EQUITABLY IDENTIFYING GIFTED STUDENTS FROM LOW-INCOME AND/OR MULTICULTURAL BACKGROUNDS: INVESTIGATION OF THE *HOPE* TEACHER RATING SCALE

by

Hyeseong Lee

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Educational Studies

West Lafayette, Indiana

May 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Marcia Gentry, Chair

Department of Educational Studies

Dr. Richard Olenchak

Department of Educational Studies

Dr. Yukiko Maeda

Department of Educational Studies

Dr. Nielsen Pereira

Department of Educational Studies

Approved by:

Dr. Janet M. Alsup

This work is dedicated to my dearest Jiwoo, Taeyong, and Family

ACKNOWLEDGMENTS

My 5-year journey as a doctoral student would have been impossible if I had not received the support, encouragement, and guidance of my committee members, mentor scholars, colleagues at GERI, as well as my family and my lifelong friends in Korea. First, I would like to express my deepest thanks to my academic advisor, Dr. Marcia Gentry. Her life as a professional scholar has been an inspiration for me to become a better educator in academia and higher education. I appreciate her sincere advice and trustworthy feedback which has always improved my work as I had challenges writing papers in a non-native language. Looking back to the day I arrived at Purdue in 2016, every step of my progress I have made has been with her, and with the continuous support of the other committee members. I thank Dr. Richard Olenchak and Dr. Nielsen Pereira for providing me with insightful suggestions and advice, both on my dissertation and with my job recruiting process. I cannot imagine that this dissertation would have been completed without Dr. Yukiko Maeda's help, guiding me in research methods and statistics. In addition, I thank Dr. James Borland at Teachers College, Columbia University who provided critical insights when I first started my academic journey in the field of gifted education. I also wish to express my appreciation to my mentors, Drs. Chang-bok Kim, Ock-ryeo Kang, and Kyung-sung Kim at Seoul National University of Education for encouraging me to continue my work when I felt exhausted.

I wish to thank my colleagues and friends who I met at Purdue under the amazing support system. Thank you to my GERI colleagues, Aakash, Abdullah, Alice, Alissa, Andrés, Anne, Bekir, Corinne, Jaret, Juliana, Jungsun, Kristen, Mehdi, Natalie, Nesibe, Ophélie, Sareh, Soohyun, Tugce, Yao, and Zafer for their friendship, motivation, and collaboration. Next, I wish to express my appreciation to my family. No words would ever exist to voice my gratitude to my daughter (Jiwoo), husband (Taeyong), parents (Geon-young and Sung-nam), parents-in-law (Jong-in and Myung-sook), and my furry soul mate (Mamma). As a parent myself, I can understand how tolerant and supportive they are. Their unconditional love, sacrifice, and encouragement have been the best supports. I would like to particularly thank my mother and mother-in-law for taking good care of Jiwoo when I was pursuing my doctoral degree. Their help enabled me to write my dissertation non-stop. My gratitude also goes to my beloved husband, Taeyong. Thanks to his unfailing support, I could keep pursuing my goals. I thank my little girl Jiwoo for being patient and growing beautiful and healthy without mommy's full existence. I also acknowledge that without my father and father-

in-law's emotional and financial help, I would not have been able to fully concentrate on my studies. Lastly, I thank my friends (Sohyun, Jiyoung, Yujin, Minjae, and Minjung) and my past students (Class 6-1 in Myeondong Elementary School, Seoul, in 2013) in Korea. Friendships and memories with them have been a big part of my life and have enabled me to continue and complete my studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

Introduction

The achievement gap in education is defined as the difference between and among the test scores for various student groups (Byers, 2018; Webb & Thomas, 2015). Despite decades of efforts to close it, the academic achievement gap among and between different ethnic and income groups persists worldwide (Mortenson, 2018). According to a Social Equity Theory (SET; McKown, 2013), two classes of social processes, "direct" and "signal," affect the achievement gap by functioning in combination with each other. "Direct" influences promote academic achievement equally for all students across diverse backgrounds. When "direct" influences are distributed differently, the achievement gap occurs. Conversely, "signal" influences are the signs that society has negative expectations and stereotypes of marginalized groups. If the students detect such signals, erosion occurs in terms of their achievement, and both components contribute to the achievement gap (Frederickson, 1990; McKown, 2013). Hence, it is important to continuously provide equal and sufficient support to students from diverse backgrounds. Fair and non-judgmental decisions by educators including school teachers, researchers, and policymakers in education are necessary.

Closing the achievement gap in public education has been a goal of United States educators highlighted by the No Child Left Behind Act (Henfield et al., 2017). The field of gifted education was not an exception and problematized its lack of equity for underidentifying and serving underrepresented students in gifted programs. Plucker et al. (2010; 2013) highlighted achievement gaps among the highest achieving groups as "excellence gaps," stating that "Low-income and minority students were much less likely to reach advanced levels of proficiency on state or national assessments, and the gaps between the top-performing disadvantaged students and White and more

affluent peers were significant" (Plucker et al., 2013, p.1). They also added that the inequities among high-ability students do not decrease and may even grow over time, despite the emphasis on high achievement for all (Plucker et al., 2013).

Although it is evident that an achievement gap has existed and endured for decades, researchers have had difficulties determining the factors that cause the differences among and between various groups of students (Ford & Moore, 2013; Robinson, 2014). Particularly, Ford and Moore (2013) argued there is no evidence showing students from any racial background or socioeconomic status are hereditarily or culturally advanced compared to other groups. The claim suggests there should not be a significant achievement gap between groups if the students from marginalized groups have equitable access to a quality education. If their hypothesis is true, then external factors contribute to academic achievement gaps and include constructs such as discrimination, bias, racism, and poverty which prevent equitable access to the education system. Furthermore, family expenditures on out-of-school enrichment activities (Duncan & Murnane, 2011), a lack of support and counseling to take advanced courses, and the accumulated differences caused by widely variable levels of access to educational opportunities have led to the existing gap (Olszewski-Kubilius & Corwith, 2018). Thus, it is critical to direct attention to the aspects of teaching and learning that educators can reform and control. For example, teachers can play significant roles affecting students' outcomes and help lead their students to fulfill their potential (Boyd, 2015; Byers, 2018).

In the field of gifted education, teachers' referrals and nominations for their students are often a first step in the identification procedures for gifted programming (Allen, 2017; Ford et al., 2008; National Association for Gifted Children [NAGC], 2015). Based on a referral from teachers who serve as gatekeepers, students normally progress to the next step of formal testing evaluating

their eligibility for gifted services. In many cases, a lack of teacher referrals for students of color and students from low-income families causes the underrepresentation of students from diverse backgrounds in gifted programs. After analyzing data from the Education Longitudinal Study of 2002 (ELS:2002), Cherng (2017) argued teachers tend to significantly underestimate the ability of students of color compared to their White peers. Kaiser et al.'s (2017) study conversely refuted the claim, stating that teachers can be accurate in their judgement of students regardless of students' backgrounds. Using computer-simulated classrooms, Kaiser et al. (2017) disentangled students' achievement from their ethnicity, allowing teachers to examine whether the achievement of ethnic minority students was being judged predominantly on their ethnic bias; Through four experimental studies, the researchers found a greater judgement accuracy from the teachers for their ethnic minority students indicating that teachers were not biased against underrepresented groups. As such, the literature shows conflicting results on teachers' bias and its effect. Therefore, in these studies, we examined teachers' patterns, influences, and perceptions of referring underserved populations for gifted programs. The findings of these studies, whether the teachers equitably identifying students from low-income or multicultural backgrounds, have the potential to address reducing the achievement gap in gifted education.

Purpose of the Studies

Since teachers' referrals involve evaluation of students through sustained observation, comprehensive features of giftedness can be identified. In 2007, a project called Having Opportunities Promotes Excellence (HOPE) was launched at Purdue University with funds from the Jack Kent Cooke Foundation (Gentry et al., 2015). This 3-year project aimed to help identify giftedness among low-income and ethnically diverse students and serve these students in advanced programs. To do so, the project team created the *HOPE Scale* (Gentry et al., 2015), an instrument

used by teachers to assess the academic and socioemotional characteristics of gifted students. Previous results from Project HOPE served as the foundation for the current studies. This dissertation is comprised of three related research papers investigating the *HOPE Scale* as an equitable measure for identifying underrepresented students for the gifted services. Following are the purpose and research questions for each of these related studies.

Study 1: Validity Evidence for the *HOPE Scale* to Identify Gifted Students from Low-Income and Multicultural Families in Korea

It is important to equitably identify students from different backgrounds, and teachers may play an important role. One way to do so is, when teachers nominate or refer their students for gifted services, they can consider comparing the students to others similar in age, experience, and environment, backgrounds (Lee & Lee, 2015; Peters & Gentry, 2010; Peters & Gentry, 2013). For instance, when teachers evaluate one student from a low-income family, they need to compare this student's outcome or potential with those of other students from low-income families who have similar backgrounds. This is particularly important when identifying students from underserved populations (Peters & Gentry, 2010, 2013). In addition, sound instrument development with strong evidence of validity and continuous updates from the validation study of the instrument is an important factor to support these populations. The research questions in this study are: (1) To what extent does the data from the *HOPE Scale – Korean* version provide evidence of construct validity?; (2) Does the measurement invariance of the *HOPE Scale – Korean* version hold across income (regular vs. low) and cultural subgroups (Korean vs. multicultural)?; and (3) How do teachers perceive the applicability and the items of the *HOPE Scale – Korean* version to identify gifted and talented students?

Study 2: Exploring Individual and Classroom Characteristics on Students' Outcome Scores from the *HOPE* Teacher Rating Scale

Since examining the extent of variation across the teachers with regard to their nomination pattern is the interest of this study, multilevel modeling was used to conduct the analysis. In this study, teachers' seniority (years of teaching), the percentage of students from low-income families, and the percentage of students from multicultural families in each classroom were used to investigate the possible differences in teacher-rating patterns on the *HOPE Scale-Korean* version as level-2 factors, along with the level 1 factors of students' ethnicity and family-income backgrounds. The primary research questions are: (1) To what extent do individual students' *HOPE Scale* scores differ by teacher?; (2) To what extent do individual students' *HOPE Scale* scores differ by individual characteristics such as ethnicity or family-income status within the group?; and (3) To what extent can the variation in the relationship be explained by classrooms' contextual variables (i.e., teachers' years of experience, percentage of multicultural students in the classroom, percentage of low-income students in the classroom)?

Study 3: The Relationship between Students' Academic Achievement and the *HOPE* Teacher-rating Scale: Exploration to Equitably Identify Underrepresented Gifted Students

Researchers have shown the types and combinations of measures affect program diversity, especially with regard to student ethnicity and socioeconomic status (Lakin, 2018; McBee et al., 2014; Peters & Gentry, 2012). To explore the different combination rules that may lead to differences in the number and diversity of students who are identified as gifted, several identification scenarios were be tested. The research questions are: (1) What is the relationship between teacher ratings of gifted behaviors (i.e., Academic, Social) as measured by the *HOPE Scale* and students' outcome performance results (i.e., school-based achievement tests, the number

of subject areas in which they show talent)?; and (2) How do various cutoff percentages and group-specific norms when applied to achievement test results and teachers' *HOPE Scale* ratings affect the identification of students from low-income or culturally diverse families for gifted programming?

CHAPTER 2 (STUDY-1)

Validity Evidence for The *HOPE Scale* in Korea: Identifying Gifted Students from Low-Income and Multicultural Families

Traditional methods of identifying students with gifts and talents have focused on students' cognitive abilities and academic achievement (Lakin, 2016). Although a student's academic ability is a critical factor in the identification process for gifted programming, measuring students' social and emotional characteristics is also important (Lee & Lee, 2015). Teacher rating scales can be an alternative pathway for identifying students with gifts and talents because rating scales enable teachers to evaluate students using sustained observation, allowing them to consider comprehensive features of giftedness such as academic and social components. This approach creates a broader pathway for identification than a one-time exam.

A teacher rating scale might help to identify students with high ability who score poorly on traditional achievement tests (McBee, 2006). Teachers' roles as nominators are critical, in the sense that they can help discover students with gifts and talents from underserved populations; however, as gatekeepers in the identification systems that require teacher nominations, they may also block students from being accepted into the program (Swanson, 2006). Hence, it is important to explore whether teachers rate their students with diverse backgrounds equitably, using a structured teacher rating scale.

Literature Review

Academic and Social Characteristics of Giftedness Perceived by Teachers

Teachers are often the gatekeepers of gifted programs, and their nominations are frequently among the primary factors that determine whether a student will be considered for and participate

in a gifted program (Moon & Brighton, 2008; Ozcan & Kotek, 2015; Szymanski & Shaff, 2013). It is, therefore, important to examine how teachers perceive their students with gifts and talents. In terms of gifted students' academic and social and emotional characteristics, a sample of ten teachers answered the survey questions from Ozcan and Kotek's (2015) study. Regarding academics, the teachers revealed they believed students with gifts and talents are curious, fast learners, easily distracted, and tended to have bad handwriting. Socially, the teachers perceived many of their students with gifts and talents lacked social skills. This study supports the claim that teachers typically think of students with gifts and talents as intellectuals who do well in school and who are intrinsically motivated. On the other hand, teachers rarely consider traditional characteristics such as signs of mental, emotional, creative, or social giftedness when evaluating their students (Lee & Lee, 2015). Results of a study by Szymanski and Shaff (2013) support this theory. They interviewed two second-grade teachers, three third-grade teachers, and one Gifted and Talented Education (GATE) teacher in the Midwest and reported the GATE teacher was concerned about how the majority of the teachers in the school primarily recommended "teacher pleasers" for GATE programs. In addition, the responses from the five classroom teachers revealed their lack of training in identifying students with gifts and talents created confusion over which characteristics to look for in students when making GATE referrals. This evidence suggests that students with gifts and talents may be more accurately and equitably identified if teachers have the proper training to look for diverse traits and characteristics of giftedness.

Lack of Training for Teachers and Their Bias toward Underserved Populations

In research studying how teachers evaluate and interact with students with gifts and talents, the majority of teachers in each study indicated they lacked training in identifying students for gifted services (Abu et al., 2017; Hammerschmidt, 2016; Moon & Brighton, 2008; Szymanski &

Shaff, 2013). Moon and Brighton (2008) suggested a lack of training in identifying students with gifts and talents leads to the under-representation of students of color and those from low-income families. Their data revealed 25% of teachers surveyed said students from a low-income family were less likely than other students to have gifted and talented traits, and 75% of the teachers stated they would not identify a student as gifted if the student had a limited vocabulary.

Elhoweris et al. (2005) conducted another study showing teachers' (83% European American, 11.1% African American, and 5.9% other) biases toward underserved populations. Using vignettes, the researchers surveyed 207 elementary teachers about how they evaluated students and asked them to decide whether each student should be referred to GATE programs. The study consisted of three vignettes: a European American student, an African American student, and a control student, whose ethnicity was not identified to the participating teachers. Even though each student in the scenarios had the same characteristics of giftedness, the researchers found teachers referred the African American student for gifted programs less frequently than they referred the European American student.

Similarly, after surveying 47 teachers, Hammerschmidt (2016) found most of them identified students with gifts and talents based on the characteristics they perceived as relevant to giftedness. In other words, the majority of teachers classified their students as gifted (or not) based on their subjective perceptions of giftedness. Fifty-seven percent of the teachers in this study also agreed their teaching experience influenced their perception and evaluation of gifted qualities and traits. Therefore, teachers not only tend to look primarily at traditional, academic signs of giftedness when evaluating students for gifted programs, but they may also possess personal, subjective perceptions of what defines gifts and talents. This pattern of variation can contribute to different, disproportionate ways of identifying students with gifts and talents.

Teacher Nominations and the Teacher Rating Scales

Educators and experts in gifted education suggest using multiple criteria when assessing student giftedness (Brown et al., 2005; Esquierdo & Arreguin-Anderson, 2012). For instance, Reynolds and Carson (2005) argued standardized assessments may be biased towards some cultures, as well as less accessible to students from different backgrounds. Similarly, Brown et al. (2005) sent out a national, 20-item survey to gifted specialists, classroom teachers, and administrators, asking them how to identify students with gifts and talents (n=2,918). They found educators agreed students should not be evaluated for giftedness using a "restricted approach" (p. 74). Instead, educators supported the notion of evaluating students based on "individual expression, ongoing assessment, multiple criteria, and context-bound expression" (p. 74). A teacher rating scale can be one of those options, because it helps teachers to provide information about their students based on various constructs (Pfeiffer & Jarosewich, 2003). Although concerns exist regarding teachers' biases, teachers can provide alternative and useful information that test-based results cannot (Peters & Gentry, 2010). This led researchers develop teacher rating scales, and the examples are; Gifted and Talented Evaluation Scales (GATES; Gilliam et al., 1996), Gifted Evaluation Scale (GES; McCarney & Anderson, 1998), Scales for Rating the Behavioral Characteristics of Superior Students (SRBCSS; Renzulli et al., 2002), Gifted Rating Scales (GRS; Pfeiffer & Jarosewich, 2003), Scales for Identifying Gifted Students (SIGS; Ryser & McConnell, 2004), Universal Multidimensional Abilities Scales (UMAS; McCallum & Bracken, 2012), Having Opportunities Promotes Excellence (HOPE, Gentry et al., 2015), and The Universal Talented and Gifted Screener (UTAGS; Bracken & McCallum, 2018), in chronical order.

The Development of the *HOPE Scale*

Much research has revealed teachers tend to refer Asian and White students to gifted programs at much greater rates than they refer Black, Latinx, and Native American youth. (McBee, 2006; Plucker & Peters, 2018; Yoon & Gentry, 2009). Teachers also infrequently refer students from low-income families for gifted services. These realities led the Jack Kent Cooke Foundation to support a 3-year project (2007-2010), Having Opportunities Promotes Excellence (HOPE, Gentry et al., 2015). Project HOPE aimed to help identify giftedness among low-income and ethnically diverse students. To do so, the project team at Purdue University created the *HOPE Scale* (Gentry et al., 2015), which is a teacher rating scale that can be used to help identify students for gifted programs. The *HOPE Scale* provides additional information in the identification process beyond simple teacher nominations and ability or achievement test scores. The *HOPE Scale* is used to assess giftedness in students from kindergarten to 12th grade. It consists of eleven items measuring academic (6 items) and social (5 items) components of giftedness. Teachers respond to these items, evaluating their students using a six-point frequency rating scale. Please see Appendix A for a copy of the *HOPE Scale*. The instrument was developed and analyzed using a sample of 5,995 students (Kindergarten through 5th grade) and their 349 teachers in the Midwest, using exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). After revisions, the construct validity of the final version of the scale was supported with good model fit statistics (CFI=.96; GFI=.91); reasonable RMSEA of .10; and internal consistency estimates of .96 for the academic subscale and .92 for the social subscale (Gentry et al., 2015; Peters & Gentry, 2010).

To ensure fair evaluation of students by their teachers, the teachers are encouraged to complete the *HOPE Scale* for every student in their class. What sets the *HOPE Scale* apart from other gifted assessments is that it asks teachers to compare their students to others with similar

backgrounds and experiences to help achieve unbiased results. This practice gives students from underserved groups a better chance of being identified and assessed for giftedness.

Although the *HOPE Scale* shows evidence of being useful for identifying students' giftedness, the research team argued it is also effective when used in conjunction with other gifted assessments and resources, such as achievement and aptitude tests, because no measure is free from error or absolute in its findings (Gentry et al., 2015). They also stressed the academic and social scores should not be combined, but rather used separately, and they recommended calculating local norms, rather than setting up a specific cut-off score. How to calculate local norms is explained in the *HOPE Scale* manual (Gentry et al., 2015). The *HOPE Scale* is a short assessment, which allows teachers to conduct an evaluation quickly for each of their students.

Gifted Education in Korea – Research Setting

South Korea is an ethnically homogeneous society, with an absolute majority of Korean-born residents (approximately 99%). However, foreign immigrants and multicultural families are consistently increasing; the multicultural population has increased about two-and-a-half times over the last 12 years, from 680,000 in 2003 to 1.74 million in 2015 (Sung, 2017). To be specific regarding the student population, in 2014 the Korean Educational Development Institute (KEDI, 2015) provided data showing that South Korean students who qualify as low-income or who are from multicultural backgrounds are 2.89% and 1.07%, respectively. Yet, multicultural students and those from low-income backgrounds are underrepresented in gifted programs, comprising only 0.68% and 0.11% (Lee & Lee, 2015). The number of multicultural students increased to 1.36% (*N*=82,536; KEDI, 2016) in 2015, and according to Lee and Lee (2016), the proportion of students with multicultural backgrounds in school will continue to increase.

Beginning in 2004, Korea began to focus on identifying students with potential giftedness from low-income families, and the Gifted Education Promotion Act (Ministry of Education & Human Resources Development, 2004) was formed in 2005, stating that the gifted program should include students from low-income families (Lee & Eum, 2018). The system then broadened to incorporate students from multicultural families. The 3rd Master Plan (2013-2017) for the Promotion of Gifted and Talented Education (2013, Ministry of Education) emphasized the inclusion of economically disadvantaged and multicultural students with gifts and talents. However, although the Gifted Education Act tried to include underserved populations, Lee and Eum (2018) insisted that the inclusion of underserved population in Korea is still only a basic stage, both in quantity and quality, and there remains a lack of evidence-based research of these populations.

To support underserved populations, teachers' referrals were included among multiple criteria to provide missing information about the students. According to KEDI (2015), after teachers were encouraged to provide referrals within the identification process in 2009, an average of 48% of gifted programs used the teacher referral process; 52.1% for the gifted classes in the regular schools; 55.4% for the Department of Education gifted programs; and 44.2% for the university-based gifted programs (Lee & Lee, 2015). A possible reason for this is teachers failed to refer them, and even if the students were referred, they may not have been identified because they had to go through the same traditional identification procedures without any further benefits (Lee & Lee, 2015). In addition, many Korean schoolteachers continued focusing only on the students' academic ability, not their social abilities, as traits of giftedness (Lee & Lee, 2015; Lee & Lee, 2016). For instance, after studying 177 regular schoolteachers in Korea, Lee and Lee (2016) found these teachers perceived creativity, task-commitment, and cognitive ability to be the greatest

factors when identifying students from underserved populations; whereas, leadership, sociality and self-esteem were seen as inefficient factors. Teachers lacked understanding of underserved populations and were unlikely to nominate them for gifted services. Hence, extra attention is required in terms of students' social and emotional characteristics when identifying students with academic and social gifts (Lee & Lee, 2016).

Purpose of Study

It is necessary and timely to offer equal opportunities to students with gifts and talents, not only from low-income families, but also from multicultural families in South Korea. The *HOPE Scale* may be a helpful tool for identifying those students; however, a validation study of the scale is necessary to ensure its efficacy with these populations. This is also the next step for a call-to-action from Lee and Lee (2015) who asserted despite the implementation of the teacher referrals in identifying students with gifts and talents, underserved populations were still underrepresented in gifted programs, and an absence of instruments exists to help identify them. They also added that a teacher rating scale, created after considering the characteristics of students from multicultural families has not yet developed in South Korea (Lee & Lee, 2015). As it is necessary to compare students with others of similar backgrounds, particularly when identifying students from underserved populations (Lee & Lee, 2015; Peters & Gentry, 2010, 2013), developing appropriate instruments and checking measurement invariance to do so is important (Lakin, 2012). By the same token, Valencia and Suzuki (2001) and Callahan (2007) pointed out that too often, the variables of socioeconomic status (SES) and race were not examined separately. This confounded attribution of effects to variables. They suggested an instrument validation study should include income level as a variable, separate from race, ethnicity, and gender, to enable clearer interpretation of results. The research questions that guided this study were:

RQ 1: To what extent does the data from the *HOPE Scale – Korean* version provide evidence of construct validity?

RQ 2: Does the measurement invariance of the *HOPE Scale – Korean* version hold across income (regular vs. low) and cultural subgroups (Korean vs. multicultural)?

RQ 3: How do teachers perceive the applicability and the items of the *HOPE Scale – Korean* version to identify gifted and talented students?

As this is the first validity investigation on the *HOPE Scale – Korean*, and the investigation is exploratory in nature in particular for the last question, we did not specify any prior hypotheses regarding expected findings for each research question in this study. However, given the validity and invariance results from the studies with U.S samples (Gentry et al., 2015; Peters & Gentry, 2010, 2013), we expected to find similar results and used their results as underlying theoretical model to guide our analyses. Thus, we hypothesized the two-factor model would hold for the construct validity evidence and tested the model as an initial step. Similarly, we also expected that measurement invariance would hold across tested groups for the *HOPE Scale* to be one of the effective tools to identify underrepresented students in Korea.

Method

Participants

The participants of this study were public elementary school teachers and their third through sixth grade students with varied socioeconomic and ethnic backgrounds. Although the *HOPE Scale* was first designed to identify underserved students with gifts and talents, it is also effective in identifying students with gifts and talents from non-underrepresented groups (Peters

& Gentry, 2010, 2013). The sample in this study, therefore, was not limited to underrepresented students. Based on their daily observations, 55 teachers from nine different elementary schools in Seoul, South Korea, completed *HOPE Scales* on their students (*n*=1,157). Table 1 describes the demographic information and Table 2 shows the descriptive statistics of the *HOPE Scale*. It is noteworthy that the students from culturally diverse background in our sample account for 14%, which reflects more than the regular composition of multicultural populations (approximately 1-2%) in Korea. In addition, the students who overlapped in two categories of low-income family and multicultural backgrounds comprised 2.85% of the total sample (n=33), and we used their demographic information separately to answer our second research question.

Table 1. Demographic Information of the Participants

Variables	Category	Student Frequency (%)
Grade level	Grade 3	266 (23) from 13 classes
	Grade 4	320 (28) from 15 classes
	Grade 5	277 (24) from 13 classes
	Grade 6	294 (25) from 14 classes
Ethnicity	Korean	990 (86)
	Multicultural	167 (14)
Family-income	Regular-income	941 (81)
	Low-income	216 (19)
Gender	Male	592 (51)
	Female	565 (49)
	Total	1157

Table 2. Descriptive Statistics of *HOPE* Scale – Korean

Construct	Items	Response Percentage (%)						Mean (*SD*)	Skewness	Kurtosis
		1	2	3	4	5	6			
Academic (6 items)	Performs or shows potential for performing at remarkably high levels. (Item 1)	5.36	11.58	25.93	25.24	20.14	11.75	3.78 (1.35)	-.12	-.68
	Is eager to explore new concepts. (Item 6)	10.20	15.59	27.48	21.00	17.20	7.52	3.41 (1.41)	.05	-.80
	Exhibits intellectual intensity. (Item 7)	11.67	18.41	27.92	19.10	14.43	8.47	3.32 (1.44)	.17	-.80
	Uses alternative processes. (Item 9)	6.40	19.19	34.23	20.92	13.83	5.45	3.33 (1.26)	.24	-.48
	Thinks “outside the box.” (Item 10)	6.83	19.88	38.20	21.35	9.85	3.89	3.19 (1.18)	.30	-.14
	Has intense interests. (Item 11)	5.62	14.69	29.30	21.43	16.25	12.62	3.66 (1.40)	.07	-.77
Social (5 items)	Is sensitive to larger or deeper issues of human concern. (Item 2)	10.63	18.06	30.34	19.45	14.52	7.00	3.30 (1.39)	.17	-.70
	Is self-aware. (Item 3)	2.94	9.51	34.23	24.89	20.66	7.78	3.74 (1.20)	.04	-.49
	Shows compassion for others. (Item 4)	3.20	10.11	25.93	27.57	21.18	12.01	3.89 (1.28)	-.12	-.60
	Is a leader within his/her group of peers. (Item 5)	21.18	20.14	21.61	16.34	12.27	8.47	3.04 (1.57)	.33	-.96
	Effectively interacts with adults or older students. (Item 8)	9.85	15.99	33.54	21.35	13.57	5.70	3.30 (1.32)	.14	-.53

Instrument Adaptation Process: *HOPE Scale - Korean* Version

The original *HOPE Scale* was translated from English into Korean in several steps (Geisinger, 1994; Hambleton, 2001). It was first translated by the first author into Korean, and the translated results were compared with the original instrument by three bilingual graduate students (i.e., English- and Korean-speaking) majoring in education. Based on their suggestions, the initial *HOPE Scale-Korean* version was drafted, and it was translated back into English by another bilingual graduate student in gifted education, who is familiar with the nature of the field. The edited English version was then shared with a panel of five graduate students, who only speak English. They compared the back-translated, English version of the instrument with the original *HOPE Scale,* and we discussed whether major differences existed between those two versions. Minor edits were made based on the feedback, and the final *HOPE Scale – Korean* version was created.

The ethnic group classification in the demographic section of the *HOPE Scale*—identifying seven different ethnic groups: American Indian/Alaska Native, Asian, Black or African American, White, Native Hawaiian or Other Pacific Islander, Mixed Race, and Hispanic/Latino—did not effectively represent the present-day multicultural population in South Korea. In South Korea, multicultural families are defined as a family that is created by the international marriage of a Korean citizen with a partner from a different country (Lee & Lee, 2016). Given the geographic location of South Korea, the majority of the multicultural families are comprised of a Korean person and their partner, who also comes from another Asian country. They would therefore still be grouped as Asian if the original categories remained. Therefore, with the consideration to the research setting, we divided the ethnicity section of *HOPE Scale – Korean* into two main groups (Korean or multicultural), and then subdivided the multicultural group into five sub-categories: White, Black, Hispanic/Latino, Asian, and Other. In addition, the Asian group is divided more into

detail, with five geographical sub-regions: Central Asia, East Asia, South Asia, Southeast Asia, and Western Asia (see Appendix B for a copy of the *HOPE Scale – Korean*).

Data Analyses

Traditionally, general linear models, including a *t-test* or Analysis of Variance (ANOVA), which do not involve any latent variables, have frequently been used to test for equivalence of mean scores between different groups (Thompson & Green, 2006). The observed mean difference may be caused by systematic differences in attributes among groups, test items that systematically yield higher or lower scores for specific groups, or biased test items. Therefore, simple equivalence of the means does not exclude the possibility that the mean is the same due to a bias toward a specific group. Alternatively, a higher group mean does not indicate the yielded scores are systematically higher for a particular group due to biased items or the differential structure of the constructs. For these reasons, we conducted more in-depth analyses in this study regarding measuring the structural invariance of constructs among different subgroups (Lakin, 2012; van de Vijver & Tanzer, 1997). To answer the first research question, we conducted a Confirmatory Factor Analysis (CFA) to support the construct validity of the data for the *HOPE Scale – Korean* as prior research showed sufficient evidence supporting the theoretically defined two-factor structure in the original *HOPE Scale* (see Gentry et al. (2015) and Peters & Gentry's (2010, 2013) study for the details of the development and revision procedures of the *HOPE Scale*). In CFA, a robust maximum likelihood (RML) estimator was applied, because it is less dependent on the assumption of multivariate normal distribution of continuous variables in comparison with ML estimation (Li, 2016; Yuan & Bentler, 2000). RML provides estimators with statistical corrections to chi-square statistics and standard errors, and it has been widely used in CFA models (Li, 2016).

Measuring invariance establishes valid information across multiple groups, by examining the extent to which items have equal meaning among the sample groups (French & Flinch, 2006). For instance, if certain test items are more advantageous for one particular group, the instrument may not yield equally valid data for all populations (Osterlind & Everson, 2009). Therefore, a Multi-group CFA (MCFA) was used as the next step to analyze invariance between students' ethnic groups and their family's socioeconomic status for the two-group comparison which would answer the second research question. MCFA is important, because the teacher rating scales including *HOPE Scale* are often applied to various subgroups without considering possible invariance issues (Brown, 2006).

In MCFA, data from 216 students from low-income families and from 167 students from multicultural families were used (see Table 1) with equal numbers of their higher-income and Korean counterparts. As parameter estimates and chi-square value are sample size dependent and sensitive to sample size (Fischer & Karl, 2019; Ullman, 2007), Fischer and Karl (2019) argued that "unless sample sizes are equal, it might be difficult to determine which samples and items are problematic when examining an overall poorly fitting multi-group model" (p. 15). Similarly, Chen (2007) and Kaplan and George (1995) stated that when the sample size is increasingly disparate among groups, it is difficult to detect factor mean differences and goodness of fit indexes across groups. One simulation study conducted by Yoon and Lai (2018) similarly revealed severely unbalanced group size (e.g., one group sample size is more than twice larger than the other) affects the invariance studies results; it leads to incorrect conclusions of invariance due to fit function in multiple-group factor analysis, weighting by group sample size. This indicates that violations of invariance can be masked, and researchers may find false invariance when there are large imbalances between groups (Yoon & Lai, 2018). Simply put, we randomly selected 216 students

from regular-income families and another 167 Korean students, creating equal sample sizes and allowing for direct chi-square comparisons in the invariance testing. Each model was then tested using several steps in an increasingly restrictive manner (Brown, 2006; Putnick & Bornstein, 2016).

To test measurement invariance, Putnick and Bornstein's (2016) four steps, measuring (a) configural; (b) metric (weak factorial); (c) scalar; and (d) residual invariance, were used as a guideline. Other studies in the field of gifted education also followed similar guidelines when studying evidence of measurement invariance across different cultural or socioeconomic status (e.g., Lakin (2012), Peters & Gentry (2010)). According to the procedures suggested by the authors (Putnick & Borstein, 2016, pp.75-77), (a) configural invariance, also known as invariance of model form, is the least stringent step designed to check whether the constructs have the same pattern of free and fixed loadings. If configural invariance is supported, testing (b) metric invariance (equivalence of the item loadings on the factors) is the next procedure. This step tests whether each item contributes to a similar degree across groups to the latent construct, and it is tested by constraining factor loadings. The metric invariance results are compared to the configural invariance model to determine its model fit; if the metric invariance model fit is not significantly worse than configural invariance model, it implies metric invariance is supported, because the results of constraining the loadings across groups do not significantly affect the model fit. If metric invariance is supported, (c) scalar invariance (equivalence of item intercepts) would be tested for the metric invariant items. This step is tested by constraining the item intercepts to be equivalent among the groups. In this process, if the scalar invariance model fit is not significantly worse than metric invariance model fit, it supports the scalar invariance model, because the results, by constraining the items across groups, do not significantly affect the model fit. On the other hand, if the scalar invariance model fit is worse than the results of the metric invariance model fit, it

implies that at least one item intercept is different across the groups. If scalar invariance is supported, the last step would be checking (d) residual invariance, which tests whether error variance (measurement error) or some of specific variance not shared with factor is similar between groups. However, although this is one of the full factorial invariance components, residual invariance is not a prerequisite, as it is not part of the latent factor (Vandenberg & Lance, 2000) and many researchers tend to omit this step (Putnick & Bornstein, 2016). In terms of the parsimony, it is also important to note that a more restrictive model typically has a higher chi-square value and degrees of freedom as there are fewer parameters to estimate which yields worse fit. If the more statistically parsimonious model is selected, it may erroneously assume within-group variability is equal across the tests (Horn & McArdle, 1992; Sprague et al., 2017). Therefore, we would judge the model based on the theoretical underpinning of the measurement with multiple indices (Williams & Holahan, 1994).

In both CFA and MCFA, diverse fit indices were reported to describe the model (Hu & Bentler, 1999); absolute fit indices (e.g., x^2, AIC, GFI), relative fit indices (e.g., IFI, NFI, NNFI (also known as TLI)), and noncentrality-based indices (e.g., RMSEA, CFI). In terms of the chi-square value, generally a large value indicates a lack of model fit, however, this is affected by the sample size; therefore, different fit indices are encouraged to be examined when explaining a model fit (Garver & Mantzer, 1999; Kline, 1998). In this sense, AIC is considered as a parsimonious adjusted fit index as it is penalized with the number of parameters to estimate, and considered as one of the most effective indices to evaluate the model fit (Williams & Holaham, 1994).Although there is no rule of thumb for each index, the value of fit indices such as NFI, NNFI, CFI, GFI, and IFI greater than .90, and RMSEA value near or less than .08 represents a good fit (Browne & Cudeck, 1993; Garver & Manter, 1999). Throughout the analyses of CFA and MCFA,

we used LISREL 9.3 (Jöreskog, n.d.) to estimate parameters and it conveniently suggests the model fit index and the orders that will improve the model fit if we accept the correlated errors between the paired items.

Following these quantitative analyses, we added the insights of six teachers who voluntarily participated in follow-up interviews regarding the *HOPE Scale – Korean* items and its applicability when identifying students with gifts and talents from diverse backgrounds. This part was designed to answer the third research question. The purpose of conducting follow-up interviews was to understand in detail (Gubrium & Holstein, 2001) how teachers felt the effectiveness of the scale which might not be captured through the quantitative analyses. This part was designed to answer the third research question and we specifically asked two main questions to the teachers (a) how they perceived using *HOPE Scale – Korean* as an alternative tool to help identify gifted students and (b) which item(s) of the 11 items on the scale they found helpful and which items they questioned. During the interview process, the teachers were asked to describe two main questions we prepared as a semi-structured interview, however, we additionally asked their opinions about non-academic characteristics of gifted students because the items teachers felt unsure during the interview were all Social items.

In the recruiting process, we invited 55 teachers who completed the *HOPE Scale – Korean* for their students and six teachers (10.91%) replied that they were willing to participate. In terms of the sample size in the phenomenological qualitative studies, the expected sample size of the interviewees ranges approximately 6 (Denzin & Lincoln, 1994), 6-8 (Kuzel, 1999), and 6-10 (Morse, 2000); however, as far as justifying the sample size of the interviewees, most qualitative methodologists openly have recognized the lack of standards for sample size and little defense exists based on the practices (Marshall et al., 2013). Although some qualitative research

methodologists describe general guidelines for the number of participants for the interviews (Subramanian & Peslak, 2010), Marshall et al. (2013) argued "the guidelines vary from methodologist to methodologist and sometimes the same methodologist has provided different ranges at different points in time" (p.13). As such, there is a subjective nature of sample size in qualitative studies (Marshall et al, 2013) and as "sample size depends on what you want to know, the purpose of the inquiry, what's at stake, what will be useful, what will have credibility, and what can be done with available time and resources" (Patton, 2002, p. 243), we considered the number of teacher participants who shared their opinions was acceptable. The interviews were conducted over the phone for average 20 minutes and we informed the teachers the conversations would be recorded. As a member-checking process, the interview data were transcribed and sent to each interviewee to review, and provide feedback concerning its accuracy. This process increased the trustworthiness of the data (Creswell & Miller, 2000). Once the responses were coded under established themes, they were compared with the quantitative findings. Based on the teachers' responses, we created four themes (a) pros of *HOPE Scale – Korean*, (b) cons of *HOPE Scale – Korean*, (c) item reviews, and (d) social and emotional characteristics of gifted students; using data from teachers' opinions regarding the *HOPE Scale – Korean*.

Results

Evidence of Construct Validity of the Data: Confirmatory Factor Analysis (CFA)

To address Research Question 1, *To what extent does the data from the HOPE Scale – Korean version provide evidence of construct validity?*, we used CFA to investigate the evidence of construct validity of the Korean data set (see Table 3 for the correlation matrix of the 11 items for overall sample). We estimated internal consistency alpha reliabilities for the data with estimates

of .916 for the academic scale and .838 for the social scale; acceptable estimates for an affective measure (McCoach et al., 2013). Table 4 shows factor loadings on each construct, ranged from .482 to .875; all within acceptable range as researchers reasonably expect to see the factor loadings with magnitudes between .40 and .90 in practice (Guadagnoli & Velicer, 1988; McNeish et al., 2018). The goodness-of-fit indices of the original model results are: NFI = .879, NNFI = .850, CFI = .883, IFI = .883, GFI = .837, and AIC=12105.391 with RMSEA = .147. These initial results required revision of allowing correlated errors between items to achieve a better fit. According to the modification indices for theta-delta index, the total chi-square statistic of the model could be decreased by allowing the errors to correlate between item #6 (*Is eager to explore new concepts*) and item #7 (*Exhibits intellectual intensity*), each from the academic construct (See Table 5 for the modification process). The results improved after the modification (NFI = .919, NNFI = .9, CFI = .924, IFI = .924, GFI=.892, and AIC=11731.490), however, the value of RMSEA at .120 was still not in the acceptable range (Browne & Cudeck, 1993; Chen et al., 2008). After the first revision, three more revisions were sequentially conducted in the same manner allowing correlation between errors of the following item pairs: item #1 (*Performs or shows potential for performing at remarkably high level*) and item #7 (*Exhibits intellectual intensity*) from the academic construct; item #9 (*Uses alternative processes*) and item #10 (*Thinks "outside-the-box"*) from the academic construct; and item #3 (*Is self-aware*) and item #4 (*Shows compassion for others*) from the social construct.

Table 3. Inter-item Correlation Matrix for Overall Sample

	Aca1	Aca6	Aca7	Aca9	Aca10	Aca11	Soc2	Soc3	Soc4	Soc5	Soc8
Aca1	1.000										
Aca6	.670	1.000									
Aca7	.672	.897	1.000								
Aca9	.626	.733	.708	1.000							
Aca10	.534	.594	.570	.716	1.000						
Aca11	.717	.580	.560	.555	.559	1.000					
Soc2	.621	.725	.729	.650	.522	.515	1.000				
Soc3	.559	.619	.594	.613	.469	.467	.645	1.000			
Soc4	.291	.348	.325	.406	.274	.211	.417	.509	1.000		
Soc5	.544	.628	.593	.617	.488	.440	.565	.553	.367	1.000	
Soc8	.447	.541	.485	.616	.468	.385	.515	.519	.463	.590	1.000

Note. Aca indicates Academic item and Soc indicates Social item; The correlation between two constructs is .78.

Table 4. Factor Loadings of the *HOPE* Scale – Korean

	Academic (α=.916)	Social (α=.838)
Aca1	.761	
Aca6	.875	
Aca7	.849	
Aca9	.843	
Aca10	.684	
Aca11	.655	
Soc2		.823
Soc4		.755
Soc5		.482
Soc6		.742
Soc8		.681

Table 5. Indices of Model Fit (Modification Process)

	Original	Revision1 Element (A7,A6) of *TD*	Revision2 Element (A1,A11) of *TD*	Revision3 Element (A9,A10) of *TD*	Revision4 (Final) Element (S4,S3) of *TD*
Chi-Square	1120.7 (*df*=43)	744.881 (*df*=42)	527.462 (*df*=41)	423.052 (*df*=40)	361.084 (*df*=39)
AIC	12105.391	11731.490	11516.071	11413.661	11353.693
RMSEA	.147	.120	.101	.091	.085
NFI	.879	.919	.943	.954	.961
NNFI	.850	.900	.929	.943	.951
CFI	.883	.924	.947	.958	.965
IFI	.883	.924	.947	.958	.965
GFI	.837	.892	.919	.934	.942

Note. P<.001; A indicates Academic item, S indicates Social item, and *TD* indicates the modification indices for *theta-delta*; Each revision allows the errors to be correlated between items and the final model contains four pairs.

The final version of the model yielded a chi-square value of 361.084 (*df* = 39), which is 32.23% smaller than the chi-square from the original model. According to the goodness-of-fit statistics, the indices were all an excellent fit, with the following results: NFI=.961, NNFI=.951, CFI=.965, IFI=.965, GFI=.942, and AIC=11353.693. The proposed criteria of these fit indices are close to .95 (Hu & Bentler, 1999; Perry et al., 2015). The RMSEA value was .0845, which is less than .1 and within an acceptance range (Browne & Cudeck, 1993; Chen et al., 2008). Figure 1 depicts the final model. Similar to the interfactor correlation results between the two constructs in the original *HOPE Scale* (*r*=.90; Gentry et al., 2015), our data also show a high interfactor correlation (*r*=.96). However, as Gentry et al. (2015) indicated, we kept two correlated factors on the instrument "because there are only two factors, a higher-order factor would not help explain the data, and a single factor in place of the two factors does not fit the theory" (p.9). In addition, it is important to note that we assumed the measured observed variables would be affected by the measurement errors because none of the observed variables are perfect or free from measurement errors. Due to the measurement errors, the correlation tends to be attenuated when investigating

interrelationship between variables. Our data showed that the correlation between the academic and social factor was .78. However, the intercorrelation between factors (the latent factors) in the final model, which is measurement error-free, yielded a higher correlation (r=.96).

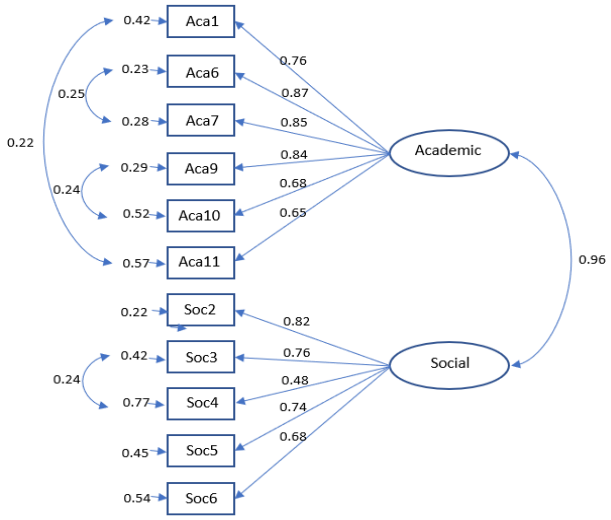

Figure 1. Final CFA model

Multi-group Confirmatory Factor Analysis (MCFA)

Once the final model was built from the general CFA, measurement invariance testing was conducted to evaluate the equivalence of different parameters for students with different race and socioeconomic status to answer the Research Question 2: *Does the measurement invariance of the HOPE Scale – Korean version hold across income (regular vs. low) and cultural subgroups (Korean vs. multicultural)?* The results of comparing the two groups are presented in Tables 6 and 7.

Ethnicity (Korean versus Multicultural)

The chi-square value for students from multicultural backgrounds (χ^2 = 111.568) was slightly higher than that of Korean students (χ^2 = 99.481). RMSEA, NFI, NNFI, CFI, and IFI values were

also very close across the two groups when analyzed separately, with a slightly worse fit for the multicultural group (see Table 6). However, the differences between the indicators above were all less than .01, indicating that the groups have very similar results. In general, the fit indices were nearly identical when indicating the model, which fits both groups well. This simultaneous model or configural invariance test was used as a baseline for the purposes of comparing the extent of the measurement invariances (RMSEA=.101, NFI=.964, NNFI=.967, CFI=.977, IFI=.977, AIC=3057.551). The test of the model with equal factor loadings (metric invariance testing) resulted in a non-significant difference in the chi-square value (ΔX^2= 22.227, *df*=9, *p* =.001, RMSEA=.099, NFI=.960, NNFI=.968, CFI=.975, IFI=0.975, AIC=3061.798). This means that adding equality constraints on factor loading across the groups did not result in a significant increase of the chi-square, meaning model fit was not affected. Thus, the metric invariance holds for the two groups. The following test results with equal indicator intercepts (scalar invariance) were also non-significant compared to the metric invariance (ΔX^2=0.565, *df*=3, *p* =.001), which means the two ethnic groups had similar item intercepts (RMSEA=.098, NFI=.960, NNFI=.969, CFI=.975, IFI=.975, AIC=3056.343).

Table 6. Measurement Invariances Tests for Korean Versus Multicultural Students (167 students each for the sample)

	X^2	*df*	ΔX^2	Δdf	AIC	RMSEA	NFI	NNFI	CFI	IFI
Single group solutions										
Korean	99.481*	39			1675.811	.096	.927	.935	.954	.954
Multicultural	111.568*	39			1382.372	.106	.926	.930	.950	.951
Measurement invariance										
Configural: Equal form (Simultaneous)	210.417*	78			3057.551	.101	.964	.967	.977	.977
Metric: Equal factor loading (Full *LX*)	232.644*	87	22.227	9	3061.798	.099	.960	.968	.975	.975
Scalar: Equal Indicator intercepts (Full *PH*)	233.209*	90	.565	3	3056.343	.098	.960	.969	.975	.975
Residual: Equal indicator error variance (Full *TD*)	271.858*	105	38.649*	15	3064.992	.098	.953	.969	.971	.951
Residual: Partial error variance (Partial *TD* of Academic)	243.830*	96	10.621	6	3054.964	.096	.958	.970	.974	.974
Residual: Partial error variance (Partial *TD* of Social)	255.085*	95	21.876*	5	3068.219	.101	.956	.968	.972	.972

* p<.001; LX defines the loadings of *x* on lambda-*x*, PH defines the covariances of the lambda-*x* variables (phi), and TD defines the covariances of the theta-delta variables; ΔX^2 is compared with the previous model at α=.001 level.

Table 7. Measurement Invariances Tests for Regular Versus Low-income Students (216 students each for the sample)

	X^2	*df*	ΔX^2	Δdf	AIC	RMSEA	NFI	NNFI	CFI	IFI
Single group solutions										
Regular Income	89.170*	39			2058.156	.077	.950	.959	.971	.971
Low Income	95.104*	39			2153.751	.082	.946	.954	.967	.968
Measurement invariance										
Configural: Equal form (Simultaneous)	183.848*	78			4211.480	.079	.975	.980	.986	.986
Metric: Equal factor loading (Full *LX*)	190.858*	87	7.010	9	4200.491	.074	.975	.982	.986	.986
Scalar: Equal Indicator intercepts (Full *PH*)	200.729*	90	9.871	3	4204.362	.076	.973	.972	.985	.985
Residual: Equal indicator error variance (Full *TD*)	210.610*	105	9.881	15	4184.243	.068	.972	.985	.986	.986

* p<.001; LX defines the loadings of *x* on lambda-*x*, PH defines the covariances of the lambda-*x* variables (phi), and TD defines the covariances of the theta-delta variables; ΔX^2 is compared with the previous model at α=.001 level.

Although an equal indicator of error variance (residual invariance) is not often investigated, assuming that the equal errors are not associated between different groups (Brown, 2006; Putnick & Bornstein, 2016), we conducted the test to further analyze structural parameters. The test of equal indicator error variances showed a significant chi-square value difference, meaning that non-equivalence (error variances across the groups) existed (ΔX^2= 38.649, *df*=15, p = .001) with RMSEA=.098, NFI=.953, NNFI=.969, CFI=.971, IFI=.951, AIC=3064.992. When examining the partial error variance of each academic and social factor as a follow-up step, the social factors showed a significant chi-square increase (ΔX^2= 21.876, *df*=5, p =.001) with RMSEA=.101, NFI=.956, NNFI=.968, CFI=.972, IFI=.972, AIC=3068.219; whereas, the academic factors had no significant chi-square value increase (ΔX^2 =10.621, *df*=6, p =.001) with RMSEA=.096, NFI=.958, NNFI=.970, CFI=.974, IFI=.974, AIC=3054.964. This indicates that the two groups are likely to have equal error variance (residual invariance) within academic factors but not within social factors. Yet, it is important to note that an indicator of equal error variance is stringent, therefore not often supported or investigated as indicated above (Brown, 2006; Putnick & Bornstein, 2016). Put simply, the tests found the overall variation between the two ethnic groups is very similar.

Income (Regular versus Low-income)

The students' family income status was used to compare measurement invariance results as well. The chi-square value for students from low-income families (χ^2=95.104) was slightly higher than that of students from regular income families (χ^2=89.17). RMSEA, NFI, NNFI, CFI, and IFI values were also very close across the two groups, when analyzed separately, with a slightly worse fit for the low-income group (see Table 7). However, the differences in the

indicators above were all less than .01, indicating that the groups had very similar results. In general, the fit indices, which were also nearly identical, indicate the model fits both groups well.

From the baseline simultaneous model (configural invariance) with RMSEA=.079, NFI=.975, NNFI=.980, CFI=.986, IFI=.986, and AIC=4211.480, the test of equal factor loadings (metric invariance) resulted in a non-significant difference of the chi-square value (ΔX^2=7.010, *df*=9) at the *alpha*=.001 level (RMSEA=.074, NFI=.975, NNFI=.982, CFI=.986, IFI=.986, AIC=4200.491). This means that the assumption of equal factor loadings holds the same for the two groups. The following tests that had equal indicator intercepts (scalar invariance) were also not significant (ΔX^2=9.871, *df*=3), *p*=.001), which means the two ethnic groups had similar item intercepts (RMSEA=.076, NFI=.973, NNFI=.972, CFI=.985, IFI=.985, AIC=4204.362). In income-group differences, even the equal error variance results (residual invariance) showed a non-significant chi-square value difference (ΔX^2=9.881, *df*=15, *p*=.001), implying equivalence of indicator error variances across the groups (RMSEA=.068, NFI=.972, NNFI=.985, CFI=.986, IFI=.986, AIC=4184.243). This reveals that the two income groups have an almost identical variation.

Teachers' Perception of the *HOPE Scale – Korean* from the Follow-up Interview

After completing the *HOPE Scale – Korean* for their students, six teachers volunteered to share their thoughts about the instrument to address research question 3: *How do teachers perceive the applicability and the items of the HOPE Scale – Korean version to identify gifted and talented students?* We asked teachers how they perceived applying the *HOPE Scale – Korean* as an additional identification tool and which items they thought were most or least helpful in selecting students for the gifted program (see Table 8).

Table 8. Demographic Information of the Interviewees (n=6)

Participant code	Age	Gender	Teaching grade	Have received professional development of gifted education?	*HOPE Scale* items teachers questioned
A	30s	Female	Grade 4	No	2, 3, 4, 8
B	20s	Female	Grade 6	Yes	8
C	20s	Male	Grade 3	No	2
D	20s	Female	Grade 6	Yes	2, 8
E	30s	Male	Grade 4	No	1, 2, 11
F	20s	Male	Grade 5	No	2, 3, 5

Note. The teachers are all from different schools and they are all Korean with no multicultural background.

Pros of HOPE Scale – Korean

Every teacher came up with pros and cons regarding the *HOPE Scale – Korean*. In terms of the merits of the *HOPE Scale – Korean*, two teachers (33%) pointed out the limitations of the traditional way of identifying students for gifted programs. For instance, one stated, "Obviously, not one tool is perfect in identifying gifted students. I think it is important to listen to teachers' voices, based on their regular observation," (Teacher E – 04/06/19) while another shared, "I believe that the traditional methods of identifying gifted students are limited in their effectiveness, so I welcome these additional tools to help us better evaluate" (Teacher F – 04/07/19). Based on the limitation of using a single instrument and the importance of applying an alternative pathway, three other teachers (50%) mentioned the advantage of *HOPE Scale – Korean* more specifically. They focused on how it would affect identifying students from multicultural backgrounds or from lower socioeconomic status, which was the original goal of the development of the *HOPE Scale* (Gentry et al., 2015). One teacher indicated that, "In other questionnaires, multicultural and/or low-income students are evaluated by comparing them to [regular] students from average backgrounds. However, the *HOPE Scale* asked us to compare a student to others from a similar

background, which are the most distinctive characteristics" (Teacher D – 04/06/19). Similarly, two others pointed out that, "The scale allows us [teachers] to identify students whose potentials are often overlooked because of external environmental factors" (Teacher C – 04/05/19), and therefore, the teachers have a chance to identify and refer students who are not prepped by tutors or test-trained, but who possess high innate potential.

> Many of the students who come to the program appear to be trained to become 'gifted'. It seems they were prepared by tutors, or given private after-school classes, with the specific goal of being accepted into these programs. I heard that they utilized study guides that are made specifically for the admission exam. If we do not have tools like *HOPE Scale – Korean*, it would be extremely difficult to recognize real "gifted" students among the average student body who did not have the chance to attend those prep schools. (Teacher B – 04/05/19)

The *HOPE Scale – Korean* can work as an identification tool that can more equitably identify students from multicultural, and/or low-income backgrounds. Using the *HOPE Scale – Korean*, teachers can refer a student who may not receive a high enough score on the traditional intelligence or achievement tests, but who may have innately high potential as compared to other students with similar backgrounds.

Cons of HOPE Scale – Korean

The teachers also indicated some drawbacks and concerns about the *HOPE Scale – Korean* in terms of its practicability and item characteristics that may not represent students' giftedness. Half of the teachers indicated that they were unsure of how it works as an effective tool in identifying students with gifts and talents, based on the practicability issue from the teachers' perspective. As *HOPE Scale – Korean* asks teachers to fill out a form for every student in their classroom, teachers argued doing so might overburden them. One stated that,

> If the questionnaire is used by the entire school, we would be able to see the average score or calculate school or group specific norms for all the students and also be able to identify who achieved high scores in the scale. However, if this is

> used for admissions into programs by district level or university-based programs outside of school, it takes a lot of time to fill these out for each student in our class if they still need all students' information. (Teacher D – 04/06/19)

This teacher added that it would be helpful if there were certain guidelines, or a possible cut-off score suggested. Another teacher similarly stated that she has questions on how it would be effective, when she is using it for recommending students for the out-of-school gifted program. She worried that, "Teachers may be concerned whether a student will get accepted, because they have initial low scores compared to other students. Plus, the students who would be willing to join the out-of-school programs tend to be higher-achievers" (Teacher B – 04/05/19). This indicated that the teacher perceived there would be no distinction among the referred students.

The remaining teachers also shared how they were uncertain about other aspects of the instrument, mostly regarding the social characteristics of students with gifts and talents. One of the teachers indicated,

> We can refer the students based on it [*HOPE Scale – Korean*], but not as a definitive indicator. The characteristics of each student vary significantly. For example, one can be a leader, one cannot be. Even autistic students can be gifted, but these students may not socialize, do not become leaders, and may not emotionally relate with others very well. Because these students exist, who are intellectually gifted but not socially, I do not think you can conclusively categorize the students as gifted or non-gifted. You can refer to it as a supplement, but I do not think its results should be heavily used. (Teacher A – 04/01/19)

This teacher's opinion is that gifted students' social and emotional abilities may vary, depending on each student's personal traits, and there could be concerns when using scores from the *HOPE Scale – Korean*. This goes along with the *HOPE Scale* manual (Gentry et al., 2015), in which the developers clearly specified that users should not use the sum of the two constructs as a total *HOPE Scale* score; rather, they recommended using the academic and social scores independently.

Item Reviews of the HOPE Scale – Korean

When teachers were asked which items they thought would be most or least helpful in identifying students with gifts and talents, all teachers surprisingly identified the social items as questionable (see Table 8). Only one teacher pointed out the academic items (item #1 *"Performs or shows potential for performing at remarkably high levels"* and item #11 *"has intense interests"*), indicating that they are too broad and vague, but even this teacher included other social items as uncertain. When explaining why they thought certain items were appropriate or questionable, the teachers provided reasons why they were uncomfortable with these items and stated that they liked the rest. Hence, the following section focuses on items about which the teachers raised concerns.

Five (83%) out of six teachers pointed out they were unsure about item #2 (*Is sensitive to larger or deeper issues of human concerns*). One teacher stated, "I do not think gifted students have particularly deep interests in the issues of our society or human concerns. Most of the gifted students I know are more interested in science or math" (Teacher D – 04/06/19). This teacher may not have understood general characteristics of students with gifts and talents but focused on talents on specific domain area. Others, however, similarly stated that these characteristics differ greatly among students. Half of the teachers also pointed out that item #8 (*Effectively interacts with adults or older students*) should not be used as a decisive indicator when describing gifted students' characteristics.

Teachers stated that, because of their high intelligence, students with gifts and talents may get along with others who are older than them; however, it is hard for teachers to observe them talking in person. They added that there are not many chances for students to meet older people at school except for the teachers, so, therefore, it is unlikely for teachers to observe and evaluate the students regarding this item. Item #3 (*Is self-aware*) and item #4 (*Shows compassion for others*) were questioned by teachers of two and one, respectively, for similar reasons. Teachers

commented that it is difficult to observe and evaluate these items, because they are a little abstract. In general, teachers identified the social items as questionable, because there are many exceptions to these characteristics. Based on their arguments, it could be summarized that although some students with gifts and talents do reveal these social characteristics, they do not necessarily apply to all gifted students, at least as the teachers perceive it.

Social and Emotional Characteristics of Gifted Students

As all interviewees questioned some of the social items from the *HOPE Scale – Korean*, we asked for more detail about their perceptions of the social and emotional characteristics of gifted students and their thoughts about including these in the general gifted identification process. The characteristics varied; teachers mentioned social and emotional traits such as resilience, passion, collaboration skills, goal orientation, concentration, and confidence. Although the social and emotional characteristics may differ from student to student, a majority of teachers (83%, *n*=5) stated that it is ideal to incorporate social and emotional characteristics when identifying or educating students with gifts and talents. One teacher, who thinks it is ideal but is uncertain about the idea, stated because no fixed social characteristics exist, "it would not be reliable to use social characteristics when identifying gifted students. However, I think it is a good idea to study the emotions and social characteristics once they are selected into the program" (Teacher A – 04/01/19). Other teachers (*n*=4), however, indicated that it is important to move beyond the traditional way of identifying students with gifts and talents to a new system by adding social characteristics. This is because, "those social characteristics such as desire, persistence, and willingness can positively affect students' talent to be further developed" (Teacher E – 04/06/19). The teachers perceived that social characteristics amplify a student's cognitive skills, as well as academic talent overall.

Discussion

Inequity issues in the field of gifted education (i.e., underrepresentation of students of color, students from low-income family backgrounds, and gifted students with disabilities) have been frequently addressed by researchers (Gentry et al., 2019; Lakin, 2016; Moon & Brighton, 2008; Peters et al., 2019; Plucker et al., 2017; Plucker & Peters, 2018). This is not only a concern in the U.S., but also in other countries, such as South Korea. Although similar studies conducted in the U.S. provide plenty of evidence regarding the inequity in identifying and serving underrepresented students with gifts and talents (Gentry et al., 2019; Neumeister et al., 2007; Yoon & Gentry, 2009), this trend has not yet been actively studied in South Korea (Lee & Lee, 2016). It is proper and timely to investigate an identification method that can equitably identify and serve students from diverse backgrounds. This also addresses Lee and Lee's (2015) call for a validation study of a teacher rating scale for use with the underserved population in South Korea.

From the quantitative findings, the revised CFA model demonstrated a strong fit, similar to the findings reported by Peters and Gentry (2010). In their study, they indicated room for improvement in the RMSEA value, which was .11 in the revised CFA model; whereas, our RMSEA value (.0845) was better than theirs, as it was within an acceptable range (Browne & Cudeck, 1993; Chen et al., 2008). Yet, as Fan and Silvo (2007) and Kenny and McCoach (2003) revealed, part of the large RMSEA value may be due to the relatively small number of items in the scale. Given that the *HOPE Scale* consists of only 11 items in total, a larger than optimal RMSEA index was not surprising, and the model fit statistics remained strong. In addition, although the alpha reliability estimates for the two factors of the *HOPE Scale – Korean* (.916-academic and .838-social) were lower than the alpha reliability estimates of the U.S data (Peters & Gentry, 2010; .97 and .95, respectively), the South Korean estimates were still high.

Findings from the MCFA revealed that teachers can effectively nominate low-income and multicultural students for gifted programs without bias and that the *HOPE Scale – Korean* is invariant among the different income and cultural groups in our sample. Academic and social items on the *HOPE Scale – Korean* were not biased against low-income and/or multicultural students as rated by their teachers, meaning that the scale provided similar information concerning students in different demographic groups. This finding was similar to those in Peters and Gentry's (2010) study, in which they reported the *HOPE Scale* validation study with U.S. populations did not demonstrate significant differences between students from low-income families and students from regular-income families on tests of equal factor loadings, equal indicator intercepts, or equal factor variances. The studies conducted in the U.S. and South Korea showed that the *HOPE Scale* and the *HOPE Scale – Korean* items measured the same constructs across both groups.

No matter which results we retrieved from the quantitative findings, we thought it was important to ask teachers how they perceive the effectiveness of the *HOPE Scale* as they are the ones who would directly use this measure. Their thoughts would not only provide perceived usefulness of the scale but also impart critical information what is missing in the current measure. This would help future researchers designing or revising teacher-rating scale. Although the quantitative study revealed promising results in our research, the qualitative data from the teachers' interview provided additional information we missed from the earlier outcomes. The findings from the follow-up interview revealed that the teachers were aware of the importance of the teacher-rating scale, as well as the two components of the scale (academic and social). However, they reported a lack of confidence using social items to identify students for gifted programs. This outcome was consistent with the previous research findings (i.e., Lee & Lee, 2015; Lee & Lee, 2016), showing that the teachers tend to focus heavily on traditional academic components, rather

than social and emotional characteristics of gifted students. Particularly, the qualitative results were similar with Lee and Lee's (2016) study, showing that teachers considered leadership, sociality, and self-esteem as ineffective factors, when identifying gifted students compared to creativity, cognitive ability, and task-commitment.

This might be explained by the paucity of teacher training (Abu et al.,2017; Hammerschmidt, 2016; Moon & Brighton, 2008; Szymanski & Shaff, 2013) related to gifted education, resulting in a lack of understanding about the characteristics of gifted students and the ways to support them. However, it is still promising that the quantitative results showed no items functioning differently among different demographic groups, indicating that the teachers were not biased against any underrepresented group when completing the *HOPE Scale* despite this lack of teacher training. The results provided hope when compared with the findings from Elhoweris et al.'s (2005) and Moon and Brighton's (2008) studies, which showed teachers tended to not recognize, nor refer to underserved populations in general. However, the findings might also imply that teachers' views may not be accurate given our quantitative results; although the teachers questioned the social scale, the data fit the model well.

In addition, it is encouraging that the teachers had a positive perception about adopting multiple criteria, including an observation tool (Brown et al., 2005; Esquierdo & Arreguin-Anderson, 2012; McBee, 2006), such as the *HOPE Scale*, and admitted its potential for identifying underrepresented students, who traditional identification measures often miss. Therefore, the consistent development and the validation of a psychometrically sound and simple instrument is needed to help with the identification of underrepresented students for gifted programs.

Limitations and Implications for Future Research

The quantitative results from 55 teachers and the qualitative results from six teachers cannot fully represent the Korean public schoolteachers' perspectives on identifying and supporting gifted students from low-income or multicultural backgrounds. Teachers may show different results when nominating and referring their students based on their exposure to gifted education training, as well as their general experiences (Hansen & Feldhusen, 1994). Our goal in this study was not aimed to differentiate the pattern or results of the teachers based on their experience with gifted education or students with gifts and talents. However, if a future study is designed considering the teacher factor, it would be meaningful to investigate. For instance, as regular schoolteachers tend to under-identify students from low-income or multicultural families (Harradine et al., 2014) as compared to a teacher with training in and knowledge of gifted education, it would be interesting to explore whether any differences exist between two groups of teachers. Due to the nested structure of our data, not only the teachers' effect, but also the general classroom characteristics (e.g., percentage of students from low-income or multicultural family background) might differentially affect students' results. The Intraclass Correlation (ICC) value of our data was .1206. This revealed that raters' effect accounted for 12.06% of the total model variance, indicating the multilevel modeling analysis might explain additional information.

In this study, we used single-level CFA as "Muthen (1997) noted, that typically, ICC values tend to range from .00 to .50 and suggested that when group sizes exceed 15 and findings yield ICC values of .10 or larger, the multilevel structure of the data should definitely be modeled" (Byrne, 2012, p. 354). In the Korean dataset, the ICCs of items range from .049 to .173, meaning that some items showed a strong nested effect, whereas other did not. Ignoring the nested effect is a limitation of the study, however, using a multilevel model may result in reduced statistical power. Future researchers may apply a two-level CFA to estimate the raters' effect and control for

examining invariance (Ryu, 2014). This approach differs from MLM analysis, which does not fully capture the raters' effects on individual items and how they interact with the invariance.

In addition, although our quantitative findings revealed no differences between the student groups (regular income vs. low income and Korean vs. multicultural) in terms of the invariance test results, teachers' perceptions of gifted students' social and emotional characteristics were comparatively uncertain and questionable. Based on the qualitative results concerning how teachers generally perceived the social and emotional characteristics of gifted students, future research is needed that focuses on developing a scale specifically designed for the social and emotional domain of giftedness. Further, such study could focus on how the gifted program effectively incorporates the affective curriculum for students with gifts and talents who might benefit from developing their collaborative or leadership skills. Future studies may include relevant populations for more in-depth focus on underserved students who show different social and emotional characteristics compared to their peers (Olszewski-Kubilius & Clarenbach, 2012).

CHAPTER 3 (STUDY-2)

Exploring Individual and Classroom Characteristics on Students' Outcome Scores from the *HOPE* Teacher Rating Scale

The identification process cannot be separated from gifted education because this process determines which students will be served in gifted programs. One of the current identification issues in the field is that intelligence tests are the primary assessment measures used to identify students for gifted programming (Sternberg, 2015). Unfortunately, these intelligence tests do not fully assess the multiple dimensions of giftedness (Jarosewich et al., 2002; Sternberg, 2015), and the field has few technically sound measures (Pfeiffer & Jarosewich, 2007). Kovacs and Conway (2019) criticized current uses of intelligence tests, stating, "while most cognitive psychologists and neuroscientists today agree that there is no such thing as general intelligence, psychometricians have become remarkably good at measuring it" (p. 268). After studying human intelligence for 45 years to understand giftedness, Sternberg (2018) similarly argued that intelligence involves more than IQ, addressing the need to expand the notions of intelligence. He perceived intelligence as more than individuals' analytical skills measured by IQ tests, and suggested the way intelligence should be measured also differs by cultures (Sternberg, 2018). As such, it is debatable whether current intelligence measurements can truly measure one's ability and potential which may be represented as giftedness.

Pfeiffer, who developed the teacher rating scale, Gifted Rating Scales (GRS; Pfeiffer & Jarosewich, 2003), highlighted that "a single test score should never be used alone in making any diagnostic or classificatory decision" (Pfeiffer, 2002, p. 43). In this respect, applying multiple pathways such as creativity tests and teacher nominations in addition to or in place of current intelligence tests or achievement tests has become more common, as the use of multiple pathways

may ensure students with high potential are not mistakenly excluded by the use of a single traditional measures (National Association for Gifted Children [NAGC], 2015). Using multiple criteria may provide information beyond what can be measured by intelligence tests (Pfeiffer, 2002, 2015; Sternberg, 2015). Well-designed teacher rating scales can be used to identify a broader range of students because they capture perceptions of students' performance based on teachers' daily observations across many classroom tasks. With a teacher rating scale that yields reliable and valid data, teachers can efficiently summarize their students' observed characteristics rather than simply referring their students based on their perceptions and opinions (Gentry et al., 2015; Jarosewich et al., 2002). Hence, it is important to investigate how a teacher rating scale functions and whether any patterns exist when it is used by teachers.

Literature Review

Teachers' Role as a Gatekeeper

A recent survey conducted by the National Research Center on the Gifted and Talented revealed that 86% of districts in their study with elementary gifted education programs used teacher nomination in their identification process (Callahan et al., 2013). This indicates that teacher nominations are a common entry point for identification (McBee et al., 2016). The use of a teacher nomination includes teachers (a) completing rating scales of their students based on their gifted behaviors observed in the classroom or (b) simply nominating or referring their students (McBee, 2006). The scores or nomination results often function as a gate for the next step in the identification process (McBee et al., 2016; Pfeiffer & Jarosewich, 2003). Accordingly, teacher nominations may limit the number of students pass through a "gate" and who are subsequently assessed with the more costly measures including standardized tests (McBee et al., 2016).

In his study, McBee (2006) argued although the teacher nominations generally showed high quality results, it was still premature to strongly advocate the accuracy and effectiveness of the teacher nominations due to lack of research in this area. For example, previous research revealed many students with high ability were overlooked by teachers; therefore, they were not nominated to be tested for the gifted program (Siegle et al., 2010). To be specific, McBee et al. (2016) insisted that the nomination stage may cause a false-negative rate of up to 60% when teacher nominations are required before testing. Based on these results, McBee et al. (2016) made three recommendations: (a) increase the validity of nominations, (b) use lowered cutoffs for nomination/screening test scores, and (c) consider the abolishment of the two-stage system and instead test all students with simple assessments, making the gifted program more inclusive. To accomplish this, it is not only necessary to provide teacher training on the characteristics of giftedness, but also to develop simple and sound teacher rating scales that teachers can easily use.

Teacher Rating Scales to Identify Students with Gifts and Talents

Although the use of teacher rating scales is controversial (Jarosewich et al., 2002), the efficiency of their use is still appealing (Callahan et al., 2013; Gentry et al., 2015; Kettler & Bower, 2017; Pfeiffer & Blei, 2008). Teacher rating scales are advantageous and frequently used because of their simple scoring procedures; low cost and time investment; and simplicity of administration (Acar et al., 2016; Ridgely et al., 2020). They are also popular because they can measure a wide range of gifted traits under several constructs in the scale (Acar et al., 2016; Peters & Pereira, 2017). After reviewing 22 studies from 1959 to 1983 which included empirical data, Hodge and Cudmore (1986) concluded that there was little basis for the negative assessment for the teacher-judgment measures when they assessed the psychometric properties of the measures. Based on

their findings, they argued the use of teacher judgments in the identification for gifted students should be expanded and continued.

Several teacher rating scales exist to evaluate students' giftedness (Peters & Pereira, 2017); Some examples include the Gifted and Talented Evaluation Scales (GATES; Gilliam et al., 1996), Gifted Evaluation Scale, Second Edition (GES; McCarney & Anderson, 1998), Scales for Rating the Behavioral Characteristics of Superior Students (SRBCSS; Renzulli et al., 2002; 2010), Gifted Rating Scales (GRS; Pfeiffer & Jarosewich, 2003), Scales for Identifying Gifted Students (SIGS; Ryser & McConnell, 2004), and *HOPE Scale* (Gentry et al., 2015). Table 9 includes brief information about the six teacher rating scales. Although widely used, some researchers contend current teacher rating scales run short of the thorough evaluation of psychometric properties (Benson & Kranzler, 2018; Pfeiffer, 2015). Pfeiffer (2015) stated, "When consulting with school districts across the United States, I have been impressed with the number of gifted programs that continue to use homegrown, non-standardized teacher rating scales with absolutely no evidence of reliability or validity" (p. 89). As such, it is important that the field of gifted education not only highlights developing a scale with sound psychometric properties, but pays attention to inform about the variety of teacher rating scales to the administrators.

Table 9. Summary of Six Teacher Rating Scales

Name	Age (Administration Time)	Subscales (Constructs)	Number of Items	Rating	Internal Consistency Reliability	Limitation and Additional Note
Gifted and Talented Evaluation Scales (GATES, 1996)	Ages 5-18 (5-10 min)	Five (Intellectual ability, Academic skills, Creativity, Leadership, and Artistic Talent)	50 (Each scale is composed of 10 items)	9-point scale	All coefficients were above .90.	Confirmatory factor analyses (CFA) were conducted, however, no analytic data were provided to confirm findings.
Gifted Evaluation Scale, Second Edition (GES-2, 2000)	Ages 5-18 (15-20 min)	Six (Intellectual, Creativity, Specific academic aptitude, Leadership ability, Performing and visual arts, Motivation [optional])	48	6-point scale (0 to 5)	Coefficients ranged from .86 to .93.	Factor analytic data were supportive. Standardization was a weakness; not nationally representative of school-age children, high proportion of students from urban areas from Southern U.S., and low proportion of ethnic minority students.
Scales for Rating the Behavioral Characteristics of Superior Students - Third Edition (SRBCSS-III, 2013)	K to G12 (Not reported)	14 (Learning, Creativity, Motivation, Leadership, Artistic, Musical, Dramatics, Communication (Precision), Communication (Expressiveness), Planning, Mathematics, Reading, Technology, and Science)*	126	6-point scale (never to always)	The range of internal consistency was .84 to .98.	National norms were not provided because the test authors highlighted the importance of calculating local norms. Different language options were available (e.g., Spanish, Chinese, Korean, and Romanian). Several CFA were conducted.

Table 9 continued

Name	Age (Administration Time)	Subscales (Constructs)	Number of Items	Rating	Internal Consistency Reliability	Limitation and Additional Note
Gifted Rating Scales (GRS, 2003)	GRS-P (Ages 4-6), GRS-S (Ages 6-13) (10 min for GRS-P and 15 min for GRS-S)	Five for GRS-P (Intellectual, Academic, Creativity, Artistic talent, Motivation) and Six for GRS-S adding Leadership. However, Motivation score is not considered as an index of giftedness, instead a measure of students' persistence and desire to succeed.	60 for GRS-P and 72 for GRS-S (Each scale contains 12 items)	9-point scale	The coefficients were equal to or greater than .97 for each scale of GRS-P and GRS-S.	It is recommended that the raters have at least 1 month of continuous contact with the student prior to the completion of GRS. The results of the factor analysis were not reported in the manual.
Scales for Identifying Gifted Students (SIGS, 2004)	Ages 5-18 (10-15 min)	Seven (General intellectual ability, Language arts, Mathematics, Science, Social Studies, Creativity, and Leadership)	84 (Each scale consists of 12 statements)	5-point scale (0 to 4)	All scale coefficients were above .85 (average greater than .90).	Two forms exist; a home and a school scale for parent, teacher, or other adult who knows the student well. Due to the limited description provided, the test users need to proceed using the measure with a caution assessing Latinx students. Spanish version of Home Rating Scale is available. CFA to support the existence of seven separate scales is recommended.
HOPE Teacher Rating Scale (2015)	K to G12 (Not reported)	Two (Academic and Social)	11(Six for Academic and five for Social)	6-point scale (never to always)	Coefficients with .96 for Academic and .92 for Social subscales.	Measurement invariance analyses were conducted based on students' socioeconomic status and ethnicity (Gentry et al., 2015) and ELL status (Pereira, 2021).

Note. * For administration, each gifted education programs are selecting only the scales that relate to specific program goals and objectives. The test authors reported that the first four scales are the most commonly used objectives. Therefore, the administration time could range from 5-40 min depending upon the number of subscales used.

Teacher Characteristics Affecting the Nomination of Students for a Gifted Program

Researchers suggest teacher rating scales can help reduce the underrepresentation of students from low-income family backgrounds and students of color with different cultural and linguistic backgrounds (Peters & Gentry, 2010, 2013; Gentry et al., 2015). However, some literature provided counterexamples showing teacher biases may keep underrepresented students out of gifted programs (Ambrose, 2002; Bianco et al., 2011; Elhoweris et al., 2005; Speirs Neumeister et al., 2007). For example, teachers were more likely to nominate boys compared to girls (Bianco et al., 2011), students from low-income families were less likely to be identified as gifted (Ambrose, 2002; Elhoweris et al., 2005; Speirs Neumeister et al., 2007), and students of color were less likely to be nominated by their teachers than White students for gifted and talented testing (Morgan, 2019). Moreover, Grissom and Redding (2016) found Black students were less likely to be referred for identification by non-Black teachers than by Black teachers. This indicates the teachers' race could be one of the factors explaining the underrepresentation of students of color in gifted programs. This result is concerning, as the majority of teachers are White (e.g., about 79% of public school teachers were White in the U.S. in 2017-2018 (National Center for Education Statistics [NCES], 2020)). Interestingly they found that Black teachers nominated proportional numbers of Black and White children, reinforcing the importance of diversifying the teaching work force.

Since teachers tend to receive little or no training about gifted education, they may lack knowledge about the characteristics gifted students display (Jarosewich et al., 2002). Teachers might feel uncomfortable nominating their students for gifted programs, and nominations may rely on teachers' subjective perceptions of giftedness (Baudson & Preckel, 2016; Mönks et al., 2000). For instance, teachers may perceive giftedness as either a resource or a detriment based on their

own understandings (Zeidner & Shani-Zinovich, 2011). In terms of the years of educational experiences as a teacher, McCoach and Siegle (2007) noted the number of years of teaching had little effect on teachers' perceptions of giftedness; however, other studies revealed experienced teachers more accurately identified students with gifts and talents (Endepohls-Ulpe, 2005). Moreover, the teachers' years of experience with gifted and talented students seemed to positively affect their nomination decisions (Bianco & Leech, 2010). Likewise, no consistent findings exist with teacher-level factors (e.g., teacher's experience with gifted students, their age, ethnicity, gender, and subjective perceptions of giftedness).

Macro Level Factors Which May Affect Students' Performance and Nominations for a Gifted Program

Research on teachers' nominations for gifted programs has mainly focused on individual factors, such as the characteristics of the students and the teachers (Rothenbusch et al., 2016). Yet, class, school, and neighborhood information may additionally deliver richer contexts. This information may have theoretical and practical importance related to the nomination process for gifted and talented programs (Lohman & Gambrell, 2012; McBee, 2010). For instance, after exploring the effect of neighborhood socioeconomic characteristics on highly competent children, Maggi et al. (2004) found that although neighborhood socioeconomic factors weakly corelated with the proportion of high-achievers in the kindergarten, it had strong positive correlation with the proportion of high achievers in Grades 4 and 7 within the school. Everson and Millsap (2004) similarly reported that school SES had strong positive correlation with the students' math and verbal SAT scores. With regard to ethnicity, the race or ethnic composition of schools may affect achievement. The possible reasons would be that (a) the relationship between race or ethnic composition and poverty and (b) the peer effects (Benson & Borman, 2010). For instance, Benson

and Borman (2010) found that the proportion of ethnic minority groups was strongly linked to first graders' reading achievement, particularly for Black students. After controlling for school quality, Hanushek et al. (2009) revealed that the proportion of Black students in a school was negatively related to achievement scores for Black students but it impacted little on White and Latinx students. Hoxby (2000) similarly found classroom proportions of Black had negative relationship on Black students' reading scores, although no effect for White and Latinx students. Correspondingly, classroom proportion of Latinx students was negatively associated with reading scores for Latinx students and had a weaker negative relationship for White students. As such, several previous studies consistently indicated the ethnic or socioeconomic composition of a school may have potential effects on students' educational outcomes which may affect gifted identification and teachers' nomination.

When considering class-level or school-level factors, the demographic compositions of the students in the unit (e.g., class, school, district) may influence the decisions of teachers when they nominate students for gifted identification testing (Baudson et al., 2014; McBee, 2010). When McBee (2010) explored the nomination probabilities with large-scale data sets of elementary school students (n = 326,352) in Georgia, he found the probability of being identified for the gifted program strongly depended not only on student factors (e.g., students' ethnicity, family socioeconomic status), but also the results varied greatly across schools with different compositions of students (e.g., the average achievement level of the students in the school, the different percentages of certain ethnicities and socioeconomic status of students as a group). Regarding the macro level factors in his study, the percentage of the student body of Black and Latinx students in the school had negative correlation with the school academic environment variable, which represents the percentage of advanced students on the Criterion Referenced

Competency Test (CRCT). He also noted it was interesting to see teacher's education level or experience did not exert any significant effects on the school's academic environment. The results from McBee's (2010) study highlighted that nomination rates for gifted programs varied greatly among schools, indicating a need to study the class-level or school-level characteristics and how they affect the probability of being identified as gifted.

Purpose of the study

A comprehensive examination has been conducted regarding income and racial disparities on students' participation rate in the gifted programs (Bernie & Beilke, 2008; Gentry et al., 2019; Gibbons et al., 2012; Grissom et al., 2019; Hamilton et al., 2018; McBee, 2010; Miller & Gentry, 2010; Naglieri & Ford, 2005; Peters & Gentry, 2010; Yoon & Gentry, 2009). The authors primarily focused on individual characteristics of students showing the underrepresentation of Black, Latinx, and Native American students has been frequently noted in the literature (Bernie & Beilke, 2008; Gentry et al., 2019; Naglieri & Ford, 2005; Yoon & Gentry, 2009). Similarly, students from low-income families or neighborhoods are another group that is widely underserved in gifted education program (Gibbons et al., 2012; Grissom et al., 2019; Hamilton et al., 2018; Miller & Gentry, 2010; Peters & Gentry, 2010). However, almost all literature regarding underrepresentation issues have focused on these two micro factors, ethnicity and income (McBee et al., 2010); whereas, not many researchers have focused on the macro level and the findings were inconsistent. More exploration is needed to determine whether other environmental factors influence the probability of students being identified as gifted.

This study aimed to study teacher's rating scale (*HOPE Scale*; Gentry et al., 2015) by accounting both environmental and student individual characteristics: (a) students' ethnicity and family-income status and (b) teachers' seniority (years of teaching), the percentage of students

from low-income families, and the percentage of students from multicultural families in each classroom to investigate the possible differences in teacher-rating patterns. The primary research questions addressed in this study are:

RQ 1: To what extent do individual students' *HOPE Scale* scores differ by teacher?

RQ 2: To what extent do individual students' *HOPE Scale* scores differ by individual characteristics such as ethnicity or family-income status within the group?

RQ 3: To what extent can the variation in the relationship be explained by classrooms' contextual variables (i.e., teachers' years of experience, percentage of multicultural students in the classroom, percentage of low-income students in the classroom)?

Method

Research Background and Samples

We collected teacher rating scale data with the translated version of *HOPE Scale* (Lee & Gentry, 2019; Lee et al., under review) from Seoul, South Korea. Although South Korea is an ethnically homogenous society, with approximately 99% of individuals identifying as Korean (Asian), multicultural families are rapidly increasing due to international marriages (Sung, 2017). Given this trend, multicultural education is gaining attention and researchers in the field of gifted education in South Korea called for an action to provide sufficient support to underserved populations including students from low-income families and from multicultural backgrounds (Lee & Eum, 2018).

The data were collected from nine elementary schools, from 55 teachers who rated a total of 1,157 students. However, two teachers did not report the required information to be used in the analysis, therefore, two classes with 44 students were excluded due to the missing data. The final

dataset included a total of 1,113 students from 53 teachers in nine elementary schools. Table 10 describes the demographic information of the students.

Table 10. Students' Demographic Information

Grade Level	*N*	%
Grade 3	266	23.9
Grade 4	276	24.8
Grade 5	277	24.9
Grade 6	294	26.4
Family Income		
Mid or High income	909	81.7
Low-income	204	18.3
Ethnicity		
Korean	949	85.3
Multicultural	164	14.7
Teacher's Years of Experience (Age)		
20s (*N* of teachers=12)	258	23.2
30s (*N* of teachers=25)	532	47.8
40s (*N* of teachers=13)	282	25.3
50s (*N* of teachers=2)	41	3.7

***HOPE Scale*- Korean**

I used the *HOPE Scale* translated in Korean (Lee & Gentry, 2019; Lee et al., under review) in this study. Two subscale scores (Academic and Social) measured by the *HOPE Scale* served as outcome variables in this study. Appendix A includes the original *HOPE Scale* (Gentry et al., 2015) and Appendix B contains the *HOPE Scale – Korean* version (see Lee & Gentry (2019) and Lee et al. (under review) for the translation process). The original *HOPE Scale* (Gentry et al., 2015) was developed to equitably identify underrepresented gifted students from kindergarten to 12th grade

(Gentry et al., 2015; Peters & Gentry, 2010). It consists of six academic items (1, 6, 7, 9, 10, and 11) and five social items (2, 3, 4, 5, and 8). Teachers are asked to rate each student on the eleven items using a six-point frequency response scale (1 = never and 6 = always).

A key idea of the *HOPE Scale* is when teachers rate individual student, they need to consider and compare the student's characteristics with other students from similar environments. For example, when rating students from low-income families, teachers were encouraged to evaluate the student based on their comparison to other students who were also from low-income families according to the *HOPE Scale* instruction. Some experts in gifted education insisted that this allows for better identification of students from diverse backgrounds (Peters & Gentry, 2010; 2013; Gentry et al., 2015; Lohman, 2005; Lohman & Gambrell, 2012). The *HOPE Scale* (Gentry et al., 2015) explored in this study follows this concept, as teachers were asked to rate all of their students in the classroom while considering their backgrounds and comparing students to similar others.

In terms of the psychometric properties, the original *HOPE Scale*, with a U.S sample of 5,995 students from 349 teachers, showed model fit statistics of CFI=.96, GFI=.91, RMSEA=.10 for their final model with an internal consistency estimates of .96 and .92 for the Academic and Social subscales, respectively (Gentry et al., 2015; Peters & Gentry, 2010). The goodness-of-fit statistics of the Korean version, with a sample of 1,157 students from 55 teachers were CFI=.97, GFI=.94, and RMSEA=.08, which showed an excellent model fit to support the two-factor model with the instrument translated in Korean. The internal consistency reliability estimates for the academic and social scales with these data were .92 and .84, respectively.

Data Analyses

Multilevel Modeling (MLM)

Whether by nature or study design, social science research increasingly involves multilevel data (Nezlek, 2008; Wagner et al., 2016). Given the classroom or teacher characteristics, the students' scores received by their teachers within the same class may have been influenced by students' individual characteristics, as well as common environmental factors they shared. The student rating data collected in this study through the *HOPE Scale* are also structured hierarchically, as students are nested in classrooms with same teacher. In addition, since the *HOPE Scale* is used by teachers to rate each student in their classroom, results might differ systematically by teachers (Wagner et al., 2016). For instance, one teacher may give most of their students high ratings; whereas, another might give high ratings to only a few students. The pattern may vary depending on individual characteristics of students and the characteristics of teachers and classrooms.

Thus, to address research questions, we used the two-level cross-sectional multilevel modeling (MLM) using students at level-1 and teachers as level-2 units. The use of MLM was appropriate because *HOPE Scale* scores of the students in the same class are more likely to be similar compared to other students in different classes because they are grouped under the same teacher (Rothenbusch et al., 2016). A nested structure (e.g., students are nested in classes, classes are nested in schools, schools are nested in the district) is an important characteristic when investigating the effectiveness of teacher rating scales for the purpose of this study as well as handling dependency among student data within class. For instance, the original study of the *HOPE Scale* (Peters & Gentry, 2010) found that 15% of the Academic scores and 13% of the Social scores from the measurement were attributed to the group characteristics rather than

individual characteristics by calculating Intraclass correlation (ICC). Thus, the data analysis started with fitting an unconditional model to calculate the ICC for evaluating the nested effect by teachers.

The ICC represents a ratio of the variation accounted by group (τ^2) to the total variation (i.e., the sum of between-groups (τ^2) and within-group variation (σ^2)) in the outcome data. Its value ranges from 0 to 1.00 (Gelman & Hill, 2007; Musca et al., 2011). An ICC value of 0 in the extreme case indicates that all variability lies within groups, while a value of 1.00 means all variability lies between the groups (Musca et al., 2011). Psychometricians reported there are no fixed cut-offs or widely accepted guidelines to interpret ICC values (Raykov, 2011; Scherbaum & Ferreter, 2009). A value between .10 and .15 provides an estimate of the ICC, showing at least 10-15% of the variance is explained at a given level, so researchers can conventionally run a MLM analysis although it is not a sole determinant to decide the needs of MLM analysis (Scherbaum & Ferreter, 2009). Based on the unconditional model not including any level-1or level-2 factors, the ICC value for the Academic and Social scores from the *HOPE Scale* were .1179 and .1098, respectively. These values indicated that 11.79% of the variance of the Academic score and 10.98% of the Social score variance are due to some systematic differences across classes. Thus, the data support the analysis with proceeding conditional MLM models to explore both student and teacher effects on *HOPE* scores.

Variables and the Specified Conditional MLM Models

As previously mentioned, two outcome variables in this study were the Academic and Social scores of the *HOPE Scale*, which students received from their teachers. Although the correlation between two constructs were .83 with our data indicating strong positive relationship, we did not use multivariate method combining two scores but rather used two univariate methods following the manual of the *HOPE Scale* (Gentry et al., 2015). After conducting multiple factor

analyses, Gentry et al. (2015) concluded two constructs existed on the *HOPE Scale* supporting the concept of giftedness and recommended not combining the scores.

Level-1 predictor variables were students' ethnicity (two categories) and family income status (two categories), and level-2 predictor variables were the teachers' years of experience (four categories), percentage of low-income students in the classroom, and percentage of multicultural students in the classroom. See Table 11 for the descriptions of the variables and how they were coded.

Table 11. Variable and Coding Description

Variable	Description	Coding	Descriptive Statistics
Outcome Variables			
Academic	Average Academic score from the *HOPE Scale (*6 items*)*	Numerical (Range 1-6)	*M*=3.51 (*SD*=1.17)
Social	Average Social score from the *HOPE Scale (*5 items*)*	Numerical (Range 1-6)	*M*=3.46 (*SD*=1.05)
Individual-level (Level-1) Variables			
Ethnicity	Student's ethnicity	Dummy coded (0=Korean, 1=Multicultural)	*M*=.15 (*SD*=.35)
Income	Student's family income status	Dummy coded (0=Mid or high-income Family, 1=Low-income Family)	*M*=.18 (*SD*=.39)
Classroom-level (Level-2) Variables			
TeacherExp	Teacher's years of experience	Ordinal (0, 1, 2, 3 for teachers in their age of 20s, 30s, 40s, and 50s)	*M*=1.11 (*SD*=.80)
%Low-income	Percentage of low-income students in the classroom	Numerical (% of Range 4.54-38.10)	*M*=18.21 (*SD*=8.58)
%Multicultural	Percentage of multicultural students in the classroom	Numerical (% of Range .00-78.95)	*M*=15.51 (*SD*=18.49)

We first tested the variation of the level-1 intercept and slope effects across classes with the random-intercept-and-slopes model with no level-2 predictor but only random effects, represented by u_{0j}, u_{1j}, and u_{2j} in the models of β_{0j}, β_{1j}, and β_{2j}, respectively. These u_j terms represent the unique effects of the *j*th classroom. Although the average *HOPE Scale* scores vary across classroom for both Academic and Social subscale scores, the variances of u_{1j} and u_{2j} were not significant, meaning the effects of income and ethnicity on two *HOPE* subscores were consistent across classrooms. Thus, we only allowed the intercept varying across groups but not for the effect of level-1 predictors in the final model. No centering method was used in the level-1 factors as they were all coded as binary variables and no centering was applied for the level-2 predictors. Our final model including level-1 and level-2 factors for the Academic score was:

Level-1 Model

$$Academic_{ij} = \beta_{0j} + \beta_{1j}*(Income_{ij}) + \beta_{2j}*(Ethnicity_{ij}) + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(TeacherExp_j) + \gamma_{02}*(\%Low\text{-}income_j) + \gamma_{03}*(\%Multicultural_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

The final model is the same with the Social outcome as the variance of u_{1j} and u_{2j} in the models of β_{0j}, β_{1j}, and β_{2j} were also not significant. All analyses were conducted with HLM 7.03 (Raudenbush et al., 2017) with Restricted Maximum Likelihood (RML) as an estimated method.

Results

Academic Score as An Outcome Variable

From the unconditional model (Intercept-only model; $\text{Academic}_{ij} = \gamma_{00} + u_{0j} + e_{ij}$), the estimation of the fixed-effects showed γ_{00}, the overall mean of the Academic score in this model, was 3.50 (*SE*=.06, t_{52}=54.39, p<.001). The random variance estimates showed u_{0j}, the class effect (variance of the mean for each class around the overall mean Academic score), was .17 (*SD*=.41, $\chi^2(52)$=202.08, p<.001), revealing that there was more variation within the classes (estimates=1.21, *SD*=.41) than among the different classes. However, as the ICC value indicated, 12.03% of the variance of the model was accounted for by class placement.

The results showed that the average Academic score of students in the school with average amount of mid or high-income and Korean students in our sample was 3.68 (*SE*=.21, t_{49}=17.58, p<.001). In terms of classroom factors, none of the variables had statistically significant effects on Academic score from the *HOPE Scale*. For instance, when the teachers' years of experience increases, they rated their students lower by .09 points (after controlling for students' demographic information), but this was not statistically significant (p =.306). Similarly, classroom characteristics of the percentage of students from low-income families and/or multicultural backgrounds did not explain the variation in the outcome variable. However, students' individual characteristics (i.e., ethnicity, family-income status) had statistically significant effects on the Academic score students received from their teachers. After controlling for other variables, students from low-income families tended to receive average .63 points (*SE*=.08, t_{1058}=-7.46, p<.001) lower than their peers from mid or high-income families. Students from multicultural families also received average .23 points lower (*SE*=.11, t_{1058}=-2.14, p=.03) than Korean students when other variables were controlled. Table 12 includes the estimates.

Table 12. Multilevel Regression Estimates of the Model of Academic Score

Fixed Effect	Coefficient	*SE*	*t*	*df*	*p*
For Average Classroom (Intercept1, β_0)					
Intercept2 (γ_{00})	3.677	.209	17.577	49	<.001
TeacherExp (γ_{01})	-.091	.088	-1.034	49	.306
% of Low-income (γ_{02})	.001	.008	.131	49	.896
% of Multicultural (γ_{03})	.003	.004	.815	49	.419
For Individual Family-income Status (Slope, β_1)					
Intercept2 (γ_{10})	-.634	.085	-7.463	1058	<.001
For Individual Ethnicity (Slope, β_2)					
Intercept2 (γ_{20})	-.225	.105	-2.144	1058	.032
Random Effect	*SD*	Variance	*df*	$\chi 2$	*p*
Level-2 Intercept1 (u_0)	.419	.176	49	207.626	<.001
Level-1 Individual Residuals (r)	1.069	1.142			

Note. Deviance = 3375.162 (*df*=8)

To address the practical significance representing an effect size, we computed a reduction of the variance statistics. By adding level-1 predictors, 5.38% of the variation in the Academic score (σ) was explained by the student's ethnicity and family income status. As the level-2 predictors were not significant contributors explaining the model, we did not calculate the reduction of τ. However, the Chi-square test tells us the significant variation in the Academic score remains as unexplained ($x^2(49)=207.626$, $p<.001$), indicating additional predictors may help to explain the remaining variation in the intercepts. Also, model fit was tested by using deviance test under Full Maximum Likelihood (FML) method. Based on the deviance estimates from the full model (estimates= 3375.162, *df*=8) and the unconditional model (estimates=3437.241, *df*=3), the difference, $x^2(5)=62.079$, was significant at $p=.01$ level. This implies the significance of the results was not based on the overpower of the data which may be caused by large sample size.

Social Score as An Outcome Variable

The unconditional model ($Social_{ij} = \gamma_{00} + u_{0j} + e_{ij}$) showed that the overall mean of the Social score in this model (γ_{00}) was 3.46 (*SE*=.06, t_{52}=60.76, p<.001). The random variance estimates (u_{0j}) indicating the class effect was .12 (*SD*=.36, $\chi^2(52)$=191.14, p<.001), revealing that there is more variation within the classes (estimates=.98, *SD*=.99) than among the different classes. However, as the ICC value indicated, 11.23% of the variance of the model was accounted for by which class each student was in.

The results from the fixed-effects estimation showed that the average Social score of the students in the school with average amount of mid or high-income and Korean students in our sample was 3.68 (*SE*=.18, t_{49}=20.67, p<.001). In terms of the classroom factors, none of the variables had a statistically significant effect on the Social outcome. For instance, when the teachers' years of experience increases, they rated their students lower by average .14 points (after controlling for students' demographic information), but it was not statistically significant (p =.068) Similarly, the classroom characteristics of the percentage of students from low-income families and/or multicultural backgrounds did not affect the outcome variable. However, students' individual demographic information significantly affected the Social score they received from their teachers. After controlling for other variables, students from low-income families tended to receive average .37 points (*SE*=.08, t_{1058}=-4.81, p<.001) lower than their peers from mid or higher-income families. Students from multicultural families also received average .30 points lower (*SE*=.10, t_{1058}=-3.10, p<.001) than Korean students when other variables were controlled. Table 13 includes detailed estimates of the fixed and random effects of the model.

Table 13. Multilevel Regression Estimates of the Model of Social Score

Fixed Effect	Coefficient	*SE*	*t*	*df*	*p*
For Average Classroom (Intercept1, β_0)					
Intercept2 (γ_{00})	3.679	.178	20.665	49	<.001
TeacherExp (γ_{01})	-.140	.075	-1.866	49	.068
% of Low-income (γ_{02})	.001	.007	.071	49	.943
% of Multicultural (γ_{03})	.002	.004	.585	49	.561
For Individual Family-income Status (Slope, β_1)					
Intercept2 (γ_{10})	-.373	.078	-4.806	1058	<.001
For Individual Ethnicity (Slope, β_2)					
Intercept2 (γ_{20})	-.297	.096	-3.099	1058	.002
Random Effect	*SD*	Variance	*df*	$\chi 2$	*p*
Level-2 Intercept1 (u_0)	.348	.121	49	180.096	<.001
Level-1 Individual Residuals (r)	.976	.952			

Note. Deviance = 3164.220 (*df*=8)

In terms of the practical significance, we computed a reduction of the variance statistics. By adding level-1 predictors, 2.96% of the variation in the Social score (σ) was explained by the student's ethnicity and family income status. Similar to the Academic score as an outcome variable, as the level-2 predictors were not significant contributors explaining the model of Social score, we did not calculate the reduction of τ. However, the Chi-square test tells us the significant variation in the Social score remains as unexplained ($x^2(49)=180.097$, $p<.001$), indicating additional predictors may help to explain the remaining variation in the intercepts. When the model fit was tested by using deviance test, the deviance estimates from the full model and unconditional model was 3164.220 (*df*=8) and 3202.501 (*df*=3), respectively under FML method. This yielded a difference of $x^2(5)=38.281$ at $p=.01$ level indicating the significance of the results was not based on the overpower of the data which may be caused by large sample size.

Discussion

The underrepresentation of students from low-income families and diverse ethnic backgrounds in gifted programming has been criticized and widely studied among equity and social justice researchers (Bernie & Beilke, 2008; Gentry et al., 2019; Gibbons et al., 2012; Grissom et al., 2019; Hamilton et al., 2018; McBee et al., 2010; Miller & Gentry, 2010; Naglieri & Ford, 2005; Peters & Gentry, 2010; Yoon & Gentry, 2009). Numerous studies focused on the students' individual factors, and it was clear that Black, Latinx, and Native American students and students from low-income families have been identified for gifted education programming less often than their counterparts. However, as students are naturally nested in classrooms under the same teacher, with each classroom's unique characteristics, class-level or school-level factors may affect students being identified for gifted programs (McBee, 2010; Nezlek, 2008; Wagner et al., 2016). It was particularly important to explore the classroom factors in our study, as our outcome variables were the students' scores received by their teacher based on the Academic and Social constructs of the *HOPE Scale*, which represent the characteristics of giftedness (Gentry et al., 2015). As the teachers rated every student in their classroom based on their daily observation, the results may be subjective (Baudson & Preckel, 2016; Mönks et al., 2000; Zeidner & Shani-Zinovich, 2011), vary from teacher to teacher, and create worse circumstances if the teachers were biased (Ambrose, 2002; Bianco et al., 2011; Elhoweris et al., 2005; Speirs Neumeister et al., 2007). On the contrary, as teachers were instructed that each student needed to be evaluated in comparison with other students with similar backgrounds, the results may show positive outcomes that the *HOPE Scale* can be used as one tool to equitably identify students from diverse backgrounds (Gentry et al., 2015; Peters & Gentry, 2010, 2013).

In terms of the level-1 individual factors, our results were consistent with previous findings with U.S. sample (Peters & Gentry, 2010) that traditionally underrepresented students received

lower scores on the *HOPE Scale* than the counterpart students with an average *Cohen's d* of .30 for family-income status and .005 to .11 for ethnic groups. However, it is important to note that Peters and Gentry (2010) also explained, "Although mean differences were found, the data indicate these differences are actual differences in the underlying constructs and not because of item bias" (pp. 308-309). The pattern was similar in the Academic and Social scores. When we explored the level-2 factors—teachers' years of experience and percentage of low-income students and students from multicultural backgrounds—we found no statistically significant influences on students' Academic and Social scores. The insignificant effects of teachers' years of experience was similar to Mcbee's (2010) findings who found years taught and teachers' degree did not influence students being identified as gifted. The percentage of low-income students and students from multicultural backgrounds did not affect the *HOPE Scale* scores. These variables were included in this study because if a class has many multicultural students, teachers may understand or be familiar with their unique characteristics. Therefore, students with these backgrounds may receive higher or lower scores from teachers based on their experiences with these students. The variable of family income was added for the same reason. However, both classroom characteristics did not create any patterns on the teachers' ratings of their students.

Although the *HOPE Scale* would be completed by individual teachers for their students in the classroom, scores will often be used and compared with the results from students rated by other classroom teachers in the school. Results can also be used in higher-level programming decisions (e.g., district-level or university-based gifted programs), indicating that diverse teachers from different schools may produce *HOPE Scale* ratings from their own perspective and compare the students in their classroom. In this case, teachers' individual characteristics or experiences may cause different patterns in the scoring. It was promising, however, that the level-2 factors including

teachers' years of experience or classroom characteristics did not affect the difference of the *HOPE Scale* results. Regardless of their years of teaching or unique educational environment, teachers seemed not to form certain patterns in rating their students. Although students from low-income families and multicultural backgrounds received lower scores from their teachers on their Academic and Social scores, the mean difference does not distinguish a genuine group difference or imply measurement or raters' bias (Camilli, 2006). Therefore, investigation beyond level-1 factors were important to be explored.

Limitations

A limitation of this study was the limited information about the teachers. We used teachers' age to represent their years of teaching experiences. However, the age may not necessarily represent the true years of teaching, and the information was not concrete enough as the teachers had few options (i.e., 20s, 30s, 40s, and 50s or more) to select their age. If they provided their actual age or years of experience in numbers, results may have differed. As indicated in the Table 14, only a few teachers who participated in this study (3.7%) reported they were more than 50 years old. If we recruited more teachers in older age groups or used stratified sampling to equally represent each age group of teachers, we might have different results. In addition, teacher factors such as if they received professional development on gifted education or ran classes for gifted students may provide additional information.

Nevertheless, this study was meaningful because it used a teacher-rating scale as an outcome variable to represent students' performance, which is different from studies that frequently use intelligence test scores or standardized achievement test results. Teacher-rating scales are an alternative pathway which provides additional information about students, and can be used to apply multiple criteria when identifying students with gifts and talents (NAGC, 2015;

Pfeiffer, 2002, 2015). Investigating the effectiveness of teacher-rating scales based on students' individual factors and classroom factors is a significant attempt to move the field of gifted education forward to equitably identifying traditionally underserved populations.

CHAPTER 4 (STUDY-3)

The Relationship between Students' Academic Achievement and the Teacher-rating Scale: Exploration to Equitably Identifying Underrepresented Gifted Students

Although experts in the field of gifted education have attempted to resolve the ever-complicated issue of defining giftedness (e.g., Dai & Chen, 2013; Ziegler & Raul, 2000), the task of identifying giftedness continues to be fraught with problems, including a lack of alignment between identification assessments and program goals, the uninformed selection of assessments that do not yield valid and reliable results, and the misuse and/or misinterpretation of assessment results used for identification (NAGC, 2008). Most researchers agree that "each indicator [measure] of giftedness used has flaws and does not fully represent giftedness, the construct of interest for identification" (Adelson, 2012, p. 47).

In response to this unfavorable predicament, researchers have promoted using multiple measures or criteria of varying weights and combinations to identify gifted students (Lakin, 2018; Lohman & Renzulli, 2007; Peters & Gentry, 2013; Pfeiffer, 2002). These measures may include achievement tests, verbal and non-verbal intelligence tests, GPA, teacher-rating scales, students' portfolios, and nominations from parents, peers, and even students themselves. The practice of using a single measure to identify giftedness promotes underidentification and has been criticized harshly; whereas, researchers have demonstrated using multiple measures increases the diversity of the population of identified students, including those from underrepresented groups (Lohman & Renzulli, 2007; Peters & Gentry, 2013). As underidentification and underrepresentation of students from low-income families and from culturally diverse background (e.g., Black, Latino, and Native American groups) have been black-eye issues in the field (Yoon & Gentry, 2009; Pereira & Gentry, 2013; Plucker & Peters, 2018), this study explored the relationship and the effect

of combining the results of two measures that reflect different criteria (academic achievement and teachers' ratings of gifted behaviors) on the identification of students from underserved populations in South Korea.

Literature Review

Underserved Populations and the Identification of Gifted Students

Much has been written about underserved populations in gifted education and how the field continues to struggle with identifying gifted students from underrepresented groups. In the United States, this issue has been at the forefront for many years, yet the National Association for Gifted Children (NAGC) and the Council of State Directors of Programs for the Gifted (CSDPG) *State of the Nation in Gifted Education* (NAGC, 2015; Rinn et al., 2020) reported that limited public accountability and the lack of data across states hamper the identification of historically underrepresented populations and the quality of services they receive. More recently, a comprehensive, landmark study of "missingness" as it relates to laws, access, and equity in gifted education in the U.S. provides valuable data and promotes public accountability for more equitable identification of giftedness (Gentry et al., 2019). By race alone, the percentages of students missing from gifted education are dismal: over half of Black (63% to 74%), Native Hawaiian/Pacific Islander (59% to 72%), and Latinx (53% to 66%) students, and only a slightly better percentage of American Indian/Alaska Native American (48% to 63%) students are missing due to nonexistent or impotent state laws, limited access, and/or inequity in identification processes (Gentry et al., 2019, p. 4-5).

One underserved population that spans all populations across all learner demographics is learners from low-income families (Gentry & Seward, 2018). Several factors work against finding

these gifted learners. First, identification systems that rely on high cutoff scores on academic assessments miss gifted learners from low-income families who often score below these sometimes arbitrarily set cutoffs (Callahan, 2005). Second, mistakenly believing that these students' above-average achievement is the best they can be expected to do, teachers who are untrained in gifted education and in identification of underrepresented populations continue to overlook these students. Sadly, continuing to overlook these learners year after year places them in jeopardy of unsatisfactory progress, underachievement, potential dropout in high school, and, if they move on to postsecondary education, college undermatching (Wyner et al., 2009). Excellence gaps increase as students with high ability from low-income families fail to keep pace with their intellectual peers from more affluent families on the National Assessment of Educational Progress (NAEP; Plucker et al., 2010; 2013). Third, teacher bias and low academic expectations for learners from low-income backgrounds influence teachers' interpretations of high achievement and performance and skews their abilities to spot academic talent in this underserved population. Too often, teachers do not consider high achieving learners from families living in poverty as qualified for gifted services; instead they view their high achievements as pleasant surprises and fail to recognize their gifted potential (Wyner et al., 2009). For these reasons, it is likely that hundreds of thousands gifted students from low-income families are unidentified (Gentry & Seward, 2018; Gentry et al., 2019).

The Need for Applying Multiple Criteria When Identifying Gifted Students

Researchers in the field of gifted education have bemoaned the inadequacies of traditional identification processes and their failure to identify learners from low-income families as well as students with diverse cultural backgrounds (Borland, 2004; Borland et al., 2000; Borland & Wright, 1994, 2000; Olszewski- Kubilius, 2007; Worrell, 2007). For instance, some controversy exists

with regard to the use of standardized achievement tests for the identification of gifted students. Many researchers have shown that scores on standardized achievement tests correlate with a students' family income. That is, students from families with higher incomes tend to score higher than students from low-income households (Stambaugh, 2007; Valencia & Suzuki, 2001; Wyner et al., 2009). Other common criticisms of achievement tests are the degree to which test bias negatively influences results for students from minoritized, low-income, or rural families (Naglieri & Ford, 2003) and ceiling effects fail to discriminate students' levels of achievement. Cultural experiences and opportunities to learn are integral to test fairness, making the measurement of their observed effects on test performance more complicated. However, the influence of test bias on identification has not been adequately examined, and the use of local and group-specific norms in the identification of gifted learners will likely compensate for most occurrences of legitimate test bias (Peters & McBee, 2019). These findings do not imply, however, that the assessments and other instruments being used in these fallible systems are useless for identifying talent. Quite possibly, a more equitable and effective use of the results of these same assessments can be determined. For example, Card and Giuliano (2015) found that universal screening can increase the number of underrepresented learners who qualify for gifted services, including those from low-income families, by 180%. Peters and Gentry (2012b) and Peters et al. (2014) argued for the use of group-specific and local norms when using test results in the identification of advanced learners.

The Application of Group-specific Norms to Identify Gifted Students

Because students are identified for gifted services at the school level, the use of national norms related to the instruments or tests used for identification may not be accurate for local identification because the national norm group may misrepresent a given school's demographics. The use of national norms to identify gifted learners might make even less sense when examining

various subpopulations to which local students belong. The uniqueness of every school with regard to the dimensions of race, ethnicity, language, socio-economic conditions, and locality requires carefully selected identification measures that are as fair as possible and whose results are reliable and valid for the students educated in that school. The use of local and group-specific norms applied to the results of identification instruments has been suggested as a more appropriate measure for underrepresented student populations, including learners from low-income families (Peters & Gentry, 2010). When practitioners apply local norms, they analyze their students' results on a given test and determine (usually by percentage or percentile) who the top performers are within local groups and identify them for gifted services. In applying group-specific norms, practitioners would analyze only those students' scores who fall into the specific group selected. For example, practitioners would analyze the results of students from low-income families and determine (by percentage or percentile) who the top performers are in that particular group of students and identify them for gifted services.

In their book, *Beyond Gifted Education: Designing and Implementing Advanced Academic Programs*, Peters et al. (2014) called for local-level practitioners to use tests differently by applying local and/or group-specific norms to identify as many learners from underrepresented groups as possible. The authors provide specific instructions with regard to creating local norms and combining multiple measures whose results are reported with different metrics (e.g., standard scores, percentile ranks, and composite scores). Peters and Gentry (2012b) showed that applying group-specific norms to grade-level achievement test scores resulted in an increase in the number of students from low-income backgrounds for gifted services.

The Role of Teacher in the Identification Process and the Teaching-rating Scale

Teachers are in the best position to spot talent. Over the course of an academic year, they come to know their students well through the academic and social-emotional experiences in their classrooms. However, "most general education teachers have no training in working with gifted students before entering the classroom or through ongoing professional development and [they are ill-prepared] to properly identify gifted students," (NAGC, 2015, p. 3) especially those from underrepresented populations (Rinn et al., 2020). The failure of all but one state legislature to prioritize teacher preparation and subsequent training in gifted education makes the goals of equity in identification and quality in services unattainable; teacher referrals for gifted identification in one U.S. state showed that learners from minoritized and low-income populations were less likely to be referred than those from White and higher-income families (McBee, 2006).

Although the use of teacher-rating scales in the identification process has been frowned upon in past years, researchers have shown that teacher ratings can yield as reliable and valid data as any other method (Callahan et al., 2013; Gentry et al., 2015; Hodge & Cudmore, 1986; Kettler & Bower, 2017; Pfeiffer & Blei, 2008). The best teacher ratings are based on evidence that is specific, observable, and verifiable. Although no human rating system is free from bias, the more teacher-rating scales focus on this evidence, the more bias-free the results will be. When teacher-rating instruments of any kind are well-constructed and provide valid and reliable data, the accuracy of teacher ratings will also improve (Hodge & Cudmore, 1986; Peters & Gentry, 2013). Further, when teachers understand the concept of giftedness, including the behaviors associated with giftedness, they are more accurate and inclusive in their ratings (Peters & Gentry, 2013; Peterson, 1999).

Purpose of the Study

Identifying and serving underrepresented populations has been a critical issue in the field of gifted education for many years (Gentry et al., 2019; Lakin, 2018; Yoon & Gentry, 2009). Similarly, researchers in South Korea have recently addressed this issue and called for an action for research exploring underserved populations (Lee & Lee, 2016; Sung, 2017). In South Korea, Gifted Education Database defines a gifted child as "a person who possesses extraordinary innate abilities or visible talents requiring special education to nurture them," and it presents an interesting case with regard to underserved populations in gifted education. Throughout its rich history, Korean school systems have educated students who are almost entirely representative of the dominant ethnic group; until recently, schools had been largely unaffected by immigration and non-native language learners. In 2019, over 99% of South Koreans identified as ethnically Korean (World Population Review, 2019). For this reason, the primary underserved population of gifted students is learners from low-income families. However, as the immigrant and/or culturally diverse families are increasing in Korea as well, and because the South Korean gifted education system has not much explored the identification of students from culturally diverse backgrounds (Lee & Lee, 2016), it is timely to consider the ethnic information when identifying and serving students for the gifted programming (Lee & Lee, 2016; Sung, 2017). As the field of gifted education continues to determine the best way to identify giftedness across race, ethnicity, family income, and locale, we believe an approach that includes multiple criteria has strong potential. In this respect, we explored whether the *HOPE Scale* (Gentry et al., 2015) can be used effectively to identify gifted students from the underrepresented groups in conjunction with achievement test scores.

Teachers' perceptions, including how they interpret the behavior of students from low-income and/or culturally diverse families, is worthy of exploration. Hence, in the first phase of this

study, we examined the relationship between students' *HOPE Scale* scores as measured by their teachers' observations of gifted behaviors (Academic and Social subscales) and the students' outcome (i.e., achievement test scores in mathematics and reading, number of students' talented subject areas marked by their teachers in the *HOPE Scale*). As researchers have also shown that the types and combinations of measures affect program diversity, especially with regard to student ethnicity and socioeconomic status (Lakin, 2018; McBee et al., 2014; Peters & Gentry, 2013), we explored the different combination rules that may lead to differences in the number and diversity of students who are identified as gifted. Several identification scenarios were tested, similar to the method used by Peters and Gentry (2013). The following research questions and hypotheses were addressed:

RQ 1: What is the relationship between teacher ratings of gifted behaviors (i.e., Academic, Social) as measured by the *HOPE Scale* and students' outcome performance results (i.e., school-based achievement tests, the number of subject areas where they show talent)?

For the first research question, we hypothesized that students' performance outcome has positive relationship with both Academic and Social scores from the *HOPE Scale* rated by the teachers.

RQ 2: How do various cutoff percentages and group-specific norms when applied to achievement test results and teachers' *HOPE Scale* ratings affect the identification of students from low-income and/or culturally diverse families for gifted programming?

For the second research question, we hypothesized that more students from low-income and/or culturally diverse family would be identified as gifted when (a) lowering cutoff percentages of

achievement test scores and (b) combining teacher-rating scale results to the achievement test outcomes.

Method

Participants

Participating elementary schools (n = 9) in Seoul, South Korea, were purposively selected using the snowball sampling method. Schools serving large numbers of low-income students (similar to Title 1 schools in the U.S.) or self-identifying as having comparatively large concentrations of culturally diverse and/or low-income students participated in this study. Based on records provided by Korean Educational Development Institute (KEDI; 2016, as cited in Sung, 2017), the percentage of culturally diverse students in Korea is 1.36%; however, this population is consistently increasing. Given that the Korean ethnic group constitutes 96-99% of the total population, many schools do not report culturally diverse students. Schools who self-identified as having comparatively large number of culturally diverse students reported that they have at least one or two students with different cultural backgrounds in their classes. Although the *HOPE Scale*, Korean version, includes different ethnic group options, 99.9% of the students who were marked as culturally diverse in this study were still of Asian descent; therefore, we used only two terms to differentiate the two groups—Korean and culturally diverse students. We did not use the term "Non-Korean" as many of them are still considered as Korean; The term "culturally diverse" in Korea typically indicates individuals from a family formed by the international marriage of a Korean citizen with a partner from a different nation (Lee & Lee, 2016). In terms of the low-income family, it includes (a) recipients of basic living (below 100% of the minimum cost of living)

and (b) near-poverty group (below 120% of the minimum cost of living), yielding families with household income of "below 50% of median income" in Korea (The Government of Korea, 2019).

Teachers (n = 55) who voluntarily participated in the study in grades 3 - 6 (n = 13, 15, 13, 14, respectively) were instructed to complete the *HOPE Scale* based on their observations, experiences, and perceptions of each of their students' (n = 1,157) academic and social behaviors. No additional training or instructions were provided other than those contained in the *HOPE Scale* manual (Gentry et al., 2015) itself because we wanted to maintain the schools' established procedures for referring students for gifted programming; it was conducted in a similar fashion with the previous research in the U.S. investigating the validity results of the *HOPE Scale* in which no additional teacher training was provided to the teachers (Gentry et al., 2015; Peters & Gentry, 2010; 2012; 2013). When the teachers were invited, however, we highlighted the sentence "compare a student with other students with similar backgrounds and/or experience" which is an important direction for using this instrument.

Teachers of grades 1 and 2 were not invited to participate in this study because the identification process in Korean schools begins when students enter the third grade. See Table 14 for students' demographic information. Although students were uniformly distributed by gender and among grade levels, over 80% of students were identified by their teachers as Korean and as paying for their own lunches at school. Overall, 14.43% (n = 167) of students were from culturally diverse backgrounds, and 18.67% (n = 216) of students were from low-income families. A total of 2.85% (n = 33) of the students were from culturally diverse, low-income families. Given the national index of culturally diverse population (1-4%) in Korea, our sample includes comparatively large number of students with different cultural backgrounds.

Table 14. Students Demographic Information

Grade Level	Frequency (%)	Gender (Male/Female)	Ethnicity (K/C)*	Income (P/FR)**
Grade 3	266 (23)	133 (50) / 133 (50)	205 (77) / 61 (23)	224 (84) / 42 (16)
Grade 4	320 (28)	159 (50) / 161 (50)	279 (87) / 41 (13)	260 (81) / 60 (19)
Grade 5	277 (24)	148 (53) / 129 (47)	240 (87) / 37 (13)	220 (79) / 57 (21)
Grade 6	294 (25)	152 (52) / 142 (48)	266 (90) / 28 (10)	237 (81) / 57 (19)
Total	1157	592 (51) / 565 (49)	990 (86) / 167 (14)	941 (81) / 216 (19)

*K = Korean. C = culturally diverse.
**P = students paying for lunches. FR = students receiving free/reduced lunches.

Instruments

The HOPE Scale

The *Having Opportunities Promotes Excellence (HOPE) Scale* (Gentry et al., 2015; Peters & Gentry, 2010, 2012, 2013) is a teacher-rating scale designed to help identify students from low-income families and/or culturally diverse backgrounds based on teachers' observations, experiences, and perceptions of students' gifted behaviors on a 6-point response scale (1 = never to 6 = always). After factor analyses including validity tests (CFI=.96, GFI=.91, RMSEA of .10), the *HOPE Scales*' 11 items describe six Academic (e.g., *Is eager to explore new concepts*) and five Social (e.g., *Shows compassion for others*) behaviors associated with giftedness, easily observed and identified by teachers in their daily interactions with students. (See Appendix A for the *HOPE Scale*, English version, and Appendix B for the *HOPE Scale*, Korean version; see Lee & Gentry (2019) and Lee et al. (under review) for the translation process). In addition, the *HOPE Scale* has been shown to provide invariant results across underrepresented populations in gifted education and, when used in conjunction with other identification measures, should be considered a reliable means to obtain more objective teacher input during the identification process (Peters & Gentry, 2010, 2013). The directions included with the *HOPE Scale* instruct teachers to rate each

student on the 11 behaviors as compared to other children similar in age, experience, and/or environment (Peters & Gentry, 2010). Teachers can also provide demographic information (e.g., student ethnicity and free/reduced lunch status, a low-income index) and the subject area(s) in which the student shows talent (i.e., math, reading, creative writing, social studies, science, foreign language, arts, other). No standardized cut-off scores are suggested; however, the authors provided a guideline to calculate local norms. See the *HOPE Scale* manual (Gentry et al., 2015) for more details.

Among many teacher-rating measures, we selected *HOPE Scale* as (a) it is one of the most recently created instrument with its clear purpose of helping to identify underrepresented gifted students, (b) it contains social construct which is unique compared to other measures traditionally focus on the academic characteristics of gifted students, and (c) it is simple with only 11-items, which may relieve a burden for the teachers. However, it is important to note that as *HOPE Scale* is a short instrument, it may lack reliability. For instance, Peters and Gentry (2010) manifested that the instrument's RMSEA value was slightly above (.10) compared to the suggested criteria (.08) which may be due to its small number of items. In addition, a replication study (Peters & Pereira, 2017) investigating the internal validity structure of three major teaching rating scales – Scales for Identifying Gifted Students (SIGS; Ryser & McConnell, 2004), Gifted Rating Scales (GRS; Pfeiffer & Jarosewich, 2003), and the *HOPE Scale* (Gentry et al., 2015) – revealed that all three instruments failed to meet traditional fit criteria with varying degrees although they all reported to fit the criteria when published. As such, the instrument itself cannot be free from yielding the absolute soundness of psychometric properties with different samples, and the measurement cannot fully address the issues of teachers' implicit bias on underserved gifted population. Nevertheless, when investigated multigroup confirmatory factor analyses (MG-CFA) on the

HOPE Scale with both U.S (Peters & Gentry, 2010) and Korean sample (Lee & Gentry, 2019), the results indicated that teachers, without bias, nominated low-income and/or students with different ethnic backgrounds.

Achievement Tests

Since 2013, Korean educational policy no longer requires national or state (province) achievement tests for elementary school students because of the excessive fervor for education that imposed unreasonable burdens for young students (Ministry of Education, 2013). Instead, teachers at each school exercise their professional judgment regarding when and how to administer achievement tests, with test results used for educational decision making at the local level only. Therefore, the mathematics and reading achievement scores used in this study were provided by the teachers and represent grade-level results from either formative assessments to evaluate students' achievement levels at the beginning of the semester or summative, content-area unit tests. Based on the achievement score results received, we standardized the raw scores to T-scores with a mean of 50, standard deviation of 10, and a range of 0 to 100. This provided a common metric, allowing us to compare students' achievement scores across grade levels and schools. We processed this standardization because we had an assumption that students were randomly spread across classes in each school; it is true that teachers refer those formative or summative assessments results to rank-order the students in each class and share the results with other teachers when dividing the students to each class at the beginning of the academic year. By doing so, the teachers evenly distribute the students into new classes with the expectation of each class to have similar average academic ability. Therefore, although the reading and mathematics scores we received were not from one common standardized test, we presumed the scores provided by each teacher fairly represented the students' academic performance.

Data Collection and Analyses

Data collection occurred in October 2018 when all teachers submitted their students' achievement scores in math and reading along with their *HOPE Scale* ratings. Data analyses were conducted in two phases: (1) application of Structural Equation Modeling (SEM) to explore associations between the teachers' *HOPE Scale* ratings and their students' reading and mathematics achievement scores, as well as the number of subject areas in which students show talent, and (2) application of various cutoff percentages and group-specific norms to determine their influences on the identification of students from culturally diverse and/or low-income families for gifted programming.

Phase I

SEM is a technique that examines the relationship among variables by analyzing covariance, enabling researchers to estimate the effects of hypothetical or theoretical constructs (Raykov & Marcoulides, 2000). Although it falls within the regression family of statistical analyses, SEM has many benefits over traditional linear regression. First, SEM allows multiple outcome variables compared to multiple regression that only allows one. Because of this, "researchers are able to distinguish between observed variables (e.g., the scores or the items) and the underlying constructs (e.g., giftedness or creativity)" (Adelson, 2012, p. 48). In this study, we are just as interested in the latent construct of giftedness as we are in the instruments used to measure it. Second, whereas multiple regression assumes that the predictors are measured without error, SEM recognizes that no measure is perfectly reliable. This characteristic was especially beneficial in this study with regard to the assumed variability in achievement test scores across participating schools. Third, SEM allows different weights to be applied to the latent construct, giving more flexibility than simply averaging and/or totaling the item scores using equal weight.

Because of these advantages, we used SEM in phase I of this study to explore the relationships among students' achievement scores in mathematics and reading and the number of students' talent domain areas as identified by their teachers (three endogenous variables) and students' Academic and Social subscale scores from the *HOPE Scale* (two exogenous variables). As hypothesized and presented in the model (see Figure 2), Academic and Social scores from the *HOPE Scale* were analyzed as predictive measures of students' performance on mathematics and reading achievement tests and on the number of subject areas where students show talent as identified by their teachers. In terms of the exogenous variables, it is important to note that the Social construct was added as it is unique value of *HOPE Scale* that the measurement developers believed to be one of the key components of giftedness (Gentry et al., 2015). With regard to endogenous variables, the number of students' talented areas marked by the teachers was added as a predictor along with the student's reading and mathematics scores, based on the assumption that gifted students tend to have multipotentiality characteristics (Kerr & Sodano, 2003). Throughout the analyses, we used LISREL 9.3 (Jöreskog, n.d.) to estimate parameters under maximum likelihood (ML) method.

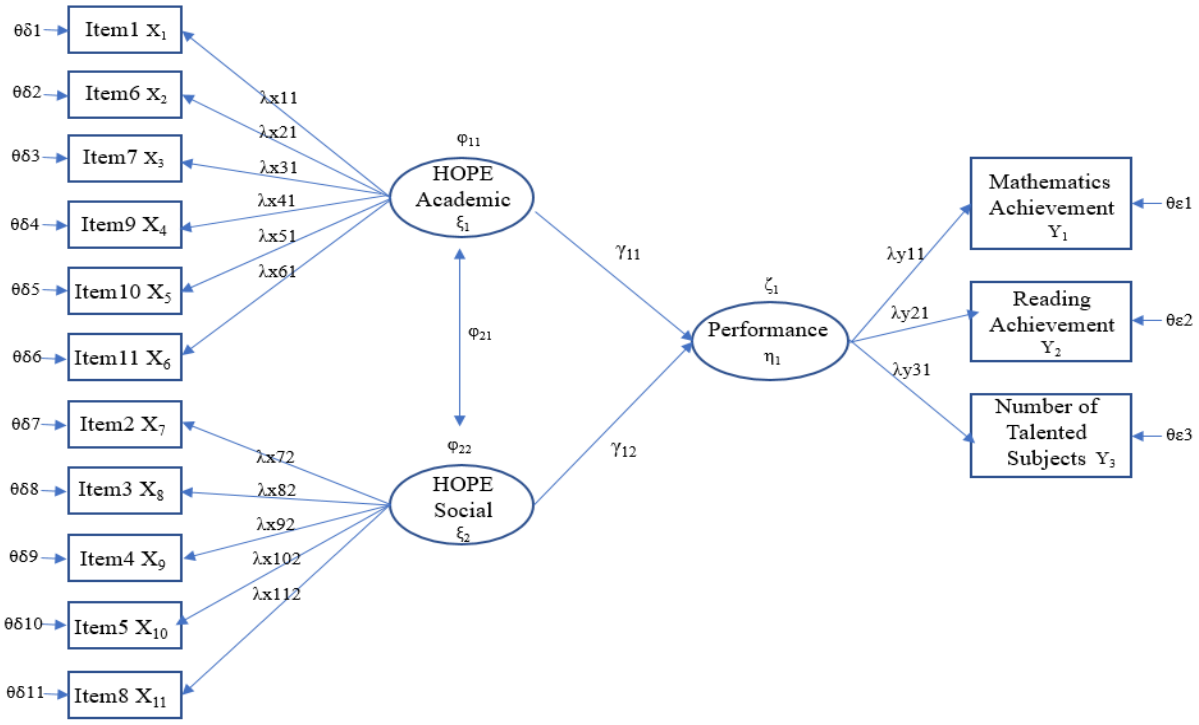

Figure 2. Specified Structural Equation Model

Phase II

Using cut-off scores or percentages with multiple measures to determine who is identified for gifted programs is common in education (Barton & Coley, 2009; Peters & Gentry, 2013). In Phase II of this study, we created eight scenarios that combined students' achievement test scores and teachers' *HOPE Scale* scores using varying percentages (i.e., 10% and 25%) to compare their influences on the populations of students who become eligible for gifted services - specifically the number of students from culturally diverse and/or low-income families. We replicated the scenarios used by Peters and Gentry (2012b) in which the top 5%, 10%, and 25% of students were identified for gifted programming based on their achievement test scores alone (mathematics and reading, separately). However, in the current study, we did not use the most restrictive 5% scenario

because it would result in fewer than two students eligible for gifted services in each of the teachers' classes given that the average class size was 21 students.

Results

Phase I – Structural Equation Modeling (SEM)

Before SEM results are described, a brief look at the analysis of the results of the *HOPE Scale* is warranted. Table 15 provides descriptive statistics for each of the 11 behaviors in the *HOPE Scale*. Item-level mean scores across grade levels ranged from 3.03 (*SD* = 1.32) to 3.89 (*SD* = 1.28). The Academic subscale items averaged 3.50 (*SD* = 1.18) and the Social subscale items averaged 3.45 (*SD* = 1.06).

Table 15. *HOPE Scale* Descriptive Statistics

Item	Response Percentages						Mean	SD	Skewness	Kurtosis
	1	2	3	4	5	6				
Academic1	5.36	11.58	25.93	25.24	20.14	11.75	3.78	1.35	-.12	-.68
Academic6	10.20	15.59	27.48	21.00	17.20	7.52	3.41	1.41	.05	-.80
Academic7	11.67	18.41	27.92	19.10	14.43	8.47	3.32	1.44	.17	-.80
Academic9	6.40	19.19	34.23	20.92	13.83	5.45	3.33	1.26	.24	-.48
Academic10	6.83	19.88	38.20	21.35	9.85	3.89	3.19	1.18	.30	-.14
Academic11	5.62	14.69	29.30	21.43	16.25	12.62	3.66	1.40	.07	-.77
Academic (all)							3.50	1.18	<.01	-.60
Social2	10.63	18.06	30.34	19.45	14.52	7.00	3.30	1.39	.17	-.70
Social3	2.94	9.51	34.23	24.89	20.66	7.78	3.74	1.20	.04	-.49
Social4	3.20	10.11	25.93	27.57	21.18	12.01	3.89	1.28	-.12	-.60
Social5	21.18	20.14	21.61	16.34	12.27	8.47	3.04	1.57	.33	-.96
Social8	9.85	15.99	33.54	21.35	13.57	5.70	3.03	1.32	.14	-.53
Social (all)							3.45	1.06	.10	-.52

Although overall Academic and Social subscales' demographic statistics are comparable, it is interesting to note teachers' ratings for particular items. Teachers rated only 3.89% of all

students across grade levels as "always" *thinking "outside the box"* (Academic 10), but they "always" observed *Performs or shows potential for performing at remarkably high levels* (Academic 1) and *Has intense interests* (Academic 11) for approximately 12% of all students, a relatively high percentage for this highest rating on the *HOPE Scale*. In the Academic subscale, teachers observed Academic 1 (above) "often, almost always, and always", 14% more than any other item, and *Is eager to explore new concepts* (Academic 6) and *Exhibits intellectual intensity* (Academic 7) received the lowest rating—"never observed"—more than any other Academic subscale item. What's more, teachers' ratings split the 6-point scale on Academic 11 (above) and reflected "sometimes, rarely, or never" observed more frequently for the remaining items—Academic 6, 7, 9 (*Uses alternative processes*), and 10. Academic 10, for example, was 30% more likely to be rated "sometimes, rarely, or never".

With regard to Social subscale items, teachers rated *Shows compassion for others* (Social4) as "often, almost always, and always" more frequently than any other Social behavior, with teachers "always" observing Social4 for just over 12% of all students, a relatively high percentage for this highest rating on the *HOPE Scale*. *Is self-aware* (Social 3) was also "often, almost always, and always" observed in their students' social behavior. *Is sensitive to larger or deeper issues of human concern* (Social 2) and *Effectively interacts with adults or older students* (Social 8) were "sometimes, rarely, or never" observed by teachers, and teachers "never" observed *Is a leader among his/her group of peers* (Social 5) for 21% of students; the students rated with this lowest Social score outnumbered all other students by 10% on the Social subscale.

Table 16 displays the correlation coefficients among these variables; the correlations ranged from .311 to .776 indicating moderate to strong relationships with no multicollinearity issues ($r > .8$). Interestingly, the *HOPE Scale* Academic and Social subscale scores represented

the highest correlation (r = .776), and the lowest correlations occurred among math and reading achievement test scores and the number of students' talented subjects as identified by their teachers (r = .311 and r = .336, respectively). Based on the inter-item correlation matrix shown in Table 17, the highest correlation (r = .897) was observed between Academic 6 (*Is eager to explore new concepts*) and Academic 7 (*Exhibits intellectual intensity*), and the lowest correlation (r = .211) occurred between Academic 11 (*Has intense interests*) and Social 4 (*Shows compassion for others*).

With regard to the acceptability of model fit indices, large chi-square values indicate a lack of model fit; however, because large sample sizes generally yield significant chi-square values, other measures need be considered (Kline, 1998). Garver and Mentzer (1999) suggested that values of Comparative Fit Index (CFI), Non-normed Fit Index (NNFI), and Root Mean Square Error of Approximation (RMSEA) be reviewed. Generally speaking, CFI and NNFI values greater than .90 indicate a good fit to the data. RMSEA values between .08 and .10 provide an acceptable fit, and values below .08 represent good fit (Browne & Cudeck, 1993).

Table 16. Correlation Matrix among Variables

	HOPE Academic	*HOPE* Social	Mathematics	Reading	No of Talented Subjects
HOPE Academic	1	.775*	.499*	.466*	.581*
HOPE Social		1	.411*	.442*	.486*
Mathematics			1	.646*	.311*
Reading				1	.336*
No of Talented Subjects					1

Note. *Correlation is significant at the .01 level.

Table 17. Inter-Item Correlation Matrix

	Aca1	Aca6	Aca7	Aca9	Aca10	Aca11	Soc2	Soc3	Soc4	Soc5	Soc8
Aca1	1										
Aca6	.670	1									
Aca7	.672	.897	1								
Aca9	.626	.733	.708	1							
Aca10	.534	.594	.570	.716	1						
Aca11	.717	.580	.560	.555	.559	1					
Soc2	.621	.725	.729	.650	.522	.515	1				
Soc3	.559	.619	.594	.613	.469	.467	.645	1			
Soc4	.291	.348	.325	.406	.274	.211	.417	.509	1		
Soc5	.544	.628	.593	.617	.488	.440	.565	.553	.367	1	
Soc8	.447	.541	.485	.616	.468	.385	.515	.519	.463	.590	1

Note. Aca = Academic; Soc = Social

Based on these criteria, the original model (Figure 2) did not fit the data well (x^2 values of 2701.455, df = 75, $p < .001$) with RMSEA = .174, NNFI = .706, and CFI = .757. Because our original model did not provide a good fit to our data, we made a series of modifications to the model and evaluated model fit after each change (Kline, 1998). See Table 18 for fit indices at each revision. Conveniently, LISREL 9.3 informs researchers about the fit index that, when allowed, will improve the model's fit to the data. Therefore, we admitted the correlated error (phi) between the *HOPE Scale*s overall Academic and Social subscales. This adjustment improved model fit, decreasing the chi-square value by 1,223.015 and improving RMSEA, NNFI, and CFI values to .128, .841, and .870, respectively, but model fit indices were still unacceptable. We continued to make three more revisions suggested by LISREL in the order that follows until model fit indices reflected acceptable levels: Revision 2) we allowed the correlation between Academic 6 and Academic 7 from the *HOPE Scale*'s Academic subscale ($\Delta\ x^2$=217.419, theta-delta); Revision 3) we admitted the correlation between reading and mathematics achievement scores ($\Delta\ x^2$=217.60,

theta-eps); and Revision 4) we allowed the correlation between Academic 1 and Academic 11 from the *HOPE Scale*'s Academic subscale (Δx^2=219.008, theta-delta). This final model shows a good fit with our data with a 74.33% reduction of x^2 value from the original model (x^2 values of 693.390, df = 71, p < .001) and other fit indices within acceptable levels—RMSEA = .087, CFI = .926, and NNFI = .926. The model's RMSEA value is marginal, indicating a slight lack of good fit; however, because our model contains a relatively small number of variables, larger RMSEA values can be expected regardless of model misfit (Kenny & McCoach, 2003).

Table 18. Model Revisions with Fit Indices

	Original (Base model)	Revision 1	Revision 2	Revision 3	Revision 4 (Final model)
Modification		Element (Aca, Soc) of PHI	Element (Aca7, Aca6) of Theta-delta	Element (Reading, Math) of Theta-eps	Element (Aca1, Aca11) of Theta-delta
Chi-Square	2701.455 (df = 75, p<.001)	1478.440 (df=74, p<.001)	1129.999 (df=73, p<.001)	912.398 (df=72, p<.001)	693.390 (df = 71, p<.001)
RMSEA	.174	.128	.112	.999	.087
NNFI	.706	.841	.878	.902	.926
CFI	.757	.870	.902	.922	.943

Note. Aca = Academic; Soc = Social

Table 19. Standardized Path Coefficients

Items	Coefficients	
LAMBDA-Y		
Performance		
Per1 (Reading)	.520	
Per2 (Mathematics)	.533	
Per3 (No. of Talented Subjects)	.615	
LAMBDA-X		
	HOPE Aca	*HOPE* Soc
Aca1	.763	
Aca6	.869	
Aca7	.850	
Aca9	.861	
Aca10	.717	
Aca11	.656	
Soc2		.822
Soc3		.772
Soc4		.516
Soc5		.741
Soc8		.688
GAMMA (Regression Matrix ETA on KSI)		
	HOPE Aca	*HOPE* Soc
Performance	1.025	-.048

Note. Per = Performance; Aca = Academic; Soc = Social

Next, standardized path coefficients in the final model were calculated for all the variables to show the relationship among variables. As a general rule, path coefficients between .10 to .20 represent small effects, between .30 to .40 indicate medium effects, and above .50 represent large effects of the variables within the fitted model (Kline, 1998). The results of the analysis showed that *HOPE Scale* Academic subscale scores had significant positive correlations with all indicators of students' performance (achievement scores in math and reading and the number of talented

subjects) with a path coefficient of 1.025. Based on the factor structure shown in Table 19, the six Academic items on the *HOPE Scale* loaded on the

Academic construct with strong correlations (range of .656 to .869), and the five Social items adequately loaded on the Social construct (range of .516 to .822). The Social construct showed a weak, negative relationship with students' academic performance with a path coefficient of -.05. Interestingly, the Social construct was positively associated with academic performance until revision 3 (path coefficient .077); however, the coefficient became negative in the final model when the correlation between Academic1 and Academic11 was allowed. This indicates that Social scores from the *HOPE Scale* may unreliably predict students' academic performance; whereas, Academic scores showed a strong relationship with performance across all modifications. The three items that represent students' academic performance (achievement test scores in mathematics and reading and the number of subjects in which teachers identified their students as talented) adequately loaded on the performance construct at .53, .52, and .62, respectively. See Figure 3 for the final model. From the analyses of the model we theorized, the performance represented by the mathematics scores, reading scores, and the number of talented subject areas which indicates the characteristics of multipotentiality have positive relationship with Academic score from the *HOPE Scale*, however, negative with Social scores. A negative relationship between the students' outcome performance and Social scores from the *HOPE Scale* was different from our hypothesis that we assumed it to have positive relationship similar to Academic score.

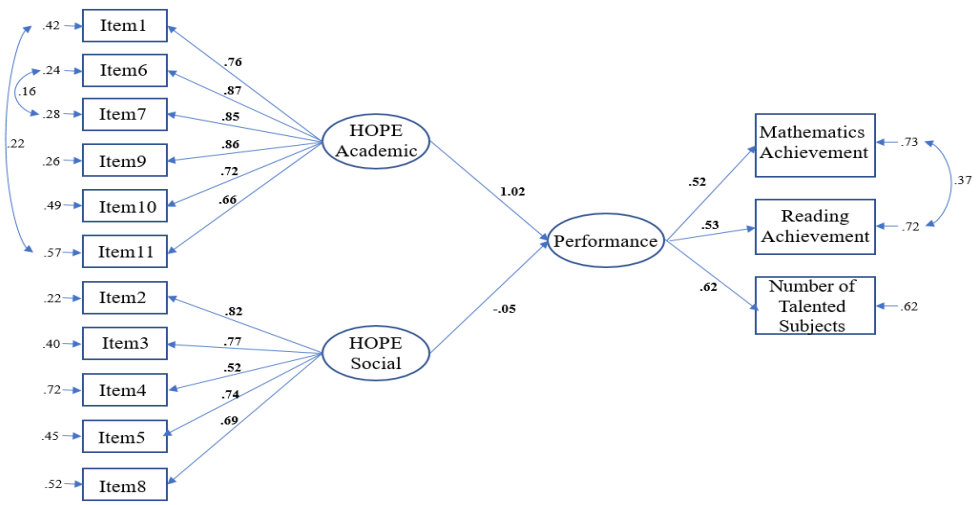

Figure 3. Path Coefficients in the Final Model

Phase II – Application of Various Cutoff Percentages and Group-specific Norms with *HOPE Scale* Scores and Achievement Test Results for Identification

In this study, we analyzed several identification scenarios (i.e., top 10%, top 25%) to determine their influences on the number of students from culturally diverse and/or low-income families identified for gifted services. For these calculations, we used the teachers' identification of students' lunch status to distinguish students from low-income families. Table 20 contains the cut scores and number of students in these populations when general norms are applied. When general norms were used with the top 10% of students' mathematics and reading achievement scores, only 9.56% of students from low-income families are identified for either subject. Similarly, only 11.30% and 10.43% of students from culturally diverse backgrounds are identified for gifted services in mathematics and reading, respectively with top 10% of scenario. Given that almost 19% of the sample in the study are from low-income families and slightly more than 14% have culturally diverse backgrounds, the proportion of low-income and/or culturally diverse students is clearly lower than the actual percentages in the total population.

Table 20. Cut Scores and Number of Students Identified Using General Norms

	Top 10%		Top 25%*	
	F/R** Lunch (%)	Paid Lunch (%)	F/R Lunch (%)	Paid Lunch (%)
Mathematics	*Cut score: 59.99*		*Cut score: 56.70*	
	11 (9.56)	104 (90.44)	35 (12.11)	254 (87.89)
Reading	*Cut score: 58.79*		*Cut score: 56.26*	
	11 (9.56)	104 (90.44)	34 (11.76)	255 (88.24)
	Culturally Diverse (%)	Korean (%)	Culturally Diverse (%)	Korean (%)
Mathematics	*Cut score: 59.99*		*Cut score: 56.70*	
	13 (11.30)	102 (88.70)	36 (13.90)	253 (86.10)
Reading	*Cut score: 58.79*		*Cut score: 56.26*	
	12 (10.43)	103 (89.57)	36 (13.90)	253 (86.10)

Note. 14% of the sample are from multicultural backgrounds and 19% are from low-income family.
* The number of students at Top 25% is cumulative and includes students in the Top 10%
**F/R = free/reduced lunch; our metric for low-income students

Not surprisingly, more students from these underrepresented groups were identified by lowering the cutoff percentage to the top 25% of student scores for both subjects which met our hypothesis. In this scenario, the cutoff scores for math and reading achievement tests dropped by 3.29 points and 2.53 points, respectively. The number of students from each underrepresented group increased by 23 or 24 students in each subject area. The percentages of culturally diverse students identified then remained as 13.90% for both math and reading which is close to the percentage of culturally diverse students participants in this study (14%). Although the number of students identified also increased for students whose families pay for their lunch and/or from Korean backgrounds, percentage gains for students from culturally diverse families outpaced their Korean peers by 4.20% (math) and 7.3% (reading), and percentage gains for students from low-income families outpaced their "paid lunch" peers by 15.57% (math) and 8.43% (reading).

When we applied group-specific norms with the same cutoff percentages to potentially yield more equitable representations of students from low-income and culturally diverse backgrounds, it was clear that the cut-scores of these students were lower for both subjects. Table 21 presents the revised cut scores using group-specific norms at the top 10% and top 25% of achievement test scores in both math and reading. The largest cutoff score gap (3.87 points) was from the top 25% mathematics achievement score between paid and F/R lunch students, and smallest cutoff score gap (.70 points) was from the top 25% reading achievement score between Korean students and students from culturally diverse backgrounds.

Table 21. Cut Scores When Using Group-Specific Norms

	Subgroups			
	F/R Lunch		Paid Lunch	
	Top 10%	Top 25%	Top 10%	Top 25%
Mathematics	57.86	53.38	60.25	57.22
Reading	58.07	54.46	58.9	56.62
	Culturally Diverse		Korean (Monocultural)	
	Top 10%	Top 25%	Top 10%	Top 25%
Mathematics	58.97	55.09	60.09	56.91
Reading	58.03	55.63	58.8	56.33

The next step in our analyses tested the effects of the addition of *HOPE Scale* teacher ratings of gifted behaviors as an additional tool to identify underrepresented populations who may be missed when using achievement data alone. As depicted in Table 22, we created 16 scenarios (A to P for mathematics achievement and A' to P' for reading achievement) with different percentage cutoffs for achievement combined with *HOPE Scale* Academic subscale scores (e.g., students from low-income families in top 10% of achievement using group-specific norms with students who scored in the top quartile of *HOPE Scale* scores and students from culturally diverse backgrounds in the top 25% of achievement using group-specific norms with

Table 22. Number of Students Identified When Combining Achievement Scores and *HOPE Scale* Scores for Placement

	HOPE Scale scores (Academic subscale)		
Mathematics achievement scores	Top quartile (score≥26)	Middle 50% (16≤score<26)	Bottom quartile (score<16)
All students in top 10% of general norm (*n*=115)	A (*n*=56)	B (*n*=52)	C (*n*=7)
Low-income students in top 10% of group-specific norm (*n*=21)	D (*n*=6)	E (*n*=15)	F (*n*=0)
Low-income students in top 25% of group-specific norm (*n*=56)	G (*n*=12)	H (*n*=36)	I (*n*=8)
Culturally diverse students in top 10% of group-specific norm (*n*=16)	J (*n*=6)	K (*n*=10)	L (*n*=0)
Culturally diverse students in top 25% of group-specific norm (*n*=41)	M (*n*=16)	N (*n*=23)	O (*n*=2)
All students who did *not* score in the top 25% of group-specific norm (*n*=869)	P (*n*=171) Low-income: 22 Culturally diverse: 21		
Reading achievement scores			
All students in top 10% of general norm (*n*=115)	A' (*n*=52)	B' (*n*=60)	C' (*n*=3)
Low-income students in top 10% of group-specific norm (*n*=21)	D' (*n*=7)	E' (*n*=10)	F' (*n*=4)
Low-income students in top 25% of group-specific norm (*n*=56)	G' (*n*=15)	H' (*n* =25)	I' (*n* =16)
Culturally diverse students in top 10% of group-specific norm (*n*=16)	J' (*n*=7)	K' (*n*=9)	L' (*n*=0)
Culturally diverse students in top 25% of group-specific norm (*n*=41)	M' (*n*=15)	N' (*n*=21)	O' (*n*=15)
All students who did *not* score in the top 25% of group-specific norm (*n*=869)	P' (*n*=165) Low-income: 19 Culturally diverse: 18		

Note. Top 10% of *HOPE Scale* cutoff score = 30

students who scored in the top quartile of *HOPE Scale* scores). We did not include the *HOPE Scale* Social subscale scores because of its weak relationship (r = -.048) to performance found in Phase I (see Table 19) and because Korean identification is based on advanced academic achievement

Based on our data using group-specific norms, *HOPE Scale* Academic subscale scores at or above 26 represented the top quartile of students who demonstrated the most gifted academic behaviors and scores below 16 represented the bottom quartile of students who demonstrated fewer gifted behaviors. With regard to students in the top 10% of general norms in Table 22, students in cells A/A,' B/B,' and C/C' would likely be identified as gifted in math or reading based on their high achievement scores alone. However, students in cells C/C' should receive special consideration for additional support services because they received very low *HOPE Scale* scores from their teachers despite their high achievement. These students may be overlooked by teachers, may lack self-regulated social and/or academic behaviors in the regular classroom setting, and/or experience conflict with their teachers. Similarly, students in cells D/D,' G/G,' J/J,' and M/M' should be considered for identification since they achieved in the top 25% in math and/or reading using group-specific norms and were rated in the top quartile for gifted behaviors by their teachers, even though their achievement test scores may be lower than their high-achieving peers who also demonstrate gifted academic behaviors.

The remaining eight cells in each scenario require further scrutiny in the identification process. Although students in these cells achieved scores in the top 10% or 25% in mathematics and reading when group specific norms are applied, they received lower *HOPE Scale* ratings from their teachers, indicating fewer gifted academic behaviors observed. According to Peters and Gentry (2012b), "All students who fall into any of the cells could potentially be labeled as

"gifted" or more appropriately could be identified for special services to help them grow based on their individual needs and current achievement level" (p.136). Put more directly, all students represented in the various scenario combinations deserve our careful attention and our intentional focus on what individual students need to be successful, academically and socially, rather than who will be identified for gifted programming.

The most promising cells of the scenarios are P/P'—171 students in mathematics (19.67%) and 165 students in reading (18.99%)—received a high score from their teachers on the *HOPE Scale* Academic subscale even though their achievement scores were below the top 25% cutoff score. These cells indicate that the *HOPE Scale* promotes the identification of students who exhibit gifted behaviors but who were missed when achievement test scores alone were used. Specifically, in our data, 22 students from low-income families (12.87%) and 21 students from culturally diverse backgrounds (12.28%) can now be considered candidates for advanced mathematics instruction, and 19 students from low-income families (11.52%) and 18 students from culturally diverse backgrounds (10.91%) can be considered for advanced reading instruction to fully develop their potentials. These students could gain access to gifted services using this alternative pathway, and they might enter the gifted program full-time or on a trial basis, where academic supports and exposure to appropriate challenge will likely promote higher academic achievement and motivation (Peters & Gentry, 2013).

Our final analysis examined students who were identified using the alternative pathway described above (i.e., cells P/P') more closely when applying the top 10% and 25% cutoffs in achievement score data and *HOPE Scale* scores in different combinations using group-specific norms. This analysis is presented graphically in Figure 4 using Venn diagrams. Students from underrepresented populations who qualified using the alternative pathway are represented within

the numbers on the right side of each Venn diagram. Four scenarios each for mathematics and reading achievement in combination with *HOPE Scale* Academic subscale scores show how many additional students could be considered for gifted programming. For example, when students scoring in the top 10% of mathematics achievement (n = 114) and the top 10% of *HOPE Scale* ratings (n = 121) were combined, an additional 94 students not within the top 10% of the achievement test can be considered for gifted programming. Among those 94 students, 11 students (11.70%) were from low-income families, and another 11 students (11.70%) were from culturally diverse backgrounds for an additional 22 students who may be identified. We recognize that students may qualify in both categories, but because only 2.85% of students live in low-income, culturally diverse families, we elected to report the numbers separately. In a similar fashion, when students in the top 10% of math achievement and the top 25% of the *HOPE Scale* ratings were combined, an additional 235 students were included for consideration creating a deeper, more representative pool from which to identify—23 students (9.79%) from low-income families and 24 students (10.21%) from culturally diverse backgrounds. Overall, when using alternative pathways at varying degrees of emphasis, an increase of 57 (top 25% math achievement and top 10% *HOPE Scale* scores) to 235 students (top 10% math achievement and top 25% *HOPE Scale* scores) may be considered for gifted services when compared with math achievement test scores alone.

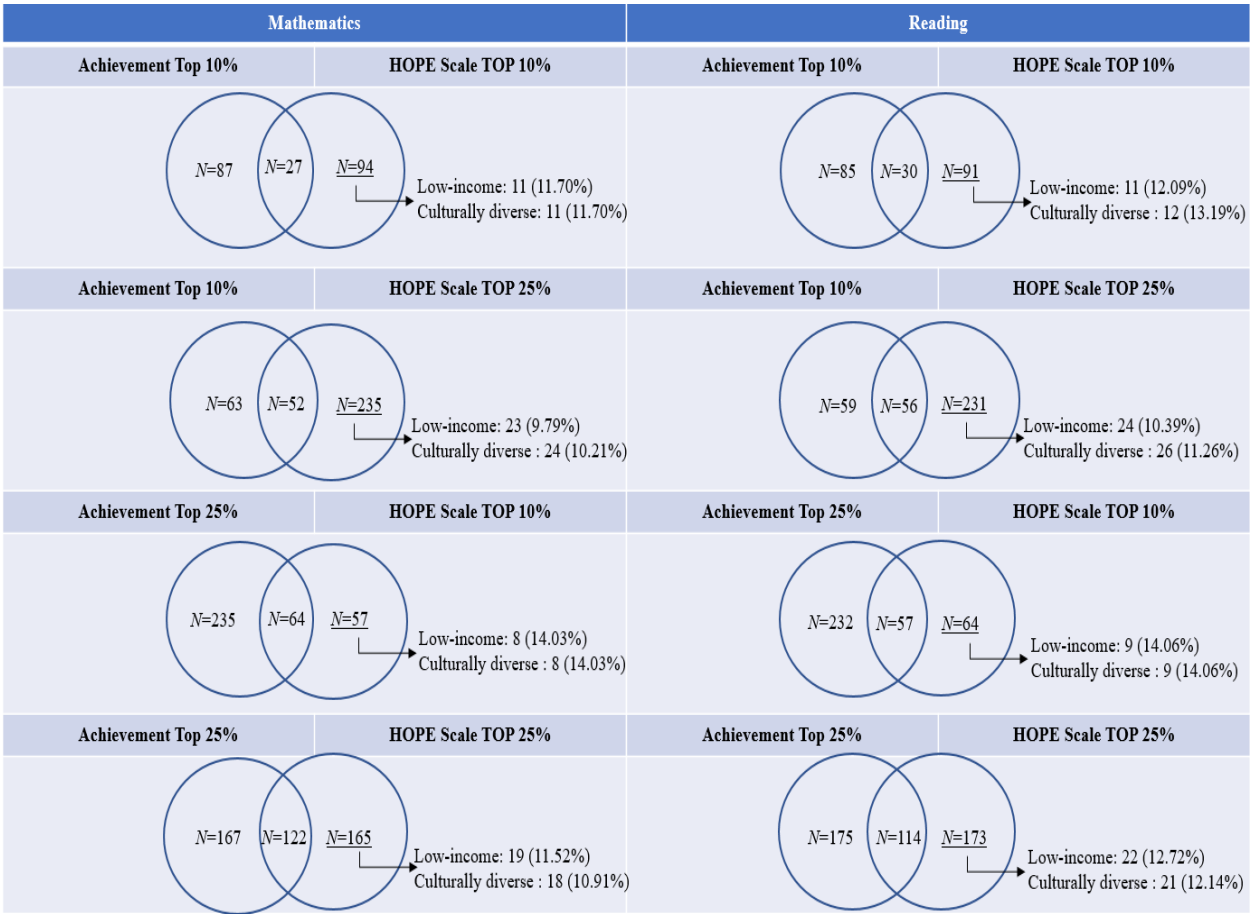

Figure 4. Venn Diagrams of Overlap between Achievement Test Scores and *HOPE Scale* Academic Scores

Results for identification for advanced reading instruction showed similar outcomes. When students scoring in the top 10% of reading achievement (n = 115) and the top 10% of *HOPE Scale* ratings (n = 121) were combined, an additional 91 students not within the top 10% of the achievement test group can be considered for gifted programming. Among those 91 students, 11 students (12.09%) were from low-income families, and another 12 students (13.19%) were from culturally diverse backgrounds for an additional 23 students who may be identified. In a similar fashion, when students in the top 10% of reading achievement and the top 25% of the *HOPE Scale* ratings were combined, an additional 231 students were included for consideration creating a

deeper, more representative pool from which to identify—24 students (10.39%) from low-income families and 26 students (11.26%) from culturally diverse backgrounds. Overall, when using alternative pathways at varying degrees of emphasis, an increase of 64 (top 25% reading achievement and top 10% *HOPE Scale* scores) to 231 students (top 10% reading achievement and top 25% *HOPE Scale* scores) may be considered for gifted services in reading when compared with reading achievement test scores alone. Clearly, all eight scenarios increased the number of students from underrepresented populations for consideration for gifted programming. Interestingly, however, the percentage increases of students from low-income families and students from culturally diverse backgrounds included for identification in math (14.03%) and, separately, in reading (14.06%) were highest when students with the top 25% of math or reading achievement scores were combined with students in the top 10% on *HOPE Scale* ratings.

The pattern of responses in the teachers' *HOPE Scale* ratings of academic behaviors for students from low-income or culturally diverse families as shown in Table 22 also merits attention. Note that when group-specific norms are applied from the top 10% to the top 25% of math achievement scores, teachers' *HOPE Scale* ratings tend toward the top quartile for students from low-income or culturally diverse families. However, when this scenario is repeated with reading achievement scores, teachers' ratings follow a more normal distribution, with more students from low-income or culturally diverse families identified in the bottom quartile of *HOPE Scale* scores. The overall results imply that when lowering the cutoff percentages of the achievement test score, more of students from low-income and/or culturally diverse backgrounds can be identified, however, when combining the results of achievement test scores and teacher-rating scales, the percentage of students from low-income and/or culturally diverse backgrounds would remain same.

This indicates that when using multiple criteria, OR method should be considered rather than AND method if the purpose of the identification is to increase more students with diverse backgrounds.

Discussion

Students from low-income families and/or culturally diverse backgrounds are less frequently identified as gifted since their achievement test scores are typically lower than their counterparts (Callahan, 2005; Peters & Gentry, 2010). The fairest representation of students from low-income and culturally diverse backgrounds was the goal for identification; when group-specific norms are used with achievement test scores in conjunction with teachers' ratings of academic behaviors using the *HOPE Scale*, we found that more underrepresented students can be identified when applying different conjunction scenarios. These additional students may be poor test takers, underachievers, students with fewer opportunities to learn outside of school, or students who were unidentified using achievement tests' general norms alone. (Peters & Gentry, 2013). Since *HOPE Scale* scores represent teachers' ratings of students' behavior when compared to other students with similar backgrounds, the combination of group-specific norms applied to achievement test results and *HOPE Scale* scores more equitably identifies students who may have been overlooked for gifted services.

Although Lohman and Renzulli (2007) claimed when teacher ratings are weighted heavily, the diversity of the population of students who are identified to a gifted program increases, our results did not completely support this idea. Recall that the largest percentage increase in students from low-income or culturally diverse families was achieved in the scenarios that included the top 25% of achievement test scores and the top 10% of *HOPE Scale* Academic scores. In this study, we used teachers' ratings of their students' academic behaviors in addition to achievement test data to determine whether more students from underrepresented populations could be considered

for gifted services. Our results indicate that using group-specific norms and lowering achievement test cut-off scores to the top 25% in combination with more restrictive, top 10% of *HOPE Scale* academic behaviors ratings yields the most representative pool of students for advanced coursework. In other words, the greatest percentage of students from low-income or culturally diverse backgrounds are identified when cut-off scores for achievement are lowered and these students are compared academically and behaviorally to their classmates similar in age, experience, and/or environment. This holds true for teachers' observations, experiences, and perceptions of students' gifted behaviors in math and reading, even when teachers observed fewer academic behaviors in reading.

In this study, the pattern of responses in teachers' *HOPE Scale* scores may indicate some behaviors may be subject-specific, with academic behaviors more observable in math than in reading. This pattern may also indicate that teachers may provide more opportunities for students to behave academically in math than in reading or that teachers are able to judge academic behaviors in math more confidently than they can in reading. On the other hand, the pattern may reflect teachers' negative perceptions of students from low-income or culturally diverse families, especially as it relates to reading. Reading is fundamental to learning, and students who struggle to read may be perceived as less capable and, therefore, not gifted. Teachers may also hold traditional conceptions about giftedness as a general intellectual ability or as indicated by excellent grades, and their ratings may reflect their failure to recognize gifted behaviors in students from underrepresented populations. Cultural expectations may have also influenced *HOPE Scale* ratings; it is well-known that Asian societies highly value education and achievement and expect students to learn and behave academically. As stated earlier, students who scored in the bottom quartile of the *HOPE Scale* and the top 10% of general norms in math (cell C, $n = 7$) and reading (cell C', *n*

= 3) should receive special consideration in the identification process. Additional programming options and/or educational supports can be offered to meet their academic and behavioral needs, especially when underachievement and/or twice exceptionality is identified.

"Alternative pathways" (Peters & Gentry, 2013) must continue to be explored for more equitable gifted identification. The limitations of state laws, unequal access, inequity, and missingness, coupled with rampant imperfections and misuses of educational assessment results has created in a broken system for identifying students' gifts and talents. This study supports previous research that calls for the use of multiple, diverse measures and the application of group-specific norms to standardized test results and teacher rating scales for more equitable identification procedures. When given proper training and high-quality assessment tools that yield valid and reliable results, teachers' participation in this process will become more effective.

Limitations and Implications for Future Research

As indicated in the Methods section, students' achievement test scores used in this study are not national- nor state-level data, but rather school- or classroom-level data. Although we standardized the scores so that students' scores across classrooms or schools could be compared and analyzed, they do not represent a standardized measure of students' achievement. In addition, although the sample in this study includes more students from culturally diverse backgrounds (14.43%) than is present in the total population of South Korea, our findings reflect more of a special case than a generalizable conclusion that can be declared across the country's schools. The lack of cultural diversity in most schools, however, does not lessen the importance of this study with regard to students from low-income families as these students can be found in every school.

A key finding in this study was that the *HOPE Scale*'s Social subscale may unreliably predict students' academic performance as measured by achievement test results and the number

of talented subjects that teachers identified for each student. Recall that the *HOPE Scale* was developed to identify academic and social behaviors that may indicate giftedness, especially for students from underrepresented populations who may not "shine" academically. Therefore, students whose Social subscale scores rate in the top quartile should be considered for identification, even in (or especially in) situations where academic achievement is lacking. In addition, students whose Social subscale scores consistently rate in the bottom quartile may be further scrutinized, especially in situations where the teacher is able to look past the socially unacceptable behavior to see students' academic potential. Further research into the teachers' role in identification using these scenarios may provide insights that improve the identification of students' gifts, talents, and learning strengths.

CHAPTER 5. CONCLUSION

Major Findings of Each Study

As the academic achievement gap among and between different income and ethnic groups has persisted worldwide for decades, diverse fields in education have tried to reduce the gap (Byers, 2018; Mortenson, 2018; Webb & Thomas, 2015). It was not an exception in the field of gifted education, and researchers and practitioners have highlighted the inequity issues of underserved populations in gifted programming and tried to decrease the "excellence gaps" (Plucker et al., 2010, 2013). This dissertation was designed with excellence gaps in mind and to and provide rationales for how the field can move forward to identify gifted and talented students more equitably. In particular, this study focused on investigating the teacher-rating scale called the *HOPE Scale* (Gentry et al., 2015) to examine whether the measure can be an effective tool to help identify underserved populations for gifted programming.

As discussed in chapters 2, 3, and 4, this research was composed of three research papers titled, (a) Validity Evidence for the *HOPE Scale* to Identify Gifted Students from Low-Income and Multicultural Families in Korea, (b) The Effects of Teacher and Classroom Characteristics on Students' Teacher-rating Scale Outcomes: MLM Analysis, and (c) Gifted Identification Using the *HOPE Scale-Korean Version* and Achievement Data: Associations, Combinations, and Group-specific Norms. Through these three papers, I first wanted to explore validity evidence the *HOPE Scale* Korean data to determine whether the *HOPE Scale* is biased against certain groups. The teachers' interviews addressing their perceptions on the *HOPE Scale* and the 11 items included in the measure were also explored to understand its practicability. Then, as a second step, I conducted an additional analysis with multilevel modeling to understand if any classroom or teacher factors affect the *HOPE Scale* scores, along with the student's individual characteristics. As the last step,

I examined the relationship of the *HOPE Scale* scores with students' reading and mathematics achievement data to understand the usefulness of combining both types of data to better identify underserved students with gifts and talents. Following are the abstracts of each study, which briefly explain the major findings.

Study 1 Abstract

The underrepresentation of students from low-income families and culturally diverse backgrounds is a long-standing and pervasive problem in the gifted education field. Teachers play an important role in equitably identifying and serving students in gifted education programs; therefore, the *HOPE Scale* was used in this study with a sample of Korean elementary school teachers (n=55) and their students (n=1,157). Confirmatory factor analysis and multi-group confirmatory factor analysis results suggested the *HOPE Scale* shows the equivalence of model form, factor loading, and factor variances across different income and ethnic groups. A follow-up interview with teachers (n=6) revealed they acknowledged the importance of using the *HOPE Scale* as an additional method for identifying gifted students; however, they indicated less confidence about rating gifted students' social characteristics compared to the academic components in the *HOPE Scale*.

Study 2 Abstract

Teachers play an important role in identifying gifted students as they are often the gatekeepers for the programming. Teachers rate the outcome variables used in this study, the *HOPE Scale* scores, to evaluate each student in their classroom. Therefore, the results might differ by teacher as students are nested in the classroom with the same teacher. From the multilevel modeling analyses investigating the pattern of the outcome based on 1,113 students from 53

teachers, we found that level-1 factors affected students' received scores and that students from low-income families and multicultural backgrounds had slightly lower scores compared to their counterparts. Students from low-income families and multicultural backgrounds received scores from their teachers averaging .63 (*SE*=.08) and .23 (*SE*=.11) points less (on an 6-point scale), respectively, than their peers on the Academic scaled score. Similar results existed for the Social scale score with students from low-income families and multicultural backgrounds receiving scores .37 (*SE*=.08) and .30 (*SE*=.10) lower than their counterparts. However, level-2 factors, such as a teacher's years of experience and the classroom characteristics (percentage of low-income students and multicultural students in the classroom), did not influence the *HOPE Scale* scores differences from the teachers.

Study 3 Abstract

Underrepresentation of students from low-income families and/or certain ethnic groups has been a serious worldwide problem in the field of gifted education. Because teachers frequently serve as gatekeepers for gifted programming, their roles in the identification process are critical. In this study, 55 elementary school teachers in South Korea completed a teacher-rating scale (*HOPE Scale*; Gentry et al., 2015) on their students (n = 1,157). In Phase I, the results were examined in relationship to the students' reading and mathematics achievement scores, as well as the number of subject areas in which students show talent as rated by their teachers; our findings showed that students' performance outcome has positive relationship with Academic score but negative relationship with Social score. In Phase II, we explored the effects of applying various cut-off scores in the combination of students' achievement scores and the *HOPE Scale* results; the results implied that underrepresented students were more identified by lowering the cutoff percentages of achievement test score, however, the percentage of underrepresented students

remained same when combining the results of achievement test scores and teacher-rating scales, indicating the needs of multiple criteria and using "OR" method rather than "AND" method for the identification.

Implications and Suggestions for Future Research

Investigating a teacher-rating scale designed to equitably identify traditionally underserved populations was one of the ways to positively move forward the field of gifted education. Diverse efforts like exploring and inventing unbiased identification measures and developing and increasing services for underrepresented students need to be continually provided to move this field forward. As noted by Peters and Pereira (2017), replication studies have rarely been conducted in the gifted education. In their study, Peters and Pereira (2017) replicated the published models of three teacher-rating scales (i.e., *HOPE Scale*, SIGS, GRS) using CFA to evaluate the internal validity structures of the instruments. Unfortunately, the findings showed all three instruments failed to meet the criteria of fit indices in the replication study. As such, instrument revision and continuous replication of the instruments with different samples are needed for the measurement to be widely used with full confidence. The Korean version of the *HOPE Scale* examined in this study met the criteria of the model fit with Korean data as evidenced by CFA and MCFA reuslts; however, continued study with different samples would enable wider application of the instrument in different contexts with different groups.

Sample

The sample collected for this study contained 1,157 students from 55 teachers in nine different public elementary schools located in Seoul, South Korea. I used these data overall in this dissertation, except for the second study in which I had to delete the data from two teachers and

their students because of missing level-2 factor information that prevented running multilevel modeling analyses. Although our sample consisted of 14% multicultural students, an over representation of South Korea's current ethnic demographics where only 1–2% of the population are from multicultural family backgrounds, it would have been interesting to have collected data of more students from multicultural or low-income family backgrounds to meet this study's purposes (Sung, 2017). For example, to make a direct comparison of the goodness of fit indices in MCFA analyses in the first study, I randomly selected cases from the general Korean student data to have the same number as students from multicultural backgrounds (n=167) and from low-income families (n=216). If more samples were from multicultural or low-income students, the study's power of invariance testing would increase. In addition, samples of multicultural families collected in this study are not fully representative of South Korea's population as they were all from metropolitan areas. Families from international marriages in rural areas might have different living patterns, which might affect students' behaviors and performance. As indicated earlier, the multicultural samples were all categorized as Asian in this study, as at least one of their parents was from another Asian country. If a future study more deeply explores education in relationship to different ethnic groups in South Korea, the researchers can selectively collect samples of different ethnic origins and incorporate more diversity.

Culture

The Korean version of the *HOPE Scale* used in this study was created through multiple steps of translating and back-translating by many educators, either from gifted education backgrounds or who were bilingual in Korean and English and who also majored in education. Then the measure was explored with multiple research methods to understand how *HOPE Scale* – Korean can be effectively used in South Korea as an additional tool to equitably identify more

students who have been traditionally underserved in the gifted programming. Throughout this dissertation I focused on investigating the one instrument, the *HOPE Scale*, but not extensively explore the value or characteristics of the gifted education system in South Korea. Nor did I investigate how teachers perceived the services and identification of students for gifted programming in general, which may also include diverse underrepresentation issues. Although brief interviews with six teachers were conducted to understand their perceptions of using the *HOPE Scale*, their preferences of items included in the *HOPE Scale*, and their perception about supporting traditionally underrepresented students in gifted programming, the data were insufficient to represent the teachers' general ideas or perceptions about the public elementary school systems in Korea.

If the study did not use the translated scale developed in the U.S. but rather aimed to invent a new teacher-rating scale reflecting South Korea's true educational value regarding gifted education, new items might have been created to represent the giftedness. Future studies, therefore, would include the Exploratory Factor Analyses to evaluate the newly-developed items. The newly-created items might function better to identify students from diverse backgrounds as they would take into account culture and values in the Korean context. For example, from the first study, although the translated *HOPE Scale* showed equal functioning for both groups (i.e., Korean vs. multicultural, low-income vs. regular or high-income families) from the invariance testing, teachers indicated that they were unsure about certain items from the Social construct. As such, it would have yielded valuable data to fundamentally understand the teachers' understanding and perception of giftedness, particularly on the social aspect, if I had conducted an exploratory study about developing items that represent the concept of giftedness. This also implies that teacher

training is needed to sensitize them to the social and emotional characteristics of gifted students as teachers' unsureness of Social items does not necessarily indicate they are inadequate.

Social Construct of Giftedness

I did not include specific research questions or extensively explore how the social scale would have affected identification of gifted students in South Korea. Future studies should explore the social construct of giftedness to investigate how effectively it can be used with a conjunction of students' academic outcomes. In my studies in chapter 2 and 4, teachers were unsure about social items and the social construct was not positively, although weakly, associated with students' academic achievement scores. This indicates that social construct may function as an alternative pathway identifying students with gifts and talents. This finding also supports that the *HOPE Scale* should keep academic and social constructs separate, rather than combining two, despite the high correlation of *HOPE* academic score and *HOPE* social score. It might be an evidence that social characteristics as its own construct uniquely contribute to understanding and finding students' giftedness, creativity, and talents.

Analyses

In this dissertation, I not only examined the validity evidence of the *HOPE Scale,* but also used MLM, SEM, and examined multiple scenarios by combining *HOPE Scale* scores with achievement data. Future studies could investigate combining the scenarios with different measures, such as intelligence tests, creativity measures, and other alternative measures, that can be used as multiple criteria for identifying gifted students. The *HOPE Scale's* relationship with other teacher-rating scales can also be compared and explored. Within a comparison study, future researchers may explore which measure better identifies the underrepresented students. In terms

of the model building in the second study, I built a two-level model based on the information I collected; however, a future study could conduct a larger study with level-3 data, such as school and district characteristics, when analyzing the patterns. In addition, the *HOPE Scale* was translated into the Korean language, however, it can be translated into different languages and introduced in other countries where teacher-rating scales have not been sufficiently developed or used. Because of its simple structure and the purpose of equitably identifying gifted and talented students from diverse backgrounds, this measure could help address inequity in other countries. The findings from chapter 2 and 4 showed similar results with U.S findings that *HOPE Scale* was invariant across the different ethnic and income groups, and more underrepresented students could be identified by using group-specific norms compared to using traditional general norm-group criteria. However, cross-country comparison of the invariance testing by including diverse countries would build a strong rationale for supporting sound identification measures.

REFERENCES

Abu, N. K., Akkanat, Ç., & Gökdere, M. (2017). Teachers' views about the education of gifted students in regular classrooms. *Turkish Journal of Giftedness and Education, 7*(2), 87-109. https://dergipark.org.tr/en/pub/tuzed/issue/58918/849469

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. http://dx.doi.org/10.1177/0016986216634438

Adelson, J. L. (2012). Examining relationships and effects in gifted education research: An introduction to structural equation modeling. *Gifted Child Quarterly, 56*(1), 47-55. https://doi.org/10.1177/0016986211424132

Allen, J. K. (2017). Exploring the role teacher perceptions play in the underrepresentation of culturally and linguistically diverse students in gifted programming. *Gifted Child Today, 40*(2), 77-86. https://doi.org/10.1177/1076217517690188

Ambrose, D. (2002). Socioeconomic stratification and its influences on talent development: Some interdisciplinary perspectives. *Gifted Child Quarterly, 46*(3), 170-180. http://dx.doi.org/10.1177/001698620204600302

Barton, P. E., & Coley, R. J. (2009). Measuring the achievement elephant. *Educational Leadership*, *66*(4), 30-34.

Baudson, T. G., Fischbach, A., & Preckel, F. (2014). Teacher judgments as measures of children's cognitive ability: A multilevel analysis. *Learning and Individual Differences, 52*, 148-156. https://doi.org/10.1016/j.lindif.2014.06.001

Baudson, T. G., & Preckel, F. (2016). Teachers' conceptions of gifted and average-ability students on achievement-relevant dimensions. *Gifted Child Quarterly, 60*(3), 212-225. https://doi.org/10.1177/0016986216647115

Benson, J., & Borman, G. (2010). Family, neighborhood, and school settings across seasons: When do socioeconomic context and racial composition matter for the reading achievement growth of young children. *Teachers College Record, 112*(5), 1338-1390.

Benson, N. F., & Kranzler, J. H. (2018). Another Look at the Construct Validity of the Gifted Rating Scales: Preschool/Kindergarten and School Forms. *Journal of Psychoeducational Assessment*, *36*(8), 782-797. https://doi.org/10.1177/0734282917710377

Bernie, V. H., & Beilke, J. R. (2008). The constraints of poverty on high achievement. *Journal for the Education of the Gifted, 31*(3), 295-321. https://doi.org/10.4219/jeg-2008-771

Bianco, M., Harris, B., Garrison-Wade, D., & Leech, N. (2011). Gifted girls: Gender bias in gifted referrals. *Roeper Review, 33*(3), 170-181. http://dx.doi.org/10.1080/02783193.2011.580500

Bianco, M., & Leech, N. L. (2010). Twice-exceptional learners: Effects of teacher preparation and disability labels on gifted referrals. *Teacher Education and Special Education, 33*(4), 319-334. http://dx.doi.org/10.1177/0888406409356392

Borland, J. H. (2004). *Issues and practices in the identification and education of gifted students from under-represented groups*. The National Research Center on the Gifted and Talented.

Borland, J. H., Schnur, R., & Wright, L. (2000). Economically disadvantaged students in a school for the academically gifted: A postpositivist inquiry into individual and family adjustment. *Gifted Child Quarterly*, *44*(1), 13–32. https://doi.org/10.1177/001698620004400103

Borland, J. H., & Wright, L. (1994). Identifying young, potentially gifted, economically disadvantaged students. *Gifted Child Quarterly*, *38*(4), 164–171. https://doi.org/10.1177/0016986294030042

Borland, J. H., & Wright, L. (2000). Identifying and educating poor and under-represented gifted students. In K. A. Heller, F. J. Mönks, R. J. Sternberg, & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (2nd ed., pp. 587–594). Pergamon.

Boyd, S, N. (2015). Teachers Perceptions of Reading Ability Groups for African American Compared to Caucasian Elementary Students (Publication No. 3714905) [Doctoral dissertation, Northcentral University]. ProQuest Dissertations Publishing.

Bracken, B. A., & McCallum, S. (2018). *UTAGS complete kit: Universal talented and gifted screener*. Prufrock Press.

Brown, S. W., Renzulli, J. S., Gubbins, E. J., Siegle, D., Zhang, W., & Chen, C. H. (2005). Assumptions underlying the identification of gifted and talented students. *Gifted child quarterly, 49*(1), 68-79. https://doi.org/10.1177/001698620504900107

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Sage.

Byers, J. (2018). The Impact of Teacher Trust on the Achievement Gap Among African American Male Students (Publication No. 10825850) [Doctoral dissertation, Northcentral University]. ProQuest Dissertations Publishing.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge.

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

Callahan, C. M. (2007). What can we learn from research about promising practices in developing the gifts and talents of low income students? In. J. VanTassel-Baska & T. Stambaugh (Eds.), *Overlooked gems: A national perspective on low-income promising learners* (pp. 53-56). National Association for Gifted Children.

Callahan, C. M., Moon, T. R., & Oh, S. (2013). *Status of elementary gifted programs.* National Research Center on the Gifted and Talented.

Camilli, G. (2006). Test fairness. In Brennan, R. L., & National Council on Measurement in Education (Eds.). *Educational measurement*, (pp. 221-256). Praeger Publishers.

Card, D., & Giuliano, L. (2015). *Can universal screening increase the representation of low income and minority students in gifted education?* https://www.pnas.org/content/113/48/13678

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological methods & research, 36*(4), 462-494. https://doi.org/10.1177/0049124108314720

Cherng, H. (2017). If they think I can: Teacher bias and youth of color expectations and achievement. *Social Science Research, 66*, 170-186. https://doi.org/10.1016/j.ssresearch.2017.04.001

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory Into Practice, 39*(3), 124-130. https://doi.org/10.1207/s15430421tip3903_2

Dai, D. Y., & Chen, F. (2013). Three paradigms of gifted education: In search of conceptual clarity in research and practice. *Gifted Child Quarterly, 57*(3). 151-168. https://doi.org/10.1177/0016986213490020

Denzin, N. K., & Lincoln Y. S. (1994). *Handbook of Qualitative Research*. Sage.

Duncan, G. J., & Murnane, R. J. (2011). *Whither opportunity?: Rising inequality, schools, and children's life chances*. Russell Sage Foundation.

Elhoweris, H., Mutua, K., Alsheikh, N., & Holloway, P. (2005). Effect of children's ethnicity on teachers' referral and recommendation decisions in gifted and talented programs. *Remedial and Special Education, 26*(1), 25-31. http://dx.doi.org/10.1177/07419325050260010401

Endepohls-Ulpe, M. (2005). Primary school teachers' criteria for the identification of gifted pupils. *High Ability Studies, 16*(2), 219-228. http://dx.doi.org/10.1080/13598130600618140

Esquierdo, J. J., & Arreguín-Anderson, M. (2012). The "invisible" gifted and talented bilingual students: A current report on enrollment in GT programs. *Journal for the Education of the Gifted, 35*(1), 35-47. https://doi.org/10.1177/0162353211432041

Everson, H. T., & Millsap, R. E. (2004). Beyond individual differences: Exploring school effects on SAT scores. *Educational Psychologist, 39*(3), 157-172. https://doi.org/10.1207/s15326985ep3903_2

Fan, X., & Silvo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research, 42*(3), 509-529. https://doi.org/10.1080/00273170701382864

Fischer, R., & Karl, J. A. (2019). A primer to (cross-cultural) multi-group invariance testing possibilities in R. *Frontiers in psychology, 10*, 1-18. https://doi.org/10.3389/fpsyg.2019.01507

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

Ford, D. Y., & Moore, J. L. (2013). Understanding and reversing underachievement, low achievement, and achievement gaps among high-ability African American males in urban school contexts. *The Urban Review, 45*(4), 399-415. https://doi.org/10.1007/s11256-013-0256-3

Frederickson, H. G. (1990). Public administration and social equity. *Public Administration Review, 50*(2), 228-237.

French, B. F., & Finch, W. H. (2006). Confirmatory factor analytic procedures for the determination of measurement invariance. *Structural Equation Modeling, 13*(3), 378-402. https://doi.org/10.1207/s15328007sem1303_3

Garver, M. S., & Mentzer, J. T. (1999). Logistics research methods: employing structural equation modeling to test for construct validity. *Journal of business logistics, 20*(1), 33-57.

Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment, 6*(4), 304-312. https://doi.org/10.1037/1040-3590.6.4.304

Gelman, A., and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press.

Gentry, M., Gray, A., Whiting, G., Maeda, Y., & Pereira, N. (2019). *Gifted education in the United States: Laws, access, equity, and missingness across the country by locale, Title I school status, and race* [Executive summary]. https://www.education.purdue.edu/geri/new-publications/gifted-education-in-the-united-states/

Gentry, M., Pereira, N., Peters, S. J., McIntosh, J. S., & Fugate, C. M. (2015). *HOPE teacher rating scale administration manual*. Prufrock Press.

Gentry, M. & Seward, K. (2018). Gifted students from low socioeconomic backgrounds. In J. L. Roberts, T. F. Inman, & J. H. Robins (Eds.), *Introduction to gifted education* (pp.353-368). Prufrock Press.

Gibbons, M. M., Pelchar, T. K., & Cochran, J. L. (2012). Gifted students from low-education backgrounds. *Roeper Review, 34*(2), 114-122. https://doi.org/10.1080/02783193.2012.715332

Gilliam, J. E., Carpenter, B. O., & Christensen, J. R. (1996). *Gifted and Talented Evaluation Scales*. Pro-Ed.

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

Grissom, J. A., Redding, C., & Bleiberg, J. F. (2019). Money over merit? Socioeconomic gaps in receipt of gifted services. *Harvard Educational Review, 89*(3), 337-369. https://doi.org/10.17763/1943-5045-89.3.337

Guadagnoli, E. and Velicer, W. F. (1988) Relation of sample size to the stability of component patterns. *Psychological Bulletin, 103*(2), 265-275. https://doi.org/10.1037/0033-2909.103.2.265

Gubrium, J. F., & Holstein, J. A. (2001). *Handbook of interview research: Context and method*. Sage Publications.

Hambleton, R. K. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment, 17*(3), 164-172. https://doi.org/10.1027//1015-5759.17.3.164

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

Hammerschmidt, M. M. (2016). *Teacher perceptions regarding the teaching of gifted students in the traditional classroom setting* [Unpublished Doctoral dissertation]. Texas Tech University. https://ttu-ir.tdl.org/handle/2346/72341

Hansen, J. B., & Feldhusen, J. F. (1994). Comparison of trained and untrained teachers of gifted students. *Gifted Child Quarterly, 38*(3), 115-121. https://doi.org/10.1177/001698629403800304

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2009). New evidence about Brown v. Board of Education: The complex effects of school racial composition on achievement. *Journal of Labor Economics, 27*(3), 349-383. https://doi.org/10.1086/600386

Harradine, C. C., Coleman, M. R. B., & Winn, D. M. C. (2014). Recognizing academic potential in students of color: Findings of U-STARS~ PLUS. *Gifted Child Quarterly, 58*(1), 24-34. https://doi.org/10.1177/0016986213506040

Henfield, M. S., Woo, H., & Bang, N. M. (2017). Gifted ethnic minority students and academic achievement: A meta-analysis. *Gifted Child Quarterly, 61*(1), 3-19. https://doi.org/10.1177/0016986216674556

Hodge, R. D., & Cudmore, L. (1986). The use of teacher-judgment measures in the identification of gifted pupils. *Teaching and Teacher Education, 20*(1), 181-196. https://doi.org/10.1016/0742-051X(86)90016-8

Horn, J. L., & McArdle, J. J (1992). A practical and theoretical guide to measurement invariance in aging research. Experimental Aging Research, 18(3), 117-144. https://doi.org/10.1080/03610739208253916

Hoxby, C. M. (2000). *Peer effects in the classroom: Learning from gender and race variation* (Working Paper No. 7867). National Bureau of Economic Research. https://www.nber.org/papers/w7867

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives, *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Jöreskog, K. G. (n.d.). LISREL 9.30. http://www.ssicentral.com/index.php/products/lisrel/free-downloads

Jarosewich, T., Pfeiffer, S. I., & Morris, J. (2002). Identifying gifted students using teacher rating scales: A review of existing instruments. *Journal of Psychoeducational Assessment, 20*(4), 322-336. https://doi.org/10.1177/073428290202000401

Kaiser, J., Südkamp, A., & Möller, J. (2017). The effects of student characteristics on teachers' judgment accuracy: Disentangling ethnicity, minority status, and achievement. *Journal of Educational Psychology*, *109*(6), 871-888. https://doi.org/10.1037/edu0000156

Kaplan, D., & George, R. (1995). A study of the power associated with testing factor mean differences under violations of factorial invariance. *Structural Equation Modeling, 2*(2), 101-118. https://doi.org/10.1080/10705519509539999

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling, 10*(3), 333-351. https://doi.org/ 10.1207/S15328007SEM1003_1

Kettler, T., & Bower, J. (2017). Measuring creative capacity in gifted students: Comparing teacher ratings and student products. *Gifted Child Quarterly, 61*(4), 290-299. https://doi.org/10.1177/0016986217722617

Kerr, B., & Sodano, S. (2003). Career assessment with intellectually gifted students. *Journal of Career assessment, 11*(2), 168-186. https://doi.org/10.1177/1069072703011002004

Kline, R. (1998). *Principles and practice of structural equation modeling*. The Guilford Press.Korean Educational Development Institute (2015). *Gifted Education Statistical Yearbook 2015*. https://ged.kedi.re.kr/stss/main.do

Korean Educational Development Institute (2015). *Gifted Education Statistical Yearbook 2015*. https://ged.kedi.re.kr/stss/main.do

Korean Educational Development Institute (2016). *Gifted Education Statistical Yearbook 2016*. https://ged.kedi.re.kr/stss/main.do

Kovacs, K., & Conway, A. R. (2019). A unified cognitive/differential approach to human intelligence: Implications for IQ testing. *Journal of Applied Research in Memory and Cognition. 8*(3), 255-272. https://doi.org/10.1016/j.jarmac.2019.05.003

Kuzel, A. J. (1999). Sampling in Qualitative Inquiry. In B. F. Crabtree & W. L. Miller (Eds.), *Doing Qualitative Research* (2nd ed, pp.33-45). Sage.

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

Lakin, J. M. (2012). Multidimensional ability tests and culturally and linguistically diverse students: Evidence of measurement invariance. *Learning and Individual Differences, 22(3), 397-403*. https://doi.org/10.1016/j.lindif.2011.12.003

Lakin, J. M. (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

Lee, H., & Gentry, M. (2019, July). *Validation study of the Hope Scale: Identifying gifted students from low-income and multicultural families*. Session presented at the 23rd World Council for Gifted and Talented Children (WCGTC) Conference, Nashville, TN.

Lee, S., & Eum, Y. (2018). Analysis of research trends on economically disadvantaged gifted education. *Journal of Holistic Convergence Education. 22*(2), 29-54. https://doi.org/10.35184/kshce.2018.22.2.29

Lee, S., & Lee, K (2015). A comparative study on gifted programs abroad for economically disadvantaged or minority students: The cases of US, UK, Germany, Australia, and Israel. *Journal of Gifted/Talented Education. 25*(3), 439-463. https://doi.org/10.9722/JGTE.2015.25.3.439

Lee, S., & Lee, K (2016). Multicultural gifted education in Korea: Current issues. *The Journal of the Korean Society for the Gifted and Talented, 15*(3), 103-129. https://doi.org/10.17839/jksgt.2016.15.3.103

Li, C. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*, 936-949. https://doi.org/10.3758/s13428-015-0619-7

Lohman, D. F. (2005). *Identifying academically talented minority students*. National Research Center on the Gifted and Talented. https://files.eric.ed.gov/fulltext/ED505376.pdf

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

Lohman, D. F., & Renzulli, J. (2007). *A simple procedure for combining ability test scores, achievement test scores, and teacher ratings to identify academically talented children.* Unpublished manuscript.

Maggi, S., Hertzman, C., Kohen, D., & D'angiulli, A. (2004). Effects of neighborhood socioeconomic characteristics and class composition on highly competent children. *The Journal of Educational Research, 98*(2), 109-114. https://doi.org/10.3200/JOER.98.2.109-114

Marshall, B., Cardon, P., Poddar, A., & Fontenot, R. (2013). Does sample size matter in qualitative research?: A review of qualitative interviews in IS research. *Journal of computer information systems, 54*(1), 11-22. https://doi.org/10.1080/08874417.2013.11645667

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

McBee, M. T. (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. http://dx.doi.org/10.1177/0016986210377927

McBee, M. T. (2010). Modeling outcomes with floor or ceiling effects: An introduction to the Tobit model. *Gifted Child Quarterly, 54*(4), 314–320. http://dx.doi.org/10.1177/001698621037909

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 rules. *Gifted Child Quarterly, 58*(1), 69-89. https://doi.org/10.1177/0016986213513794

McCallum, R. S., & Bracken, B. A. (2012). *Universal multidimensional abilities scales*. Pro ED.

McCarney, S. B., & Anderson, P. D. (1998). *The Gifted Evaluation Scale–technical manual* (2nd ed.). Hawthorne Educational Services.

McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument development in the affective domain*. Springer.

McCoach, D. B., & Siegle, D. (2007). What predicts teachers' attitudes toward the gifted? *Gifted Child Quarterly, 51*(3), 246-255. https://doi.org/10.1177/0016986207302719

McKown, C. (2013). Social equity theory and racial-ethnic achievement gaps. *Child Development, 84*(4), 1120-1136. https://doi.org/10.1111/cdev.12033

McNeish, D., An, J., & Hancock, G. R. (2018). The thorny relation between measurement quality and fit index cutoffs in latent variable models. *Journal of Personality Assessment, 100*(1), 43-52. https://doi.org/10.1080/00223891.2017.1281286

Miller, R., & Gentry, M. (2010). Developing talents among high-potential students from low-income families in an out-of-school enrichment program. *Journal of Advanced Academics, 21*(4), 594-627. https://doi.org/10.1177/1932202X1002100403

Ministry of Education. (2013). *National level achievement test and evaluation plan*. Seoul, Korea: Ministry of Education. https://www.moe.go.kr/boardCnts/view.do?boardID=294&lev=0&statusYN=C&s=moe&m=0204&opType=N&boardSeq=43470

Ministry of Education. (2013). *3rd Master Plan for the Promotions of Gifted and Talented Education (2013-2017)*. http://gifted.kedi.re.kr/images/gifted/policy/report04.pdf

Ministry of Education & Human Resources Development. (2004). *Plans for the Promotions of Gifted and Talented Education to 5% of K-12 Populations*. http://gifted.kedi.re.kr/images/gifted/policy/report02.pdf

Mönks, F. J., Heller, K. A., & Passow, A. H. (2000). The study of giftedness: Reflections on where we are and where we are going. In K. A. Heller, F. J. Mönks, R. J. Sternberg, & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (2nd ed., pp. 839-864). Elsevier.

Moon, T. R., & Brighton, C. M. (2008). Primary teachers' conceptions of giftedness. *Journal for the Education of the Gifted, 31*(4), 447-480. https://doi.org/10.4219/jeg-2008-793

Morgan, H. (2019). The lack of minority students in gifted education: Hiring more exemplary teachers of color can alleviate the problem. *The Clearing House: A Journal of Educational Strategies, Issues and* Ideas, 92(4-5), 156-162. https://doi.org/10.1080/00098655.2019.1645635

Morse, J. M. (2000). Determining sample size. *Qualitative Health Research*, *10*(1), 3-5.

Mortenson, B. C. (2018). The Role of Teacher Implicit Bias in the Racial Achievement Gap (Publication No. 10791341) [Doctoral dissertation, Georgetown University]. ProQuest Dissertations Publishing.

Musca, S. C., Kamiejski, R., Nugier, A., Méot, A., Er-Rafiy, A., & Brauer, M. (2011). Data with hierarchical structure: impact of intraclass correlation and sample size on type-I error. *Frontiers in Psychology*, *2*, 74. https://doi.org/10.3389/fpsyg.2011.00074

Muthén, B. (1997). Latent variable modeling of longitudinal and multilevel data. In Raftery. A. E. (Eds.), *Sociological methodology* (pp. 453–480). Blackwell.

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 for Gifted Children. (2008). *The role of assessments in the identification of gifted students*. http://www.nagc.org/sites/default/files/Position%20Statement/Assessment%20Position%20Statement.pdf

National Association for Gifted Children. (2015). *2014-2015 State of the state in gifted education*. National Association for Gifted Children & The Council of State Directors of Programs for the Gifted.

National Center for Education Statistics. (2020). *Characteristics of public school teachers.* Retrieved from https://nces.ed.gov/programs/coe/indicator_clr.asp#:~:text=In%202017%E2%80%9318%2C%20about%2079,1%20percent%20of%20public%20school

Neumeister, K. L. S., Adams, C. M., Pierce, R. L., Cassady, J. C., & Dixon, F. A. (2007). Fourth-grade teachers' perceptions of giftedness: Implications for identifying and serving diverse gifted students. *Journal for the Education of the Gifted*, *30*(4), 479-499. https://doi.org/10.4219/jeg-2007-503

Nezlek, J. B. (2008). An introduction to multilevel modeling for social and personality psychology. *Social and Personality Psychology Compass, 2*(2), 842-860. https://doi.org/10.1111/j.1751-9004.2007.00059.x

Olszewski-Kubilius, P. (2007). Working with promising learners from poverty: Lessons learned. In J. VanTassel-Baska & T. Stambaugh (Eds.), *Overlooked gems: A national perspective on low-income promising learners* (pp. 43–46). National Association for Gifted Children.

Olszewski-Kubilius, P., & Clarenbach, J. (2012). *Unlocking emergent talent: Supporting high achievement of low-income, high ability students*. National Association for Gifted Children. https://files.eric.ed.gov/fulltext/ED537321.pdf

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

Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning*. Sage.

Ozcan, D., & Kotek, A. (2015). What do the teachers think about gifted students? *Procedia-Social and Behavioral Sciences, 190*, 569-573. https://doi.org/10.1016/j.sbspro.2015.05.044

Patton, M. Q. (2002). *Qualitative research & Evaluation methods*. Sage.

Pereira, N. (2021). Finding Talent Among Elementary English Learners: A Validity Study of the HOPE Teacher Rating Scale. *Gifted Child Quarterly, 65*(2), 153-166. https://doi.org/10.1177/0016986220985942

Pereira, N., & Gentry, M. (2013). A qualitative inquiry into the experiences of high-potential Hispanic English language learners in Midwestern schools. *Journal of Advanced Academics, 24*(3), 164-194. https://doi.org/10.1177/1932202X13494204

Perry, J. L., Nicholls, A. R., Clough, P. J., & Crust, L. (2015). Assessing model fit: Caveats and recommendations for confirmatory factor analysis and exploratory structural equation modeling. *Measurement in physical education and exercise science, 19*(1), 12-21. https://doi.org/10.1080/1091367X.2014.952370

Peters, S. J., & Gentry, M. (2010). Multi-group construct validity evidence of the HOPE Scale: Instrumentation to identify low-income elementary students for gifted programs. *Gifted Child Quarterly, 54*(4), 298-313. https://doi.org/10.1177/0016986210378332

Peters, S., & Gentry, M. (2012). Group-specific norms and teacher-rating scales: Implications for underrepresentation. *Journal of Advanced Academics, 23*(2), 125-144. https://doi.org/10.1177/1932202X12438717

Peters, S. J., & Gentry, M. (2013). Additional validity evidence and across-group equivalency of the HOPE Teacher Rating Scale. *Gifted Child Quarterly, 57*(2), 85-100. https://doi.org/10.1177/0016986212469253

Peters, S., Matthews, M., McBee, M., & McCoach, D. B. (2014). *Beyond gifted education: Designing and implementing advanced academic programs.* Prufrock Press.

Peters, S., & McBee, M. (2019, April). *The application of differential normative criteria to the gifted education screening phase: Implications for demographic representation*. Paper presented at the American Education Research Association annual meeting, Toronto, Canada. Retrieved from https://osf.io/v5sdg/.

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

Peterson, J. S. (1999). Gifted—through whose cultural lens? An application of postpositivistic mode of inquiry. *Journal for the Education of the Gifted, 22*(4), 354-383. https://doi.org/10.1177/016235329902200403

Pfeiffer, S. I. (2002). Identifying gifted and talented students: Recurring issues and promising solutions. *Journal of Applied School Psychology*, *19*(1), 31-50. https://doi.org/10.1300/J008v19n01_03

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

Pfeiffer, S. I., & Blei, S. (2008). Gifted identification beyond the IQ test: Rating scales and other assessment procedures. In S. I. Pfeiffer (Ed.), *Handbook of giftedness in children: Psychoeducational theory, research, and best practices* (pp. 177-198). Springer.

Pfeiffer, S. I., & Jarosewich, T. (2003). *Gifted Rating Scales* (3rd ed.). Pearson.

Pfeiffer, S. I., & Jarosewich, T. (2003). *The Gifted Rating Scales*. Psychological Corporation.

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

Plucker, J. A., Burroughs, N., & Song, R. (2010). *Mind the (other) gap! The growing excellence gap in K-12 education*. Center for Evaluation and Education Research.

Plucker, J., Hardesty, J., & Burroughs, N. (2013). *Talent on the sidelines: Excellence gaps and America's persistent talent underclass.* Retrieved from https://www.nagc.org/sites/default/files/key%20reports/Talent%20on%20the%20Sidelines%20(2013).pdf

Plucker, J. A., Makel, M. C., Matthews, M. S., Peters, S. J., & Rambo-Hernandez, K. E. (2017). Blazing new trails: Strengthening policy research in gifted education. *Gifted Child Quarterly, 61*(3), 210-281. https://doi.org/10.1177/0016986217701838

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

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental review, 41*, 71-90. https://doi.org/10.1016/j.dr.2016.06.004

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2017). *HLM 7.03 for Windows* [Computer software]. Scientific Software International, Inc. https://ssicentral.com/

Raykov, T. (2011). Intraclass correlation coefficients in hierarchical designs: Evaluation using latent variable modeling. *Structural Equation Modeling, 18*(1), 73-90. https://doi.org/10.1080/10705511.2011.534319

Raykov, T., & Marcoulides, G. A. (2000). A method for comparing completely standardized solutions in multiple groups. *Structural equation modeling, 7*(2), 292-308. https://doi.org/10.1207/S15328007SEM0702_9

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.

Reynolds, C. R., & Carson, A. D. (2005). Methods for assessing cultural bias in tests. In C. L. Frisby & C. R. Reynolds (Eds.), *Comprehensive handbook of multicultural school psychology* (pp. 795-823). John Wiley.

Ridgley, L. M., DaVia Rubenstein, L., & Finch, W. H. (2020). Issues and opportunities when using rating scales to identify creatively gifted students: Applying an IRT approach. *Gifted and Talented International*, 1-13. https://doi.org/10.1080/15332276.2020.1722041

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

Robinson, C. L. M. (2014). North Carolina High School African American males perceptions of the external and internal factors contributing to academic achievement (Publication No. 3581426) [Doctoral dissertation, Fayetteville State University]. ProQuest Dissertations Publishing.

Rothenbusch, S., Zettler, I., Voss, T., Lösch, T., & Trautwein, U. (2016). Exploring reference group effects on teachers' nominations of gifted students. *Journal of Educational Psychology, 108*(6), 883-897. http://dx.doi.org/10.1037/edu0000085

Ryser, G. R., & McConnell, K. (2004). *Scales for identifying gifted students: Ages 5–18*. Prufrock Press.

Ryu, E. (2014). Factorial invariance in multilevel confirmatory factor analysis. *British Journal of Mathematical and Statistical Psychology, 67*(1), 172-194. https://doi.org/10.1111/bmsp.12014

Scherbaum, C. A., & Ferreter, J. M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods, 12*(2), 347-367. https://doi.org/10.1177/1094428107308906

Siegle, D., Moore, M., Mann, R. L., & Wilson, H. E. (2010). Factors that influence in-service and preservice teachers' nominations of students for gifted and talented programs. *Journal for the Education of the Gifted, 33*(3), 337-360. https://doi.org/10.1177/016235321003300303

Speirs Neumeister, K. L., Adams, C. M., Pierce, R. L., Cassady, J. C., & Dixon, F. A. (2007). Fourth-grade teachers' perceptions of giftedness: Implications for identifying and serving diverse gifted students. *Journal for the Education of the Gifted, 30*(4), 479-499. https://doi.org/10.4219/jeg-2007-503

Sprague, B. N., Hyun, J., & Molenaar, P. C. (2017). Revisiting measurement invariance in intelligence testing in aging research: Evidence for almost complete metric invariance across age groups. *Journal for Person-oriented Research, 3*(2), 86-100. https://doi.org/10.17505/jpor.2017.08

Stambaugh, T. (2007). Next steps: An impetus for future directions in research, policy, and practice for low-income promising learners. In J. VanTassel-Baska & T. Stambaugh Eds.), *Overlooked gems: A national perspective on low-income promising learners* (pp. 83-88). National Association for Gifted Children.

Sternberg, R. J. (2015). Successful intelligence: A model for testing intelligence beyond IQ tests. European Journal of Education and Psychology, 8(2), 76-84. https://doi.org/10.1016/j.ejeps.2015.09.004

Sternberg, R. J. (2018). 21 Ideas: A 42-Year Search to Understand the Nature of Giftedness. *Roeper Review, 40*(1), 7-20. https://doi.org/10.1080/02783193.2018.1393609

Subramanian, G. H., & Peslak, A. R. (2010). User perception differences in enterprise resource planning implementations. *Journal of Computer Information Systems, 50*(3), 130-138. https://doi.org/10.1080/08874417.2010.11645415

Sung, E. (2017). Identification and education for the multicultural gifted and talented students. *The Journal of the Korean Society for the Gifted and Talented. 16*(2), 5-21. https://doi.org/10.17839/jksgt.2017.16.2.5

Swanson, J. D. (2006). Breaking through assumptions about low-income, minority gifted students. *Gifted Child Quarterly, 50*(1), 11-25. https://doi.org/10.1177/001698620605000103

Szymanski, T., & Shaff, T. (2013). Teacher perspectives regarding gifted diverse students. *Gifted Children, 6*(1), 1-27. https://docs.lib.purdue.edu/giftedchildren/vol6/iss1/1

The Government of Korea. (2019). *The support of low-income family*. https://www.gov.kr/portal/service/serviceInfo/000000013450

Thompson, M. S., & Green, S. B. (2006). *Evaluating between group differences in latent variable means. Structural equation modeling: A second course* (pp. 119-169). Information Age.

Ullman, J. B. (2007). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (5th ed., pp. 676–780). Allyn & Bacon.

Valencia, R. R., & Suzuki, L. A. (2001). *Intelligence testing and minority students*. Sage.

van de Vijver, F., & Tanzer, N. K. (1997). Bias and equivalence in cross-cultural assessment. *European review of applied psychology, 47*(4), 263-279.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods, 3*(1), 4-70. https://doi.org/10.1177/109442810031002

Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B., & Trautwein, U. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology, 108*(5), 705-721. https://doi.org/10.1037/edu0000075

Webb, M., & Thomas, R. (2015). Teachers' Perceptions of Educators' and Students' Role in Closing the Achievement Gap. *National Forum of Teacher Education Journal. 25*(3). 1-8.

Williams, L. J., & Holahan, P. J. (1994). Parsimony-based fit indices for multiple-indicator models: Do they work?. *Structural Equation Modeling: A Multidisciplinary Journal*, 1(2), 161-189. https://doi.org/10.1080/10705519409539970

World Population Review (2019). *South Korea Population 2019*. Retrieved from http://worldpopulationreview.com/countries/south-korea-population/

Worrell, F. C. (2007). Identifying and including low-income learners in programs for the gifted and talented: Multiple complexities. In J. VanTassel-Baska & T. Stambaugh (Eds.), *Overlooked gems: A national perspective on low-income promising learners* (pp. 47–51). National Association for Gifted Children.

Wyner, J.S., Bridgeland, J.M., & DiIulio, Jr., J.J. (2009). *Achievementrap: How America is failing millions of high-achieving students from lower-income families*. A Report of the Jack Kent Cooke Foundation and Civic Enterprises.

Yoon, M., & Lai, M. H. (2018). Testing factorial invariance with unbalanced samples. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(2), 201-213. https://doi.org/10.1080/10705511.2017.1387859

Yoon, S. Y., & 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

Yuan, K. H., & Bentler, P. M. (2000). Robust mean and covariance structure analysis through iteratively reweighted least squares. *Psychometrika, 65*(1), 43-58.

Zeidner, M., & Shani-Zinovich, I. (2011). Do academically gifted and nongifted students differ on the big-five and adaptive status? Some recent data and conclusions. *Personality and Individual Differences, 51*(5), 566-570. https://doi.org/10.1016/j.paid.2011.05.007

Ziegler, A., & Raul, T. (2000). Myth and reality: A review of empirical studies on giftedness. *High Ability Studies, 11*(2), 113-136. https://doi.org/10.1080/13598130020001188

APPENDIX A. *HOPE* TEACHER RATING SCALE

HOPE TEACHER RATING SCALE

TEACHER'S NAME/CODE: _____

Marcia Gentry, Ph.D. • Nielsen Pereira, Ph.D. • Scott J. Peters, Ph.D. • Jason S. McIntosh • C. Matthew Fugate, Ph.D.

Developed with funding from the Jack Kent Cooke Foundation 2007

Student Name/ID#:			Grade:
Date of Birth:	Age:	Sex: ☐ Male ☐ Female	☐ Free/Reduced Lunch

☐ American Indian/Alaska Native ☐ Asian ☐ Black or African American ☐ White
☐ Native Hawaiian or Other Pacific Islander ☐ Mixed Race ☐ Hispanic/Latino

When rating students on each item below please think about the student *compared to other children similar in age, experience, and/or environment.* Use the following scale to indicate how frequently you observe the traits and behaviors listed in items 1–11.

6 = Always 5 = Almost Always 4 = Often 3 = Sometimes 2 = Rarely 1 = Never

The student demonstrates . . .	1	2	3	4	5	6
1. Performs or shows potential for performing at remarkably high levels.						
2. Is sensitive to larger or deeper issues of human concern.						
3. Is self-aware.						
4. Shows compassion for others.						
5. Is a leader within his/her group of peers.						
6. Is eager to explore new concepts.						
7. Exhibits intellectual intensity.						
8. Effectively interacts with adults or older students.						
9. Uses alternative processes.						
10. Thinks "outside the box."						
11. Has intense interests.						

12. Please indicate all content areas where the student shows talent.

☐ Math ☐ Reading ☐ Creative Writing ☐ Social Studies
☐ Science ☐ Foreign Language ☐ Arts ☐ Other: _____

Please provide additional information concerning this child's potential:

_____

_____

_____

Additional copies of this form may be purchased from Prufrock Press Inc., P.O. Box 8813, Waco, TX 76714-8813; Phone 800/998-2208; Fax 800/240-0333; http://www.prufrock.com; ProdCode: 4546

APPENDIX B. KOREAN VERSION - *HOPE* TEACHER RATING SCALE

교사이름/코드(학교이름):_____

HOPE[1]교사용
학생 추천 체크리스트

학생이름/ID (번호): _____ 학년:_____ 날짜:_____

학생생년월일:_____ 성별: ☐ 남자, ☐ 여자 소득: ☐ 일반, ☐ 차상위/저소득계층

인종(민족): ☐ 한국인, ☐ 다문화 (☐ 유럽/미국계, ☐ 아프리카, ☐ 히스패닉(라틴),

☐ 아시아 - ☐ 중앙아시아 ☐ 동북아시아 ☐ 동남아시아 ☐ 남아시아 ☐ 서아시아, ☐기타_____)

다음 체크리스트를 작성하실 때에, 해당 학생을 *비슷한 나이또래, 경험,환경을 가진 아이들과 비교하여* 평가해 주세요.학생이 얼마나 자주 1-11 문항의 특성과 행동을 보이는지 체크해 주시기 바랍니다.

1 =전혀, 2=가끔, 3 = 보통, 4 = 자주, 5 = 거의 항상, 6 = 항상

	1	2	3	4	5	6
1. 관심분야에 높은 수준의 능력을 보이거나 *잠재력*을 가지고 있다.						
2. 인류/사회 문제에 민감하고 관심이 많다.						
3. 자기 자신을 잘 알고 있다.						
4. 타인에 대한 연민/동정심을 가지고 있다.						
5. 또래 그룹에서 리더 역할을 한다.						
6. 새로운 개념에 대한 탐구심이 강하다.						
7. 강한 지적 욕구를 보인다.						
8. 어른들이나 자신보다 나이가 많은 학생들과 활발히 어울린다.						
9. 대안을 활용할 수 있다.						
10. 고정관념의 틀에서 벗어나 생각한다.						
11. 강한 관심분야가 있다.						

12. 학생이 재능을 보이는 분야에 모두 체크해주세요.

☐수학 ☐읽기/독서 ☐작문 ☐사회

☐과학 ☐외국어 ☐예술 ☐기타 _____

해당 학생의 잠재력에 관한 추가적인 정보를 제공해 주세요:

_____
_____
_____
_____
_____

[1]2007Jack Kent Cooke Foundation Funding으로 개발됨
2015년 개정

PUBLICATIONS

Chapter 2 was submitted to *Gifted Child Quarterly* and Chapter 4 was submitted to *Journal of Advanced Academics* when the dissertation was defensed on March, 30, 2021.