., & McConnell, K. (2004). *Scales for identifying gifted students.* Prufrock Press.

*Sarouphim, K. M. (2000). Internal structure of DISCOVER: A performance-based assessment. *Journal for the Education of the Gifted*, *23*(3), 314–327. https://eric.ed.gov/?id=EJ609790

*Sarouphim, K. M. (2004). DISCOVER in middle school: Identifying gifted minority students. *Journal of Secondary Gifted Education*, *15*(2), 61–69. https://files.eric.ed.gov/fulltext/EJ682704.pdf

*Sarouphim, K. M. (2009). The use of a performance assessment for identifying gifted Lebanese students: Is DISCOVER effective? *Journal for the Education of the Gifted*, *33*(2), 275–295. https://doi.org/10.1177/016235320903300206

*Sarouphim, K. M., & Maker, C. J. (2010). Ethnic and gender differences in identifying gifted students: A multi-cultural analysis. *International Education*, *39*(2), 42–56. https://trace.tennessee.edu/internationaleducation/vol39/iss2/4

Siegle, D., Gubbins, E. J., O'Rourke, P., Langley, S. D., Mun, R. U., Luria, S. R., Little, C. A., McCoach, D. B., Knupp, T., Callahan, C. M., & Plucker, J. A (2016). Barriers to underserved students' participation in gifted programs and possible solutions. *Journal for the Education of the Gifted*, *39*(2), 103-131. https://doi.org/10.1177/0162353216640930

Smith, J. K., Smith, L. F., & Smith, B. K. (2017). The reproducibility crisis in psychology: Attack of the clones or phantom menace. In M. C. Makel & J. A. Plucker (Eds), *Toward a more perfect psychology: Improving trust, accuracy, and transparency in research* (pp. 273-287). American Psychological Association.

Starko, A. J. (2018). *Creativity in the classroom: Schools of curious delight.* Routledge.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* Cambridge University Press.

Sternberg, R. J. (2005). WICS: A model of positive educational leadership comprising wisdom, intelligence, and creativity synthesized. *Educational Psychology Review, 17*(3), 191-262. https://doi.org/10.1007/s10648-005-5617-2

*Subhi, T. (1997). Who is gifted? A computerized identification procedure. *High Ability Studies*, *8*(2), 189-211. https://doi.org/10.1080/1359813970080205

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.org/10.1177/1529100611418056*

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2019). Environmental factors and personal characteristics interact to yield high performance in domains. *Frontiers in psychology*, *10*(2804), 1-8. https://doi.org/10.3389/fpsyg.2019.02804

Tannenbaum, A. J. (1983) *Gifted children: Psychological and educational perspectives.* Macmillan Publishing.

Tannenbaum, A. J. (2003). Nature and nurture of giftedness. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (3rd ed., pp. 45-59). Allyn & Bacon.

Torrance, E. P. (1971). Are the Torrance Tests of Creative Thinking biased against or in favor of the “disadvantaged” groups? *Gifted Child Quarterly, 15*(2)*,* 75–80. *https://doi.org/10.1177/001698627101500201*

Torrance, E. P. (1974). *Torrance Tests of Creative Thinking: Norms-technical manual (Research ed.).*. Scholastic Testing Service.

Torrance, E. P. (2004). The role of creativity in identification of the gifted and talented. In D. J. Treffinger (Ed.), *Creativity and Giftedness* (pp. 79-86). Corwin Press.

Torrance, E. P. (2017). *Torrance test of creative thinking (TTCT) norms-technical manual for figural forms A and B.* Scholastic Testing Service.

Treffinger, D. J. (2004). Research on creativity. In D. J. Treffinger (Ed.), *Creativity and Giftedness* (pp. 87-96). Corwin Press.

Treffinger, D. J., & Feldhusen, J. F. (1996). Talent recognition and development: Successor to gifted education. *Journal for the Education of the Gifted*, *19*(2), 181-193. https://doi.org/10.1177/016235329601900205

Udall, A. J. (1987). Peer referral as a process for locating Hispanic students who may be gifted [Doctoral dissertation, University of Arizona]. Dissertation Abstracts International. https://repository.arizona.edu/handle/10150/184312

U. S. Department of Education. (2020). *Civil rights data collection: 2017-18 State and national estimations.* https://ocrdata.ed.gov/estimations/2017-2018

Wallach, M. A., & Kogan, N. (1965). A new look at the creativity-intelligence distinction. *Journal of Personality, 33*(3), 348–369. https://doi.org/10.1111/j.1467-6494.1965.tb01391.x

Williams, F. E. (1980). *Creativity assessment packet (CAP)*. D.O.K. Publishers

Worrell, F. C., & Dixson, D. D. (2018). Recruiting and retaining underrepresented gifted students. In S. I. Pfeiffer (Ed.), *Handbook of Giftedness in Children* (pp. 209-226). American Psychological Association. https://doi.org/10.1007/978-3-319-77004-8_13

Worrell, F. C., & Dixson, D. D. (2020). Diversity and gifted education. In J. A. Plucker & C. M. Callahan (Eds.), *Critical Issues and Practices in Gifted Education* (3rd ed., pp. 169-183). Prufrock Press.

*Wu, I. C., Pease, R., & Maker, C. J. (2019). Students' perceptions of a special program for developing exceptional talent in STEM. *Journal of Advanced Academics, 30*(4), 474-499. https://doi.org/10.1177/1932202X19864690

*VanTassel-Baska, J., Feng, A. X., & Evans, B. L. (2007). Patterns of identification and performance among gifted students identified through performance tasks: A three-year analysis. *Gifted Child Quarterly*, *51(3)*, 218-231. https://doi.org/10.1177/0016986207302717

VanTassel-Baska, J. (2008). *Alternative assessments with gifted and talented students.* Prufrock Press.

*Zhbanova, K., Rule, A., & Stichter, M. (2015). Identification of gifted African American primary grade students through leadership, creativity, and academic performance in curriculum material making and peer-teaching: A case study. *Early Childhood Education Journal*, *43*(2), 143–156. https://doi.org/10.1007/s10643-013-0628-z

Zotero. (2020). Zotero (Version 5.0.93) [Computer software]. *https://www.zotero.org/*

2.7 Data Tables

Table 2.1

Database Search Return

Database	Use Advanced Search Each term on its own line	Years; Search Return Years	Options	Returns
Academic Search Complete via Ebscohost	TX “creative ability” OR “creative potential” OR creativity AND AB gifted OR talent AND AB identification OR assessment	1973-Oct 2020	English; Full Text; Peer Reviewed; Academic Journals	252
ERIC via Ebscohost	TX “creative ability” OR “creative potential” OR creativity AND AB gifted OR talent AND AB identification OR assessment	1989 - Oct 2020	English; Full Text; Peer Reviewed; Academic Journals	51
Education Source via Ebscohost	TX “creative ability” OR “creative potential” OR creativity AND AB gifted OR talent AND AB identification OR assessment	1980 - Oct 2020	English; Full Text; Peer Reviewed; Academic Journals	359
Psychology and Behavioral Sciences Collection	TX “creative ability” OR “creative potential” OR creativity AND AB gifted OR talent AND AB identification OR assessment	1984-Aug 2020	English; Full Text; Peer Reviewed; Academic Journals	149
Professional Development Collection	TX “creative ability” OR “creative potential” OR creativity AND AB gifted OR talent AND AB identification OR assessment	1984 - Oct 2020	English; Full Text; Peer Reviewed	224
SocINDEX with Full Text	TX “creative ability” OR “creative potential” OR creativity AND AB gifted AND AB identification OR assessment	1959-Oct 2020	English; Full Text; Peer Reviewed; Academic Journals	16
APA PsycINFO	TX “creative ability” OR “creative potential” OR creativity AND AB gifted AND AB identification OR assessment	1959-Oct 2020	English; Peer Reviewed; Linked to Full Text; Academic Journals;	54
APA PsycArticles	TX “creative ability” OR “creative potential” OR creativity AND AB gifted AND AB identification OR assessment	1959-Oct 2020	English; Full Text; Peer Reviewed; Academic Journals	17
			Total	1122

Table 2.2

Inclusion and Exclusion Criteria

Inclusion	Exclusion
• Equitable access to gifted programs • Underrepresented Groups in Gifted Education • Group Differences (Gender, Race/Ethnicity, Socioeconomic status, rurality, special education, learning disabilities, twice-exceptionality) • K-12 Students (Outcome focused on students) • Measure of Creativity: Creativity, Problem-solving, Divergent Thinking, Performance-Based, other Alternative Assessment • Gifted Identification • Talent Development • Cross-cultural (outside of the US)	• Does not discuss gifted and talented • Type of publication: Books, Book Sections, Dissertation/Thesis, Conceptual/Theoretical, Editorials, Practitioner Articles. • GT teachers or classroom application without students • Workplace creativity • Higher education • Does not report how students were identified for a service/program • Unspecified creativity assessment • Conceptions of creativity or giftedness by teachers and does not focus on students • Cross-cultural and does not compare group differences

Table 2.3

Included Studies from Quality Screening

Author/Year	Screening Questions		Qualitative Studies					Quantitative Descriptive Studies				
	Are there clear research questions? Or is there a designated purpose?	Do the collected data allow to address the research questions?	Is the qualitative approach appropriate to answer the research question?	Are the qualitative data collection methods adequate to address the research question?	Are the findings adequately derived from the data?	Is the interpretation of results sufficiently substantiated by data?	Is there coherence between qualitative data sources, collection, analysis and interpretation?	Is the sampling strategy relevant to address the research question?	Is the sample representative of the target population?	Are the measurements appropriate?	Is the risk of nonresponse bias low?	Is the statistical analysis appropriate to answer the research question?
Anderson (2020)	Yes	Yes	Yes	Yes	Yes	Yes	Can't tell					
Chambers (1980)	Yes	Yes						Yes	Yes	Yes	Can't tell	Yes
Cunningham et al. (1998)	Yes	Yes						Yes	Yes	Yes	Yes	Yes
Dewing (1970)	Yes	Yes						Yes	Yes	Can't tell	Can't tell	Yes
Diener et al. (2014)	Yes	Yes						Yes	Yes	Yes	Yes	No
Elliot et al. (1986)	Yes	Yes						Yes	Yes	Yes	Yes	Yes
Freund & Holling (2008)	Yes	Yes						Yes	Yes	Yes	Yes	Yes
Harris et al. (2009)	Yes	Yes	Yes	Yes	Yes	Yes	Yes					
Harty et al. (1984)	Yes	Yes						Yes	Yes	Yes	Yes	Yes
Karadag & Pfeiffer (2016)	Yes	Yes						Yes	Yes	Yes	Yes	Yes
Kim et al. (2009)	Yes	Yes						Yes	Yes	Yes	Can't tell	Yes
Labar & Frumos (2013)	Yes	Yes						Yes	Yes	Yes	Can't tell	Yes
LaFrance (1995)	Yes	Yes						Yes	Yes	Yes	Yes	Yes
Li et al. (2009)	Yes	Yes						Yes	Yes	Yes	Can't tell	Yes
Maker (2020)	Yes	Yes						Yes	Yes	Can't tell	Yes	Yes
Maker & Zimmerman (2020)	Yes	Yes						Yes	Yes	Can't tell	Yes	Yes
Mann (2009)	Yes	Yes						Yes	Yes	Yes	Can't tell	Yes
Nakano et al. (2016)	Yes	Yes						Yes	Yes	Yes	Yes	Yes
Oreck et al. (2003)	Yes	Yes						Yes	Yes	Yes	Yes	Yes
Peters & Pereira (2017)	Yes	Yes						Yes	Yes	Yes	Yes	Yes
Pfeiffer & Jarosewich (2007)	Yes	Yes						Yes	Yes	Yes	Yes	Yes
Pfeiffer et al. (2007)	Yes	Yes						Yes	Yes	Yes	Yes	Yes
Reid et al. (2000)	Yes	Yes						Yes	Yes	Yes	Yes	Yes
Romanoff et al.(2009)	Yes	Yes						Yes	Yes	Can't tell	Yes	Yes
Rosado et al. (2008)	Yes	Yes						Yes	Yes	Yes	Yes	Yes
Rostan (2005)	Yes	Yes						Yes	Yes	Yes	Yes	Yes

(table continues)

Author/Year	Screening Questions: Are there clear research questions? Or is there a designated purpose?	Screening Questions: Do the collected data allow to address the research questions?	Qualitative Studies: Is the qualitative approach appropriate to answer the research question?	Qualitative Studies: Are the qualitative data collection methods adequate to address the research question?	Qualitative Studies: Are the findings adequately derived from the data?	Qualitative Studies: Is the interpretation of results sufficiently substantiated by data?	Qualitative Studies: Is there coherence between qualitative data sources, collection, analysis and interpretation?	Quantitative Descriptive Studies: Is the sampling strategy relevant to address the research question?	Quantitative Descriptive Studies: Is the sample representative of the target population?	Quantitative Descriptive Studies: Are the measurements appropriate?	Quantitative Descriptive Studies: Is the risk of nonresponse bias low?	Quantitative Descriptive Studies: Is the statistical analysis appropriate to answer the research question?
Sarouphim (2000)	Yes	Yes						Yes	Yes	Yes	Can't tell	Yes
Sarouphim (2009)	Yes	Yes						Yes	Yes	Yes	Yes	Yes
Sarouphim (2004)	Yes	Yes						Yes	Yes	Yes	Yes	Yes
Sarouphim & Maker (2010)	Yes	Yes						Yes	Yes	Yes	Yes	Yes
Suhbi (1997)	Yes	Yes						Yes	Yes	Yes	Can't tell	Yes
VanTassel-Baska, et al. (2007)	Yes	Yes						Yes	Yes	Yes	Can't tell	Yes
Wu et al. (2019)	Yes	Yes	Yes	Yes	Yes	Can't tell	Can't tell	Yes	Yes	Yes	Yes	Yes
Zhbanova et al. (2015)	Yes	Yes	Yes	Yes	Yes	Can't tell	Can't tell	Yes	Yes	Yes	Yes	Yes

Table 2.4

Themes, Subthemes, and Coding Structure

Themes	Subthemes	Second Phase Codes	First Phase Codes
Overcoming identification barriers through inclusive practices for developing exceptional talent	Barriers to Talent Development	• Overcoming Barriers • Nomination bias • Test bias • Cut off scores • Two-phase system	• Identification • Two-phase system • Cut off scores • Nomination bias • Test bias • Underrepresentation • Deficit thinking
	Inclusive Identification Practices	• Achieving Excellence • Multiple criteria systems • Identification Process • Universal screening	• Equitable Access • Universal screening • Excellence/high achievement • Strengths • Multiple criteria • Identification process • Inclusive practices
	• Broadened Conceptions of Giftedness for Identifying and Developing Exceptional Talent	• Broadened Conceptions of Giftedness • Developing Exceptional Talent • Psychosocial Factors and Benefits	• Conceptions of giftedness • Intellectual ability • Leadership • Motivation/Task Commitment • Exceptional talent • Gifted programs • STEM program • Enrichment activities • Excellence/high achievement

(table continues)

Themes	Subthemes	Second Phase Codes	First Phase Codes
Conceptions of Creativity Used in Gifted Identification Processes	• Domain-General versus Domain-Specific Creativity in Gifted Identification • Identifying Creativity as a Product or Performance • Identifying Creative Processes: Divergent Thinking & Problem-Solving • Identifying Creative People: Creative behaviors or personality characteristics • The Creative Press: Environments that Encourage Creative Potential	• Conceptions of Creativity • Creativity - Arts-based conception • Creativity - Creative problem-solving • Creativity – process • Creativity – product • Creativity – person • Creativity – press • Creativity - domain-specific • Creativity - domain-general • Creativity - divergent thinking	• Creativity • Creative ability • Problem Solving • Artistic ability • Creative achievement or production • Divergent thinking • Creative process • Performance-based assessments • Expert ratings • Creative environment • Domain-specificity
Research Focused on Validity of Creative Measures	Content, Construct, & Criterion Validity of Creative Measures	Types of validity	• Validity • Content Validity • Construct Validity • Convergent Validity • Discriminant Validity • Criterion-Related Validity • Concurrent Validity • Predictive Validity • Reliability
	Group differences on creative measures	Group differences	• Gender/sex • Disability • Special Education • Autism Spectrum Disorder • Race/Ethnicity • Socioeconomic status • English Language Learning

Table 2.5

Summary of Study Characteristics

Author/Year	Purpose	Country	Participants	Group	Research Design	Methods	Major Findings
Anderson (2020)	Retrospectively explores the narratives and experiences of gifted Black girls	United States	3 gifted-identified Black women with doctoral degrees.	Ethnicity: Gender	Qualitative: Ethnography, Interpretive; Single Case Design	Semi-structured Interviews	Encourages using tools like Frasier's TABS, which incorporates a component on problem-solving ability, how educators need to create a conducive learning environment, and the need to discuss the experiences of intersectionality in gifted Black girls.
Chambers (1980)	Develop a simple, effective method to identify gifted Mexican American children for future intellectual/creative contributions.	United States	298 3rd-6th Mexican American students in two California public schools.	Ethnicity	Quantitative: Descriptive	Descriptive; Covariance, t-tests	Socioeconomic level and language spoken in the home had little effect on performance. Encourages the use of multiple sequential methods of identification that includes measures related to creativity.
Cunningham et al. (1998)	Establish reliability and validity of the peer nomination form. Specifically, authors evaluate racial and gender differences of using a peer nomination form.	United States	670 4th-6th grade students within three large southwestern U.S. districts.	Gender, Race	Quantitative: Correlational	Factor Analysis, ANOVA, Descriptive Discriminant Analysis	Evidence of construct validity with higher correlations of items for creative and artistic behaviors in comparison to arts and intellectual ability items. No significant difference between Hispanic and Caucasian students within this sample. Small gender differences found that favor males.
Dewing (1970)	Examine the relationship between scores on four of the Minnesota Tests (two verbal and two non-verbal) and five measures of creative performance.	Australia	394 children in the 7th grade.	Sex	Quantitative: Descriptive	Descriptive; *t*-tests	No sex differences found for creative thinking or performance within the study. No relationship found between measures of IQ and creative performance.

(table continues)

Author/Year	Purpose	Country	Participants	Group	Research Design	Methods	Major Findings
Diener et al. (2014)	Develop a visual-spatial measure of creativity that would assess the strengths of youth with autism spectrum disorder (ASD)	United States	Random sampling of 27 projects from students with Autism Spectrum out of 100 projects.	Autism Spectrum Disorder (ASD)	Quantitative: Correlational	Inter-rater reliability; Bivariate Statistics.	The strengths of this creativity assessment are that it (a) utilizes 3D-modeling computer software, (b) takes place in a natural learning environment, (c) shows promise for autistic youth. Evidence of content validity from the relation between google design coders conceptions of creativity and total creativity scores.
Elliot et al. (1986)	(a) the overall usefulness of the SRBCSS, and (b) whether the SRBCSS predicts scholastic achievement for different cultural or socioeconomic groups.	United States	402 students in the 3rd-6th grade in a Southwestern district	Ethnicity; SES	Quantitative: Correlational	Descriptive; Regression	Correlations were higher for a group of low socioeconomic Hispanics, with the highest correlations being between the Reading Comprehension subtest of the SAT and the Creative scales of the SRBCSS.
Freund & Holling (2008)	Examines if creativity is related to scholastic achievement.	Germany	1133 students grouped within 60 classrooms.	Gender	Quantitative: Correlational	Multilevel modeling	The effects of creativity and reasoning ability are both highly significant, as is the effect for the interaction term (at the 5% level), but the effect of reasoning ability is largest. In general, girls outperform boys, and there are differences between the school types.
Karadağ, Karabey, & Pfeiffer (2016)	Evaluates the psychometric properties of the GRS-P scale with a Turkish sample.	Turkey	The sample consisted of 30 preschool teachers (of a total of 390 preschoolers) working in one of the 15 preschools across Izmir during the 2014-2015 school year.	Gender	Quantitative: Correlational	Confirmatory Factor Analysis	Results found evidence of a five-factor model that the subscales scores for intellectual, academic, and creative ability did not differ by gender. There was no significant difference across age groups.
Kim, Shim, & Hull (2009)	Examines giftedness through a focus on implicit theories and the second study examines the criteria for identifying gifted students in Korea.	Korea	The total number of participants were 328 (245 males & 83 females), including 71 scientists, 73 parents of school-age children, 104 teachers and 80 college students in Daejeon, Korea.	Gender	Study 1 - Qualitative: Descriptive Study 2 - Quantitative: Correlational	MANOVA/ ANOVA; Open-ended questionnaire	For study 1: Major conceptions of giftedness included: intelligence, task commitment, creativity, interpersonal relationship, moral sense, and artistic talent. Small to moderate percentages of students, parents, teachers, and scientists consider creativity an important component. For study 2: No statistically significant differences on artistic talent and creativity among students in a science gifted program, humanities gifted program, or regular students.
Lâbar & Frumos (2013)	Adapt an original version of SRBCSS for self- assessment and assess the reliability and validity of the Romanian version of the SRBCSS.	Romania	The data were obtained from a number of 180 students (7th and 10th graders). The sample includes 76 boys and 104 girls, 99 students from gymnasium and 81 from high school.	Gender	Quantitative: Correlational	Principal Component Analysis, T-tests, Bivariate Correlations	All scales discriminated between the Olympic and non-Olympic sample. There were no significant gender differences found.
LaFrance (1995)	Analyze differences between students with learning disabilities, gifted students, and students with learning disabilities and identified as gifted.	Canada	The sample, drawn from four school districts or Boards of Education across Ontario, consisted of 30 children who were gifted, 30 who were learning disabled and 30 who were gifted with learning disabilities.	Gender; Learning Disabled	Quantitative: Correlational	MANOVA, Descriptive Discriminant Analysis (DDA), Unspecified qualitative method for criterion-referenced analysis.	Found evidence that the Torrance Test of Creative Thinking can be helpful in identifying gifted students who have a learning disability.

(table continues)

Author/Year	Purpose	Country	Participants	Group	Research Design	Methods	Major Findings
Li, Lee, Pfeiffer, Kamata, Kumtepe, & Rosado (2009)	Examines measurement invariance of the Gifted Rating Scales—School Form (GRS–S) across five countries.	United States, Puerto Rico, China, South Korea, & Turkey	A total of 1,817 students were rated by 287 teachers using either translated versions of GRS–S or the original English GRS–S.	Cross-Cultural; Gender	Quantitative: Correlational	Multigroup Confirmatory Factor Analysis; Measurement Invariance	Patterns of factor loadings and the factor variances and covariances are invariant across the five groups, meaning that these measures yielded equivalent results across five cultures.
Maker (2020)	Evaluates the effectiveness of existing instruments and others created specifically for a 4-year project over the Cultivating Diverse Talent in STEM (CDTIS)	United States	43 high school juniors split within two groups. White (*n*=10), Hispanic (*n*=11), American Indian (*n*=14), African American (*n*=3), Asian American (*n*=5).	Ethnicity; Gender; SES	Quantitative: Descriptive	Chi-square, *t*-tests	M2 students scored higher on all the performance assessments of creative problem-solving and at similar levels on concept maps and mathematical problem-solving.
Maker & Zimmerman, (2020).	Developed a concept mapping assessment to identify STEM talent in high school students from traditionally underrepresented groups.	United States	62 students field tested the life science concept map assessment; 61 students field tested the physical science concept map.	Ethnicity; SES	Quantitative: Descriptive	Descriptive statistics	When used separately or in combination with other concept maps and other assessments of creative problem solving, they provide a measure of the strengths and creative abilities of students of color in STEM.
Mann (2009)	Analyze assessments for creative potential in mathematics.	United States	89 7th graders from Connecticut	Gender	Quantitative: Correlational	Regression	Revealed that the correlation between CAMT and SRBCSS–Creativity was not a significant predictor; thus, scores on the SRBCSS-creativity measure were not predictive of scores of mathematical divergent thinking tasks. Gender had a small negative relationship with performance on mathematical divergent thinking tasks.
Nakano et al. (2016)	Assess criterion validity of the Battery of Giftedness Assessment (BaAH/S) and group differences in gifted students in the areas of academic and artistic talents.	Brazil	987 children and adolescents; Gender: 464 boys, 523 girls; Two groups: regular students (n=866), gifted students; 67 academic abilities, 34 artistic abilities, 20 no-domain identified.	Sex/Gender (control)	Quantitative: Correlational	Descriptive, Structural Equation Modeling (MIMIC)	Found positive associations between creativity (verbal and figural) and intelligence measures. Fluid reasoning predicts the different identified giftedness and shows a relationship of creativity and giftedness. Authors argue to add different abilities within the identification process. Small gender differences found in quality of drawings in figural fluency, where girls did better than boys.
Oreck, Owen, & Baum (2003)	Discusses the issues involved in designing and administering a talent assessment in schools with diverse student populations.	United States	Schools were selected for the initial study based on their existing participation with Arts Connection's Young Talent Program (YTP).	Ethnicity; Gender	Quantitative: Correlational	Exploratory Principal Axis Factor Analysis, Interrater Reliability, Stability, Multivariate probability	The students identified through D/M/T TAP, unlike those selected for gifted and talented programs through IQ or academic test scores, accurately represented the demographics of the schools, including students in self-contained special education and bilingual classrooms.
Peters & Pereira (2017)	Replicate the internal validity structure of three teacher rating instruments (SIG, GRS-S, HOPE Scale)	United States	Following this philosophy, we recruited a single, diverse school district based on size (approximately 25,000 K-12 students), diversity (53% non-Caucasian), and family income (49% on Free or Reduced-Price Meals).	Ethnicity; Family Income	Quantitative: Correlational	Confirmatory Factor Analysis (CFA); MGCFA	Lack of fit across race/ethnicity and income groups is consistent with what was found with the HOPE Scale and was dissimilar with previous research on the GRS which found no race or ethnic differences.

(table continues)

Author/Year	Purpose	Country	Participants	Group	Research Design	Methods	Major Findings
Pfeiffer, Petscher, & Jarosewich (2007)	Analyze the standardization sample for the Gifted Rating Scales - Preschool/Kindergarten form to assist in the identification of gifted students.	United States	Standardization sample; 188 boys (50%); 187 girls (50%).	Gender; Race	Quantitative: Correlational	MANOVA	Support for the internal structure of the scale, gender differences found for artistic talent, where females score slightly higher on artistic talent. No gender differences found on the other subscales. Ethnic/racial differences were found on each of the subscales. For the creativity subscale, Caucasians were rated higher than Hispanic and African American children.
Reid, Romanoff, & Algozzine (2000)	Investigated the effectiveness of the problem-solving assessment (PAS) procedure, an application of multiple intelligence theory that focuses on identifying students for programs of gifted education.	United States	1100 2nd grade elementary students; 434 recommended for gifted placement; Males (54%); Females (46%)	Gender	Quantitative: Descriptive	Chi-Square, Unspecified	Placement recommendations were predicted better by PSA scores than by MAT scores, PSA scores were not independent of each other, and moderate concurrent validity was evident for MAT and PSA scores. More than twice as many students were identified using PSA compared to MAT criterion.
Romanoff, Algozzine, & Nielson (2009)	Evaluate the performance of elementary school children in a Southeastern state identified as gifted using an assessment process based on MI theory, the Problem-Solving Assessment, with a comparable group of students referred for assessment but not identified as gifted.	United States	The district was the 25th largest school district in the United States with 86 elementary schools, 28 middle schools, 14 high schools, and 11 special schools at the time of the study. The district is primarily urban, although visits to schools in the county reveal inner city, suburban, and even rural areas.	Ethnicity; Gender	Quantitative: Causal-Comparative/ Quasi-Experimental	ANOVA	Scores for African American students were below those for the Caucasian peers in both groups, the difference between the groups was smaller in reading and mathematics for students identified and placed in gifted programs. Findings from this longitudinal study demonstrate that students selected as a result of the Problem-Solving Assessment and placed in a program for the gifted score on end-of-grade tests in reading and math significantly higher than students referred for assessment but not selected.
Rosado, Pfeiffer, & Petscher, (2008)	Preliminary examination of the psychometric properties of a newly developed Spanish translated version of the Gifted Rating Scales-School Form (GRS-S).	Puerto Rico	153 Puerto Rican students; 72 males; 81 females. Grades 1-8.	Gender	Quantitative: Correlational	Descriptive, Reliability, Confirmatory Factor Analysis	High internal consistency, strong intercorrelations between subscales, and evidence of support for a 6-factor model. Unknown reasons for why creativity and academic ability strongly correlate within this sample. Found motivation scores a significant predictor of academic performance.
Rostan, (2005)	Examines the role of personal expertise in a student's development of problem finding, domain-specific technical skill, perseverance, evaluation, and creative ideation.	United States	59 children enrolled in a private afterschool art program in New York City.	Gender	Quantitative: Correlational	MANCOVA	Gender was significantly related to creativity in the drawings from imagination and life- drawing erasures. Male participants tended to produce drawings from imagination eliciting higher assessments of creativity and tended to make fewer life-drawing erasures. Within the drawing from imagination task, gender provided significant adjustment to problem finding.
Sarouphim (2000)	Investigated the internal structure of DISCOVER, a relatively new performance-based assessment based upon the theory of multiple intelligences.	United States	Participants were 257 K–5 Navajo Indian and Mexican American students.	Ethnicity	Quantitative: Correlational	Spearman Rank Correlation	Results were inconclusive. Low and nonsignificant internal correlations provide preliminary evidence in support of the sound internal structure of the DISCOVER assessment. Further examination of the results seems to provide evidence in the opposite direction giving rise to a multitude of issues that need to be addressed in further research before one can call for the use of the DISCOVER assessment on a wider scale.

(table continues)

Author/Year	Purpose	Country	Participants	Group	Research Design	Methods	Major Findings
Sarouphim (2004)	Examine the validity of the 6-8 version of the DISCOVER assessment and investigate the effectiveness in identifying culturally diverse students.	United States	295 participants; 214 males; 81 females	Ethnicity; Gender	Quantitative: Correlational	Bivariate Correlations, Chi-Square Tests, MANOVA, ANOVA	The results indicate that males were not significantly higher than females. Likewise, no significant ethnic differences on DISCOVER performance were found.
Sarouphim (2009)	Investigate the effectiveness of DISCOVER, a performance-based assessment in identifying gifted Lebanese students.	Lebanon	248 students (121 boys, 127 girls) from Grades 3–5 at two private schools in Beirut, Lebanon.	Gender	Quantitative: Correlational Qualitative: Descriptive	Bivariate Correlations, Interviews, MANOVA	Found significant gender differences in math ratings in 5th grade, but not in the younger students. However, non-significant gender differences across students indicates evidence that the DISCOVER is not gender biased.
Sarouphim & Maker (2010)	Examine ethnic and gender differences in using DISCOVER, a performance-based assessment, for identifying gifted students.	United States	The sample consisted of 941 students from Grades K-5 belonging to six ethnicities: White Americans, African Americans, Hispanics, Native-Americans, South Pacific/Pacific Islanders, and Arabs.	Ethnicity; Gender	Quantitative: Correlational	MANOVA	Plots of the interaction showed that South Pacific/Pacific Islanders scored highest on Oral Linguistic whereas White Americans scored highest in Math and Native Americans scored highest in Spatial Artistic activity. No gender differences in identification were found. All ethnic groups were well represented among identified students, suggesting that DISCOVER might be used in different countries and with culturally diverse students.
Subhi (1997)	Identify a sample of gifted primary school children in Jordan and to ascertain the efficiency and effectiveness of each of the identification methods.	Jordan	4583 third graders in the 25 private primary schools in Amman Educational District	Sex/Gender	Quantitative: Correlational	Descriptive, bivariate statistics	217 pupils were identified as gifted through the multiple criteria identification procedure. There were no significant differences between the mean scores of males and females on the study's measures.
VanTassel-Baska et al. (2007)	Examined the trend of identification and achievement patterns of performance task-identified students over a span of 6 years.	United States	Across 20 school districts in the State of South Carolina (25% of districts); 30,526 gifted students; 22,671 (74.3%) traditionally identified students; 7,855 (25.7%) performance task-identified students.	Ethnicity; SES; Gender	Quantitative: Descriptive	Descriptive statistics	A higher proportion of uneven identification profiles among performance task-identified students in comparison to their counterparts identified through traditional methods suggests that flexible instructional accommodations might be more appropriate for performance task-identified students. Also, performance-based protocols yielded more female students in gifted programs.
Wu et al. (2019)	Focuses on how underrepresented students with exceptional talent in STEM perceive their research program and experiences.	United States	13 11th grade students; 4 females; 9 males; selected by the CDTIS team from assessments of mathematics, spatial analytical ability, life science concept maps, naturalist, physics concept maps, and mechanical–technical abilities.	Ethnicity	Qualitative: Descriptive	Thematic analysis; General inductive approach	Students demonstrated strengths within and across disciplines. The overarching theme found was active involvement in problem-solving inspired and motivated students with exceptional talent from student perceptions of the STEM program.
Zhbanova, et al. (2015)	Investigated whether some minority students who achieved at an average level in their daily work and on standardized tests might be identified for gifted services through an alternative means	United States	The eight children who participated in the African Animal enrichment program.	Ethnicity	Qualitative: Case Study	Observation, Pre/Post-test Sociogram, Curriculum Intervention	Two African American students were identified as gifted through the study. The three minority students not only benefitted academically and socially from the project, but two were able to display strong enough leadership and creative skills in addition to solid academic achievement to convince the researchers and classroom teacher that they should be identified as gifted students.

Table 2.6

Included Studies Creativity Measures

Type of Assessment	Instruments/Intervention	Author
Creative Person	Frasier's Traits, Aptitudes, and Behaviors (TABS)	Anderson (2020)
	Udall Peer Nomination Form	Cunningham et al. (1998)
	Torrance Creative Leisure Interests Checklist	Dewing (1970)
	Golann Creative Motivation Scale (Golann, 1962)	Dewing (1970)
	Unspecified Peer rating	Dewing (1970)
	Scales for Rating the Behavioral Characteristics of Superior Students (SRBCSS)	Elliot et al. (1986); Harty et al. (1984); Lâbar & Frumos (2013); Mann (2009); Subhi (1997)
	Gifted Rating Scales - Preschool/Kindergarten	Karadağ et al. (2016); Peters & Pereira (2017); Pfeiffer et al. (2007); Pfeiffer & Jarosewich (2007); Rosado et al. (2008)
	Questionnaire: Implicit Theories of Giftedness	Kim et al., (2009)
	Torrance Creative Strengths Checklist	LaFrance (1995)
	Gifted Rating Scales - School Form	Li et al. (2009);
	What Kind of Person Are You? (WKOPAY?; Khatena & Torrance, 1976).	Mann (2009)
	Sociogram survey based on the Alpha Project Peer Nomination Simulation (Renzulli et al. 1981)	Zhbanova et al. (2015)
Creative Process	Wallach-Kogan Tests of Creativity	Chambers (1980)
	Torrance Test of Creative Thinking - Figural/Verbal	LaFrance (1995); Subhi (1997)
	Barron-Welsh Art Scale	Chambers (1980)
	Berlin Structure of Intelligence Test for Youth: Assessment of Talent and Giftedness (BIS-HB)	Freund & Holling (2008)
	Battery for Giftedness Assessment (BaAH/S)	de Cassia Nakano et al. (2016)
	Creative Ability in Mathematics Test (CMAT)	Mann (2009)

(table continues)

Type of Assessment	Instruments/Intervention	Author
	Problem-Solving Assessment	Reid et al. (2000); Romanoff et al. (2009)
	Matrix Analogy Test- Short Form	Reid et al. (2000)
Creative Products or Performance	Performance-based Assessments	Maker (2020); Maker & Zimmerman (2020); VanTassell-Baska et al. (2007); Wu et al. (2019)
	The Talent Assessment Process in Dance, Music, and Theater (D/M/T TAP)	Oreck et al. (2003)
	Consensual Assessment Technique (CAT) for Drawing Tasks: Life Drawing, Imagination Drawing	Rostan (2005)
	Observation of Enrichment Lesson: Renzulli Enrichment Triad (Type I, II, III)	Zhbanova et al. (2015)
	DISCOVER	Sarouphim (2000); Sarouphim (2004); Sarouphim (2009); Sarouphim & Maker (2010)
	Portfolio	Harris et al. (2009)
	Modified Creativity Assessment Packet: 3D Modeling Projects using the design program SketchUp and assessed by Google Design Experts.	Diener et al. (2014)
	Imaginative Writing Task	Dewing (1970)