

8-8-2017

Novel Methods to Address Measurement Error Issues in Gifted Identification

Huihui Yu

University of Connecticut - Storrs, huihui.yu.wang@gmail.com

Follow this and additional works at: <https://opencommons.uconn.edu/dissertations>

Recommended Citation

Yu, Huihui, "Novel Methods to Address Measurement Error Issues in Gifted Identification" (2017). *Doctoral Dissertations*. 1524.
<https://opencommons.uconn.edu/dissertations/1524>

Novel Methods to Address Measurement Error Issues in Gifted Identification

Huihui Yu, PhD
University of Connecticut, 2017

This study applied statistical simulation techniques to posit a practical situation of gifted identification based on students' performance on the intelligence test and academic achievement tests of math and reading. Three tests were generated based on the one-parameter item response theory model. The marginal reliabilities were .95, .90, and .80 for observed intelligence and observed abilities in math and reading, respectively. Totally, 1,000,000 pairs of true and observed abilities in each of the three domain areas were generated. Results suggest that using different combination rules is conceptually aimed to identify different gifted populations. Using the conjunctive rule to combine a high standard for all three measures is aimed to identify the gifted population who are gifted in all three domain areas; however, using the complementary rule is aimed to identify the gifted population who are gifted in at least one domain area. Therefore, the consequences on the identified gifted group and the performance of gifted identification using different combination rules should not be compared directly. Given the differences in the reliabilities and correlations, results suggest that the conjunctive and compensatory rules identify more students who are gifted in the domain area measured with higher reliabilities and more correlated with the other domain areas; and, the complementary rule favors the students who are gifted in the domain area measured with higher reliabilities. Using any test as the conclusive test favors the students who are gifted in the domain area more correlated with the domain area measured by the conclusive test. In general, gifted identification using different combination rules may perform relatively well in terms of the positive predictive rate (PPR) or sensitivity but rarely both. This study explored new methods of gifted

identification under the Bayesian framework to systematically address the measurement error issues. Results suggest that using the posterior probability of being gifted given multiple observed abilities to identify gifted students has the potential to simultaneously improve the PPR and sensitivity of gifted identification.

Novel Methods to Address Measurement Error Issues in Gifted Identification

Huihui Yu

B.E., Harbin University of Engineering, 1998

M.E., Harbin University of Engineering, 2001

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2017

Copyright by

Huihui Yu

2017

APPROVAL PAGE

Doctor of Philosophy Dissertation

Novel Methods to Address Measurement Error Issues in Gifted Identification

Presented by

Huihui Yu, B.E., M.E.

Major Advisor _____
D. Betsy McCoach

Associate Advisor _____
Hariharan Swaminathan

Associate Advisor _____
H. Jane Rogers

Associate Advisor _____
Eric Loken

University of Connecticut

2017

Acknowledges

First, I would like to thank my advisor, Betsy McCoach, for her inspiration, encouragement, understanding, and tremendous amounts of time, patience, and persistence through this process. I would not have accomplished this study and my Ph.D. journey without her.

I would also like to thank my committee members, Hariharan Swaminathan, Jane Rogers, and Eric Loken. I really appreciate Swami always having an open door policy with me and always having an answer to my question. Jane helps me to build up all my knowledge, skills, and confidence in statistical programing. I am very grateful that Eric joins the MEA program before I leave. He generously provides so many inspiring and helpful suggestions to me on this study.

Finally, I would like to thank my husband, Zeqi, for investing in me and supporting me to pursue my academic goals. I would also like to thank my beloved sons, Ben and Braydon, for their love, understanding, and support.

Contents

Chapter 1	1
Statement of the Problem	1
<i>What Measures to Use</i>	<i>1</i>
<i>How to Use Multiple Measures in Gifted Identification</i>	<i>2</i>
<i>Evaluating the Performance of Gifted Identification.....</i>	<i>4</i>
<i>The Current Study.....</i>	<i>7</i>
Chapter 2	9
Review of the Literature.....	9
<i>Definitions of Gifted Children</i>	<i>9</i>
<i>Gifted Identification in the Public School System.....</i>	<i>11</i>
<i>Influence of Measurement Error on the Identification of Gifted</i>	<i>13</i>
<i>Suggestions on Using Multiple Measures in Making Educational Decisions</i>	<i>15</i>
<i>Combination Rules.....</i>	<i>18</i>
<i>Performance of Gifted Identification Based on Multiple Measures</i>	<i>24</i>
Chapter 3	28
Methods.....	28
<i>Notations</i>	<i>28</i>
<i>Data Generation</i>	<i>30</i>
<i>Defining Groups with Different Gifted Characteristics.....</i>	<i>41</i>

<i>Plan of Data Analysis for Research Question One</i>	44
<i>Plan of Data Analysis for Research Question Two</i>	48
<i>Plan of Data Analysis for Research Question Three</i>	50
Chapter 4	53
Results	53
<i>Results for Research Question One</i>	53
<i>Results for Research Question Two</i>	59
<i>Results for Research Question Three</i>	63
Chapter 5	77
Discussion	77
<i>Defining the Target Population of Gifted Identification</i>	77
<i>Two Qualities of “Good” Measures</i>	78
<i>Understanding the Identified Gifted Group</i>	81
<i>Better Performance of Gifted Identification by Using Bayesian Methods</i>	82
<i>Implications for Practice</i>	84
<i>Limitations and Future Study</i>	85
References	87
Appendix A	92
<i>The Syntax of Generating Observed abilities</i>	92

Chapter 1

Statement of the Problem

In the field of gifted education, how to identify gifted and talented students has been one of the most widely discussed and debated topics (Renzulli, 2004). Answers to this question depend on how one conceptualizes who are gifted and talented students. However, the definition of gifted students has been changing with the evolution of theories of intelligence and has never been uniform (Hunt, 2010; National Association for Gifted Children [NAGC], 2015; Renzulli, 2004; U.S. Department of Education [USDE], 1993; Sternberg & Davidson, 2005). With the advent of new theories of intelligence (e.g., Gardner's Multiple Intelligences Theory, 1983; Renzulli's Three-Ring Concept of Giftedness, 1973; Sternberg's Triarchic Theory of Intelligence, 1985) and the increase in the knowledge of gifted students, it has been widely accepted that giftedness is not a unidimensional construct and cannot be fully addressed by a single measure (Friedman-Nimz, 2009; Worrell, 2009). Therefore, it is broadly recommended to use multiple measures in the practice of gifted identification (NAGC, 2015). However, the instruction for the use of multiple measures in gifted identification is lacking. The instruction should include what measures can be used in gifted identification and how the multiple measures should be used in gifted identification. To develop the instruction of using multiple measures in gifted identification calls for data-driven studies (Johnsen, 2004; McCoach, Kehle, Bray, & Siegle, 2001; NAGC, 2015; Renzulli, 2004).

What Measures to Use

The measures used for gifted identification are high-stakes tests in that they determine whether students are qualified for gifted services. Therefore, gifted identification should use quality measures. Reliability is a broadly accepted statistical parameter indicating the quality of

measurement. The present study suggests that observed scores from the measures used for gifted identification should be adequately reliable. According to the classic test theory (CTT), the discrepancy between true scores and observed scores is measurement error (Gulliksen, 1950). Reliability is the ratio of the true variance to the total variance composed of the true variance and the non-systematic measurement error variance (Silva, 1993), which reflects the degree of consistency between true abilities and observed abilities (Lord & Novick, 1968). Statistically, the higher the reliability is the more predictive observed abilities are of students' true abilities (Kelley, 1947). Observed scores with low reliabilities are not a good indicator of students' true abilities and therefore should not be used in gifted identification, especially during the decision-making process. Given that, the present study only investigates the usage of observed measures with adequate reliabilities ($p^2 \geq .8$). Further, for gifted identification based on multiple measures, if the multiple measures measure different abilities, the present study suggests that the abilities should be moderately strongly correlated with each other ($r \geq .7$), and therefore it can be expected that 50% or more of variance is common across different abilities underlying the multiple measures.

How to Use Multiple Measures in Gifted Identification

The methods of gifted identification based on multiple measures can be specified by two factors: (1) the criteria of gifted for different measures; and (2) the rules used to combine the criteria of gifted for different measures. A scaled score of two standard deviations above the population mean (e.g., an IQ score of 130) is commonly used as the criterion of gifted, which is corresponding to the 97.5th percentile. The NAGC (2015) suggests that 10% of children are gifted in at least one domain area, which is corresponding to the 90th percentile or a scaled score of 1.28 standard deviations above the mean. Concerning the combination rules, the conjunctive

rule, the compensatory rule, and the complementary rule are commonly used in making decisions in the field of education (Chester, 2003; Douglas, 2007; Douglas & Mislevy, 2010). The combination rules can be used to combine an identical criterion of gifted for repeated measures or to combine the same or different criteria of gifted for different measures. When different combination rules and different criteria of gifted are applied, the group of students identified as gifted may differ in the size and gifted characteristics (e.g., intellectually and mathematically gifted, or intellectually or mathematically gifted). Also, the performance of gifted identification may vary in terms of the precision of identification results and the sensitivity to the target gifted group (McBee, Peters, & Waterman, 2014).

Further, the combination rules can be combined with each other. For example, in Georgia, students who perform at or above the 90th percentile on the norm-referenced math or reading achievement tests are selected as potential candidates for the gifted program; and the potential candidates need to perform at or above the 96th percentile on a standardized mental ability test to be qualified for the gifted program. In this case, the complementary rule is used during the screening process and the conjunctive rule is used during the decision-making process. The complementary rules, also known as the “Or” rule, requires the attainment of the minimum standard on any one of the multiple measures. The conjunctive rule, also known as the “And” rule, requires the fulfillment of the minimum standards on all multiple measures. Using the conjunctive rule to combine the screening results and the performance on the mental ability test is common among the four states (e.g., Florida, Oklahoma, Georgia, Iowa), where gifted education is mandatory and fully funded, to make the final decision on who are qualified for the gifted program. (Florida State Department of Education, 2013; Oklahoma State Department of Education, 2016; Georgia State Department of Education, 2016; Iowa State Department of

Education, 2011). The conjunctive rule makes students' performance on the standardized mental ability test as a necessary condition for entering gifted program, and therefore students who pass the screening process but fail to achieve the high standard on the mental test will not be eligible for gifted programs. This phenomenon is called the "multi-criteria smoke" by Renzulli (2004), who thought that using the mental ability test during the decision-making process conflicted with the essential purpose of introducing multiple measures in gifted identification. However, what is the real source of the "multi-criteria smoke", the mental ability test or the way to use mental ability tests in gifted identification? This study thoroughly investigated the consequences of using different tests as the conclusive test on the identified gifted group to determine whether gifted identification should use a conclusive test during the decision-making process and further explored the factors that may introduce unfairness into gifted identification.

Evaluating the Performance of Gifted Identification

The performance of gifted identification can be evaluated in terms of the reliability, precision, sensitivity, and validity of the results of identification (Douglas & Mislevy, 2010). The present study is concerned primarily with precision and sensitivity of the identification results. To calculate the precision and sensitivity, this study defines six ability groups based on students' true and observed scores. If a student's true scores fulfill the criteria of gifted, the student is defined as "truly"¹ gifted (TG); if a student's observed scores fulfill the criteria of gifted, the student is defined as identified gifted (IG). Further, if a student's true scores and observed scores

¹ Given the present study is a simulation study, students' true scores are known. The term "truly" gifted (TG) in the present study simply refers to the group of students whose true scores fulfill the criteria of gifted. The main purpose of using TG is to distinguish gifted students defined based on true scores from students identified based on observed scores.

both satisfy the criteria of gifted, the student is defined as true positive gifted (TPG); if the student's true scores fulfill the criteria of gifted but the student's observed scores do not fulfill the criteria of gifted, the student is defined as false negative gifted (FNG); if neither of a student's true scores nor the student's observed scores satisfy the criteria of gifted, the student is defined as true negative gifted (TNG); and, if a student's true scores do not satisfy the criteria of gifted but the student's observed scores fulfill the criteria of gifted, the student is defined as false positive gifted (FPG).

TG is the sum of TPG and FNG; non-TG is the sum of TNG and FPG; and, IG is the sum of TPG and FPG.

$$TG = TPG + FNG$$

$$non-TG = TNG + FPG$$

$$IG = TPG + FPG$$

The precision is the ratio of TPG to IG. The precision, also known as the positive predictive rate (PPR), is the probability that a student is TG given the student is IG. Also, the precision indicates the likelihood of making the same identification decision based on true scores and observed scores.

$$PPR = \frac{TPG}{IG} = \frac{TPG}{TPG + FPG}$$

The sensitivity to giftedness is the ratio of TPG to TG. The sensitivity, also known as the true positive rate (TPR), is the probability that a student can be identified as gifted given the student is TG.

$$Sensitivity = TPR = \frac{TPG}{TG} = \frac{TPG}{TPG + FNG}$$

Another widely applied statistical term is the specificity or the true negative rate (TNR), which is not examined in the present study. The TNR is the ratio of TNG to non-TG.

$$TNR = \frac{TNG}{TNG + FPG}$$

Given the small proportion (e.g., $\leq 10\%$) of gifted within the entire population, the variation in the TNR^2 due to measurement error will be much smaller than the variation in the PPR^3 and TPR^4 . The TNR does not provide additional information to better evaluate the performance of gifted identification beyond the PPR and TPR . Therefore, the present study only uses the PPR and sensitivity to evaluate the performance of gifted identification. High $PPRs$ and sensitivities are both indicators of quality performance of gifted identification.

Further, the present study estimated the posterior probability of being TG given observed ability. To be specific, let x refers to a observed ability, T refers to the vector of possible true abilities of x , and t_1, \dots, t_k refers to true abilities no less than the criterion of gifted given the observed ability x , then the posterior probability of true abilities (T) fulfilling the criterion of gifted (c) given a specific observed ability (x) is

$$P(T \geq c|x) = \frac{\sum_{i=1}^k P(x, t_i)}{P(x)}$$

² Assuming that the proportions of TG and IG are both equal to 10%, then FPG is no greater than .1 and TNG is no less than .8. Therefore, the maximum TNR is 1.0 when FPG is equal to 0.0; the minimum TNR is $\frac{0.9-0.1}{(0.9-0.1)+0.1}=.89$ when FPG is equal to 0.1.

³ Assuming the proportion of IG is equal to 10%, then both TPG and FPG are no greater than .1. Therefore, the maximum PPR is $\frac{0.1-0.0}{(0.1-0.0)+0.1} = 1.0$ when FPG is equal to 0.0; the minimum PPR is $\frac{0.1-0.1}{(0.1-0.1)+0.1}=.0$ when FPG is equal to 0.1.

⁴ Assuming the proportion of TG is equal to 10%, then both TPG and FNG are no greater than .1. Therefore, the maximum TPR is $\frac{0.1-0.0}{(0.1-0.0)+0.1} = 1.0$ when FNG is equal to 0.0; the minimum TPR is $\frac{0.1-0.1}{(0.1-0.1)+0.1}=.0$ when FPG is equal to 0.1.

Theoretically, the probability of a continuous variable equal to a specific value is zero. In this study, observed abilities and true abilities are both continuous variables. To calculate the posterior probabilities, both observed and true abilities were rounded to two decimal places. Therefore, in the posterior probability equation, $P(x)$ refers to the probability of observed abilities ranging from $(x - 0.005)$ to $(x + 0.005)$ and $P(x, t_i)$ refers to the probability of the small square of $(x \pm .005)$ and $(t_i \pm .005)$. I shall return to this point in more detail.

The Current Study

This study applied statistical simulation techniques to posit a practical situation of gifted identification based on observed intelligence and observed abilities in two academic fields (math and reading). Given that both true and observed abilities are known in this simulation study, I can thoroughly investigate the performance of gifted identification using different rules to combine multiple measures. Findings about the consequence and performance of gifted identification using different methods will provide valuable information to instruct the use of multiple measures in the practice of gifted identification.

Further, this study is aimed to discover a novel method of gifted identification under the Bayesian framework based on the posterior probability of being TG given observed abilities. Given that true abilities are not directly observed and observed abilities are contaminated with measurement error in reality, the posterior probability of being TG given observed abilities indicates that the probability of being TG given different observed abilities should be continuous, ranging between 0 and 1, instead being dichotomous, only taking values of 0 and 1. This study is engaged in exploring how to use the posterior probabilities of being TG given observed abilities to improve the performance of gifted identification.

To be specific, this study is aimed to address three research questions.

1. Among the methods of gifted identification using different rules to combine multiple measures, how differently do those methods perform in terms of the PPR and sensitivity?
2. Considering gifted identification based on multiple measures, what are the consequences of using different combination rules on the size and gifted characteristics of the identified gifted (IG) group?
3. Can the posterior probabilities of being TG given observed abilities be used to effectively improve the PPR and sensitivities of gifted identification?

Chapter 2

Review of the Literature

Definitions of Gifted Children

Over a century ago, with the emergency and success of norm-referenced intelligence tests, gifted children specifically referred to very intelligent children (Hunt, 2010; Johnsen, 2004; Renzulli, 2004). The traditional method of gifted identification merely relied on observed scores on intelligence tests. Consequently, people were identified as gifted if they achieved super high scores (e.g. 130 scale scores on an individually administered standardized test of intelligence with the mean of 100 and the standard deviation of 15) on a high profile standardized intelligence test (e.g., the *Wechsler Intelligence Scale for Children, WISC-V*, 2003; the *Stanford-Binet Intelligence Scale, SB*; Roid, 2003), though the traditional method of gifted identification has gradually fallen out of favor (Hunt, 2010; Johnsen, 2004; Renzulli, 2004).

About a half century ago, Sidney Marland officially specified the diversity in giftedness in his August 1971 report to Congress, which was revised in 1978 and 1993 (Johnsen, 2004). In the U.S. Department of Education (USDE) report of 1993, gifted children are defined as:

Children and youth with outstanding talent perform or show the potential for performing at remarkably high levels of accomplishment when compared with others of their age, experience, or environment. These children and youth exhibit high capability in intellectual, creative, and/or artistic areas, possess an unusual leadership capacity or excel in specific academic fields. They require services or activities not ordinarily provided by the schools. Outstanding talents are present in children and youth from all cultural groups, across all economic strata, and in all areas of human endeavor. (p. 26)

With the advent of the No Child Left Behind Act (NCLB) in 2001, the federal definition of giftedness was modified as:

Students, children, or youth who give evidence of high achievement capability in areas such as intellectual, creative, artistic, or leadership capacity, or in specific academic fields, and who need services and activities not ordinarily provided by the school in order to fully develop those capabilities. (P.L. 107-110, Title IX, Part A, SEC. 9101 (22), 2002)

Further, the NAGC provided more details of the gifted and talented students in the position statement of 2010:

Gifted individuals are those who demonstrate outstanding levels of aptitude (defined as an exceptional ability to reason and learn) or competence (documented performance or achievement in top 10% or rarer) in one or more domains. Domains include any structured area of activity with its own symbol system (e.g., mathematics, music, language) and/or set of sensorimotor skills (e.g., painting, dance, and sports). (p. 1)

Although there is a federal definition of giftedness, states have the authority to determine their own definitions of gifted and talented students. In the school year of 2014-2015, the NAGC conducted the State of the States in Gifted Education survey to investigate the policies and practices of gifted education across the country. Among the 39 responding states, 37 states provided definition of gifted children in statute (13), regulations (23), or other resources (1). For example, Florida defined “gifted” in state statute as “One who has superior intellectual development and is capable of high performance” (6A-6.03019, 2002). Colorado defined “gifted children” in state rules (1 CCR 301-8, [2220-R-12.00, 12.01(16)], 2015) as:

Those persons between the ages of four and twenty-one whose aptitude or competence in abilities, talents, and potential for accomplishment in one or more domains are so

exceptional or developmentally advanced that they require special provisions to meet their educational programming needs. (p.100)

Among the five domain areas of giftedness (e.g., intelligence, academic achievement, creativity, performing/visual arts, and leadership) that are commonly included in the state definitions of “Gifted” or “Gifted Children”, intellectual giftedness (34 out of 37 states, 81%) and academic giftedness (24 out of 37 states, 57%) are the top two forms of giftedness addressed in the state definitions of gifted children. In addition, among the 37 states that provided the definition of gifted children, 20 states emphasized giftedness in specific academic areas. For example, in Kentucky, there are magnet schools that serve students in grades 4 through 12 who are gifted in science and mathematics.

Gifted Identification in the Public School System

Currently, there is no federal mandate for gifted identification or services. Therefore, decisions on placement in gifted programs are made at the state level or, sometimes, at the district or even school level (Johnsen, 2004). Among the 42 states that responded the NAGC 2014-2015 survey, 32 states reported a mandate related to gifted and talented education, for gifted identification, services, or both. However, 8 out of the 32 states that had mandates did not provide funding for gifted education. Among the 42 responding states, 33 states provided information on the criteria or methods required for the identification of gifted and talented students. Further, among the 33 states that provided information about gifted identification, 19 states required the application of a multiple criteria model. In terms of the required information for gifted identification, scores on an individually administered standardized test of intelligence (IQ scores) and academic achievement data are the two types of information that were required for gifted identification by most of the states (13 out of 19) that required the use of a multiple

criteria model. Consequently, students who are intellectually gifted or academically gifted represent the majority of students who are identified as gifted and eligible for gifted services.

Gifted education in the public school system pays great attention to giftedness in the domain areas of mental abilities and academic achievement, given the goal of gifted education is to maximize students' academic potential. States commonly employ standardized tests of academic achievement and mental abilities in the practice of gifted identification (Birch, 1984; Johnsen, 2004; NAGC, 2015). There is a strong relationship between academic achievement and mental abilities, and therefore most states use quantitative assessments of both mental abilities and academic achievement in the identification of giftedness in either domain area, intelligence or academic. Therefore, the current practice of gifted identification based on both IQ scores and achievement test scores cannot clearly distinguish these two forms of giftedness from each other. Further, there is a difference in the relationship between intelligence and abilities in different academic fields. McCoach, Yu, Gottfried, and Gottfried (2017) applied latent variable modeling techniques to investigate the relationship between intelligence and abilities in math and reading. They found that the correlation between latent early childhood intelligence and elementary math achievement was .85 and the correlation between latent early childhood intelligence and elementary reading achievement was .70. Both correlations were strong but not perfect. Additionally, the correlation between intelligence and the abilities in math was evidently stronger than the correlation between intelligence and the abilities in reading. This finding raises two questions about the use of intelligence test scores in the identification of giftedness in different academic fields: (1) should intelligence test scores be used in the identification of giftedness in all academic fields; and (2) should the same standard for intelligence test scores be used in the identification of giftedness in different academic fields? No study has systematically

investigated how to determine the standards for multiple measures of different abilities in the identification of gifted.

Influence of Measurement Error on the Identification of Gifted

The definition of gifted children is based on children's true abilities, though children's true abilities cannot be observed directly. In practice, gifted identification uses observed abilities. For gifted identification based on observed quantitative assessments, the observed gifted status is traditionally determined by a super high cutoff score (e.g., a cutoff score of 130 IQs corresponding to two standard deviations above the population mean or the 97.5th percentile). Students who perform higher than the cutoff score are identified as gifted. Regardless of the issues of validity, students' observed abilities are not their true abilities. The discrepancy between the students' observed abilities and their true abilities is called measurement error (Lord & Novick, 1968). Evidently, the size of measurement error in observed abilities can greatly affect the performance of gifted identification based on observed abilities.

According to the classic test theory, the observed score is the sum of the true score and a piece of random measurement error (Lord & Novick, 1968). However, neither true scores nor measurement error is observed. By assuming that measurement error is independent of true scores, the variance of observed scores is the sum of the variance of true abilities and the variance of measurement error. The proportion of variance in observed scores explained by true scores is the well-known reliability (ρ_{XT}^2), which represent the consistency of observed scores across occasions. The variance in observed scores unexplained by true scores is measurement error variance. Therefore, the higher the reliability is the smaller proportion of variance in observed scores is due to measurement error. For a given true score, the smaller measurement error is the closer the observed score is to the true score. However, even a very small

measurement error may change the result of gifted identification of students whose true abilities are around the cutoff scores. Further, since the sign of measurement error is unknown, it is unknown that students are identified as gifted correctly, students are identified as gifted due to a positive measurement error, or students are not identified as gifted due to a negative measurement error. Therefore, measurement error makes it uncertain whether students are correctly classified into gifted or non-gifted groups. This concern about measurement error is especially severe for students whose true scores are around the cutoff score.

In the CCT, the variance of measurement error is assumed consistent across the entire range of students' abilities. Given this assumption, Kelley (1947) provided the equation of estimating true scores given observed scores:

$$R(T | X) = \rho_{XT}^2 X + (1 - \rho_{XT}^2) \mu_X = X + (1 - \rho_{XT}^2) (\mu_X - X)$$

Kelley suggested that the estimated true score is the weighted sum of the observed score and the population mean. For an observed score above the population mean, the estimated true score is lower than the observed score and closer toward the population mean. On the contrary, for an observed score below the population mean, the estimated true score is higher than the observed score and also close toward the population mean. For a given reliability, the larger the observed score is the more the estimated true score will be pulled toward the mean. Further, given this formula, the true score given an observed score lower than the criterion of gifted but above the population mean will be surely lower than the criterion of gifted. If this is true, then gifted identification need not to consider students whose observed scores are lower than the criterion of gifted. Kelley also provided the equation of the standard error of estimated true scores as below:

$$\varepsilon = \rho_{XT}^2 (X' - \mu_{X'}) - (T - \mu_{X'})$$

Therefore, given an observed score, the lower the test reliability is the less certain the true score is estimated. Kelley's estimation of true scores given observed scores is only useful if the mean and reliability are both known.

According to item response theory (IRT, Hambleton & Swaminathan, 1991), the reliability and standard error variance are actually conditional on true abilities and observed scores of a specific true abilities do not follow a certain distribution. In this case, the mean and reliability conditional on a true score are both unknown. Therefore, Kelley's method of estimating true scores given observed scores cannot be used. Currently, the one parameter IRT (1P-IRT) model is commonly used in designing the state standardized achievement (e.g., AZ, CT, MN). The observed scores on the test designed based on the 1P-IRT model have bigger standard error for true abilities at the two tails of the distribution (Hambleton & Swaminathan, 1991). In the situation of gifted identification, the observed gifted status is determined by a super high cutoff score that locates at the right tail of the distribution, where the standard error is larger than the average standard error across the entire range of abilities. Further, the conditional standard error increases with the increase in abilities, which means the reliabilities of observed abilities decrease with the increase in abilities. For an extremely high ability (e.g., 4 or 4 standard deviations above the mean), the less reliable observed abilities may not change the observed gifted status. However, for abilities near to the cutoff score, the relatively larger standard error increase the risk for misidentification of gifted (Crocker & Algina, 1986).

Suggestions on Using Multiple Measures in Making Educational Decisions

In the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], 2014), Standard 12.10 stated:

In educational settings, a decision or characterization that will have major impact on a student should take into consideration not just scores from a single test but other relevant information. (p.198)

The committee on Appropriate Test Use, formed by the National Research Council, also emphasized that “An educational decision that will have a major impact on a test taker should not be made solely or automatically on the basis of a single test score” (Heubert & Hauser, 1999, p. 3). Further, the modern testing philosophy proposes that intellectual giftedness should not be identified using only one score from a particular instrument (Johnsen, 2004). Using multiple measures is expected to improve the quality of decisions. However, the improvement does not automatically happen only by including multiple measures (Chester, 2003). Compared with the usage of multiple measures in the context of academic achievement (Ryan, 2002; Chester, 2003; Douglas & Mislevy, 2010), the usage of multiple measures in the context of gifted identification has been rarely explored (McBee, Peters, & Waterman, 2014).

The best practices in gifted identification also underscored the importance of using multiple measures to make better decisions about identifying gifted students. The NAGC suggested using both subjective and objective assessments in gifted identification. Further, the NAGC also suggested that tests used for gifted identification should always be administered by trained professionals. Renzulli (2004) also promoted the central role of professionals in gifted identification especially when both subjective and objective assessments contributed to the final decision on students’ qualification for gifted services or placement in gifted programs. Given the focus of the present study, using quantitative assessments in gifted identification, the NAGC required that tests served as object assessment tools in gifted identification should be align with the characteristics of the target domain area of giftedness. Therefore, if there is no evidence that

abilities in a specific domain area (e.g., sports, arts) correlate with abilities in another domain area (e.g., academic achievement), then the quantitative assessment of abilities in one domain area should not be used or play a determining role during the identification of giftedness in the other domain area. Further, the objective assessments or quantitative measures of students' abilities should provide the relative standing of each student within a specific population (e.g., different age or grade groups). Therefore, it is very common among states that standardized tests (e.g., state mastery tests, intelligence tests) are used in gifted identification.

Theoretically, using multiple measures, assuming that the multiple measures are correlated and credible, should result in better decisions than using single measures because of more evidence. However, Cronbach, Linn, Brennan, and Haertel (1997) called for caution in combining multiple measures based on complex decision rules. After investigating the sources of error in observed abilities and the effects of measurement error on the decisions, they warned people, who believed that combining multiple measures based on complex decision rules would result in better decisions, to be aware that "measurement error is likely to make decisions highly fallible" (p. 381). The main concern is that the compound effect of measurement error in different measures on the final decision is greater than the effect of measurement error in each individual measure under certain combination rules (e.g., the conjunctive rule). For a student whose true scores are above the criterion of gifted on two tests, if the probabilities of the student's observed scores on the two tests passing the criterion of gifted are .8 and .9 due to measurement error and the measurement error in observed scores on the two tests is independent from each other, then the probability of the student passing the criterion of gifted on both tests is the product of .8 and .9, or .72, which is lower than the probability of the student's observed

score on either test passing the criterion of gifted. In this example, using two tests does not increase the precision of the decision.

Combination Rules

Chester (2003) presented a framework to combine multiple measures to make high-stake decisions. To combine multiple measures, Chester suggested three combination rules, the conjunctive rule, the compensatory rule, and the complementary rule. The conjunctive rule, also known as the “And” rule, requires the fulfillment of the minimum standards on each of the multiple measures. The compensatory model, also known as the “Mean” rule, applies the standard of gifted to the mean of observed scores on the multiple measures. Under the compensatory rule, weaker performance on some measures can be counterbalanced by stronger performance on the other measures. The complementary rules, also known as the “Or” rule, requires the attainment of the minimum standard on any one of the multiple measures. Given the use of combination rules in the identification of gifted, the minimum standard of gifted is a super high cutoff score (e.g., a scale score corresponding to the 97.5th percentile or two standard deviations above the mean). Using a super high cutoff score as the standard of gifted, mathematically, the probability of fulfilling the standard across all measures (gifted identification based on the conjunctive rule) should be lower than the lowest probability of fulfilling the standard on any one measure; on the contrary, the probability of fulfilling the standard on at least one of the multiple measures (gifted identification based on the complementary rule) should not be lower than the highest probability of fulfilling the standard on any one measure. Considering gifted identification based on the compensatory rule, the standard of gifted is applied to the mean. It is true that high performance on some measures may compensate low performance on the other measures; however, it is possible that the performance on some measures is too low to

be counterbalanced by the high performance on the other measures. Logistically speaking, if a student is identified as gifted by the conjunctive rule, then the student surely will be identified as gifted by the complementary rule and compensatory rule. If a student is identified as gifted by either conjunctive rule or the compensatory rule, then the student surely will be identified as gifted by the complementary rule.

The combination rules used to combine multiple measures greatly influences the quality of a high-stakes decision (Chester, 2003; Douglas, 2007). Further, the correlations between the multiple measures and criteria of classification both influence the results (Douglas, 2007). Different combination rules have their unique strengths as well as weaknesses in comparison with the other combination rules. Potentially, the combination of combination rules may enhance the advantage and simultaneously weaken or eliminate the shortcoming.

The conjunctive rule. The conjunctive rule is the most restrictive rule among the three combination rules (Chester, 2003; Douglas, 2007; Douglas & Mislevy, 2010). The conjunctive rule is appropriate in the situation where specific performance on multiple measures is simultaneously required to reach a decision. For example, students need to attain the minimum standards on tests of all required subjects (e.g., reading, math, writing) to earn a high school diploma. The failure to meet the minimum standard on the test of one subject can result in the failure to earn the high school diploma. Therefore, the “minimum standard” on each test should be decided with great caution. Further, it is critical to determine whether the “minimum standard” should be the same or different for different subjects. As the most restrictive one among the three combination rules, the conjunctive rule may relax its restriction by applying different standards on different measures. In the field of gifted education, some scholars suggested that students who were qualified for gifted programs should have above average

abilities across all relevant domains and simultaneously excel in one or more areas to qualify for gifted programs (Page, 2006; Renzulli & Reis, 1994). This suggestion on the practice of gifted identification used the conjunctive rule to combine an outstanding standard on the measure of the target form of giftedness and a standard of the population mean on the measures of relevant abilities. Conceptually, using the conjunctive rule to combined different standards on measures of different abilities suggests that students who are gifted in one domain area are not necessarily gifted in all relevant domain areas. Therefore, gifted identification should not apply the same outstanding standard to measures of all relevant abilities but only to the measure of the target ability. It is critical to decide appropriate standards on measures of relevant abilities for gifted identification. The present study has applied statistical simulation to systematically explore the appropriate standards on measures of relevant abilities that will improve the performance of gifted identification using the conjunctive rule to combine multiple measures.

The reliability of the decision based on the conjunctive rule is determined by the least reliable measure among the multiple measures (Chester, 2003). Considering the use of conjunctive rule in gifted identification, the reliability of the identification of gifted should be lower than the lowest reliability and reduce by adding more test scores. Cronbach et al's (1997) concern about the compound error in results of the conjunctive rule should be more severe in identifying gifted because the standard error for true abilities locating at the right tail of the distribution is relatively larger than the average standard error across the whole range of true abilities. Alternatively speaking, the reliability conditional on a super high true ability is lower than the marginal reliability averaged across the distribution of true abilities. Therefore, the conjunctive rule should be only applied to observed abilities with adequate reliabilities (e.g., $r < .70$; Peters, 2014) to identify gifted. If the reliability is consistent across measures and time,

using the conjunctive rule to combine different measures of the same construct or repeated measures across time can help to draw validating or confirming inference. However, if the reliabilities are not consistent across measures or time, using the conjunctive rule to combine multiple measures or repeated measures of the same construct may not be effective to improve the performance of gifted identification. As the reliability is not consistent across multiple measures, using multiple measures of the same construct combined by the conjunctive rule does not perform better than using the most reliable measure only.

The compensatory rule. The compensatory model is widely used in the field of education (Chester, 2003; Douglas & Mislevy, 2010). The flexibility in using the compensatory rule could be the weight for each measure as combining them to make decisions. The weight for different measures could be the same or different. The total score of multiple measures can be viewed as using the compensatory model to combine multiple measures with equal weights (e.g., Wechsler full scale IQ scores, the SAT composite scores). It is also very common to combine multiple measures with different weights. For example, in college, it is common that the final grade of a course is determined by in-class quiz and participation (worth 20%), mid-term exam (worth 30%), and the final exam (50%). In both cases that multiple measures were combined by the compensatory rule, weaker performance on one exam or section could be offset by stronger performance on the other exam and in-class quiz. Further, the “Mean” rule may be a solution to the issue of measurement error when it is used to combine multiple measures of the same ability or similar abilities (Sternberg, Grigorenko, & Bundy, 2001). However, the compensatory model will not be effective to deal with the issue of measurement error if reliabilities are low across measures, because, in such a scenario, the reliability of mean scores would still be low.

However, gifted identification using the compensatory rule to combine multiple measures may diverge from the definition of gifted students. By definition (NAGC, 2010), students are gifted if they present outstanding capabilities in at least one domain area, which should be unconditional on their abilities in the other domain areas. When multiple measures of different abilities are combined by the “mean” rule to identify gifted students and the high standard (e.g., a z-score of two corresponding to the 97.5th percentile) is not set for the measure of a specific ability (e.g., math) but for the mean of multiple measures (e.g., the mean or sum of math and reading scores), then students can be identified as gifted only if their scores on all the measures fulfill the high standard for the mean or their scores their score on one measure is extremely high and simultaneously their scores on the other measures are not too low. For example, a mean of two could be achieved by averaging two scores of two on both math and intelligence tests or by averaging a score of three on the math test and a score of one on the intelligence test. Further, reducing the criterion for the mean of multiple measures may potentially confound averagely highly capable students and gifted students.

The complementary rule. A popular example of the typical application of the complementary rule is the way that *SAT* scores are used during the college admission process. Students may take the *SAT* and they only need to submit their single best test scores to prospective colleges. Therefore, the best performance on the *SAT* is regarded as a sufficient indicator of students’ readiness for college and likelihood of success in college (Wyatt, Kobrin, Wiley, Camara, Proestler, 2011).

Among the three combination rules, the complementary rule best matches the definition of gifted students given by the NAGC (2015) or the USED (1993, 2001). To be specific, the definition of gifted children given by the NAGC suggests that children who present high aptitude

(at or above the 90th percentile) in any one or more domain areas are gifted. That is to say, conceptually, meeting the criterion of giftedness for the measure of abilities in any one domain area is sufficient to identify a student as gifted. Using the complementary rule to combine different criteria for multiple measure conveys the same idea that meeting the criterion for any one measure is sufficient for making the decision. Using the complementary rule also identifies more students as gifted than the other two rule will. However, it is challenging for the complementary rule to improve the precision of the identification of gifted and distinguish students gifted in different domain areas from each other.

The combination of combination rules. In the field of education, it is common to combine the combination rules to make decisions (Douglas and Mislevy, 2010). For example, composite SAT scores combine SAT math and reading scores by the compensatory rule and then decisions on college admission apply the complementary rule to students' repeated composite SAT scores. In the field of gifted education, scholars suggested that gifted students who are identified for gifted programs need to excel in one or more domain areas and simultaneously present above average abilities in all relevant domain areas (Page, 2006; Renzulli & Reis, 1994). This suggestion on gifted identification applied a mixed conjunctive-complementary approach, which combined the conjunctive rule and the complementary rule. To be specific, this suggestion applied the conjunctive rule to combine a very high criterion for the measure of abilities in a specific domain area of giftedness and a relatively lower criterion (e.g., the population mean) for the measures of abilities in the other relevant domain areas to identify each form of giftedness and then applied the complementary rule to capture different forms of giftedness. In such a situation, the complexity of decision rules increased. The practice of gifted identification abased on such complex combination rules calls for a very detailed guidance regarding the process of

gifted identification, such as the order of applying combination rules and the criterion for each measure.

Performance of Gifted Identification Based on Multiple Measures

The main challenge of examining the performance of gifted identification is that students' true abilities are not directly observed and observed abilities are fraught with measurement error. Given that, McBee, Peters, and Waterman (2014) conducted a simulation study to thoroughly investigate the performance of gifted identification using different rules (e.g., conjunctive, compensatory, and complementary) to combine multiple measures of different constructs in terms of the incorrect identification rate (the probability of being non-TG given IG) and sensitivity.

McBee et al. (2014) assumed that test reliabilities and correlations between true abilities (true correlations) were known parameters. Based on the assumption of homoscedasticity, they calculated the correlations corrected for attenuation between true abilities and observed abilities in different domain areas and the correlations corrected for attenuation between observed abilities in different domain areas. Further, they assumed that true abilities and observed abilities in different domain areas followed the standard multivariate normal distribution with means of zero and the covariance matrix of true correlations and corrected correlations. However, intelligence tests and achievement tests are widely designed based on the Rasch model (Rasch, 1980) or the 1P-IRT model. Therefore, the information of each item and the entire test is conditional on true abilities and varies across the range of true abilities (Hambleton, Swaminathan, & Rogers, 1991; Kim & Feldt, 2010). The test is most informative for the abilities around the population mean but least informative for the abilities at the two tails. Given that, the conditional standard error is bigger in the observed abilities of the high abilities at the two tails.

The assumption of homoscedasticity underestimates the standard error in the observed abilities of super high abilities and overestimates the correlations between observed abilities.

Regardless of the potential violation of assuming homoscedasticity, McBee et al. (2014) evaluated the performance of gifted identification using three combination rules to combine multiple measures in terms of the incorrect identification rate and sensitivity. As mentioned earlier, the incorrect identification rate is the probability of being non-TG given IG and sensitivity is the probability of being IG given TG. Based on multiple measures, they tested the performance of gifted identification using three combination rules: the conjunctive rule, the complementary rule, and the compensatory rule (Chester, 2003). In brief, they evaluated the performance of gifted identification that (1) used the conjunctive rule or the complementary rule to combine two strongly correlated true scores ($r = .80$) and corresponding observed scores with the same reliability; (2) used the compensatory rule to combine multiple strongly correlated true scores ($r = .80$, $n = 2, 3$, or 4) and corresponding observed scores with the same reliability; and (3) used the conjunctive rule to combine multiple true scores ($n = 2$ or 3) correlated at different levels and corresponding observed scores with different reliabilities. Their findings suggested that, in terms of sensitivity, the complementary rule performs the best, the conjunctive rule performs the worst, and the compensatory rule performs in the middle. Using observed scores with lower reliabilities resulted in less sensitivity and higher incorrect identification rate. Further, using higher criterion of gifted could reduce the incorrect identification rate but simultaneously reduce the sensitivity. For example, given the cutoff of the 90th percentile, the sensitivity of the conjunctive rule based on two observed scores with a reliability of .95 is .79, which misses about 20% of TG students; the sensitivity of the complementary rule based on two observed scores with a reliability of .95 is .87, which misses about 13% of TG students. However, the issue is

that the TG students defined based on the conjunctive rule and the complementary rule are different. The TG students defined based on the conjunctive rule are actually a subgroup within the TG students defined based on the complementary rule. Given that the TG used to calculate the sensitivity is different for the conjunctive rule and complementary rule, the sensitivity cannot be compared across combination rules.

McBee et al. (2014) also examined the effect of the reliability and the criterion of gifted on the performance of gifted identification. For a cutoff score corresponding to the 90th percentile, as the reliability reduced from .95 to .70, the sensitivity of gifted identification using the conjunctive rule reduced from .79 to .45, the sensitivity of gifted identification using the complementary rule reduced from .87 to .71, and the sensitivity of gifted identification based on a single-measure reduced from .84 to .60. Based on observed scores with a reliability of .95, as the criterion of gifted increased from the 90th percentile to the 99th percentile, the sensitivity of gifted identification using the conjunctive rule decreased from .79 to .69, the sensitivity of gifted identification using the complementary rule decreased from .87 to .79, and the sensitivity of gifted identification using the single-measure method decreased from .84 to .76.

The uses of combination rules in gifted identification can have a great variability by varying the combination rule and criterion of gifted. In public school system, it is very common that students are selected as candidates for gifted programs based on their high performance on an academic achievement test (e.g., math test or reading test); however, the candidates must fulfill the criterion of gifted for the mental ability test to be qualified for the gifted program. In such a practical situation, the achievement tests are combined by the complementary rule and further combined by the conjunctive rule with the mental ability test. Given the popularity of this method of gifted identification in practice, it is worthy of investigation on its performance. It is

also worthy of thorough investigation on the performance of gifted identification using the combination rule to combine different criteria of gifted for different measures. As mentioned earlier, some have suggested that students selected for gifted programs should present outstanding abilities in one or more domain areas and above average abilities in relevant domain areas (Page, 2006; Renzulli & Reis, 1994). According to this suggestion, different criteria for different measures are combined by the conjunctive rule to identify students who are gifted in one specific domain area. The identification results of the conjunctive rule can be further combined by the complementary rule to identify a larger group of gifted students.

Chapter 3

Methods

This study applies statistical simulation techniques to generate true and observed abilities in three domain areas: intelligence, math, and reading. The simulation is conducted in RStudio (V1.0.143; 2017), an integrated development environment (IDE) for R (V3.3.1).

Notations

To make the description of the process of simulation and analyses readable, all the notations of statistical terms used throughout this study are summarized here.

1. T : T refers to students' true abilities. To be specific, T_1 , T_2 , and T_3 stand for students' true abilities in intelligence, math, and reading.
2. X' and E' : X' refers to students' observed abilities. E' refers to measurement error. To be specific, X'_1 , X'_2 , and X'_3 stand for students' observed abilities in intelligence, math, and reading, respectively. E'_1 , E'_2 , and E'_3 stand for the measurement error in the students' observed abilities in intelligence, math, and reading, respectively.

According to the classic test theory (CTT), X' is the sum of the true abilities and measurement error.

$$X' = T + E'$$

3. X : X refers to standardized students' observed abilities. To be specific, X_1 , X_2 , and X_3 stand for the standardized students' observed abilities in intelligence, math, and reading, respectively. I shall describe the process of standardization further on.
4. N : N refers to the sample size.
5. \bar{X} and S : \bar{X} refers to sample mean. S refers to sample standard deviation.
6. n : n refers to the test length.

7. C : C refers to a 3-element vector of criteria for the three measures, which are the cutoff scores used for gifted identification. Further, c refers to the criterion for a single measure. To be specific, c_1 , c_2 , and c_3 refers to the criteria for the measures of students abilities in intelligence, math, and reading.

$$C = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

8. P : P refers to probability.
9. μ and σ : μ refers to mean and σ refers to standard deviation. The subscripts of μ and σ specify the mean and standard deviation of which variable they stand for. For example, $\mu_{X'}$ refers to the mean of the unstandardized students' observed abilities.
10. ρ : ρ refers to the correlation between observed abilities and true abilities. Therefore, ρ^2 refers to the test-level reliability.
11. θ : θ refers to ability.
12. i : i refers to item.
13. u_i : u_i refers to a binary score on item i .

$$u_i = \begin{cases} 1, \text{correct answer;} \\ 0, \text{wrong answer;} \end{cases}$$

14. b_i : b_i refers to the difficulty of item i
15. $P_i(\theta)$ and $Q_i(\theta)$: $P_i(\theta)$ refers to the probability of students with the ability of θ whose answers to item i are correct ($u_i = 1$).

$$P_i(\theta) = P(u_i = 1|\theta, b_i) = \frac{1}{1 + e^{-1.7(\theta - b_i)}}$$

$$Q_i(\theta) = 1 - P_i(\theta)$$

16. $I_i(\theta)$ and $I(\theta)$: $I_i(\theta)$ refers to the information of item i conditional on the ability θ .

$I(\theta)$ refers to the test information conditional on the ability of θ .

$$I(\theta) = \sum_{i=1}^n I_i(\theta) = \sum_{i=1}^n \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}$$

17. $\varphi(\theta)$: $\varphi(\theta)$ refers to the probability density function of the standard normal distribution.

$$\varphi(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\theta^2}{2}}$$

18. SE : SE refers to standard error.

19. r : r refers to correlation. To be specific, r_{12} , r_{13} , and r_{23} stand for the correlation between students' true ability in intelligence and math, the correlation between students' true ability in intelligence and reading, and the correlation between students' true ability in math and reading.

20. Σ : Σ refers to the variance-covariance matrix.

21. N : N refers to the normal distribution. For example, $N(0, \Sigma)$ refers to a normal distribution with means of zeros and the variance-covariance matrix of Σ .

Data Generation

To evaluate the performance of gifted identification, both true abilities and observed abilities in the three domain areas (intelligence, math, and reading) are generated and therefore both observed and true abilities are known in this simulation study.

Generate true abilities. This study assumes that students' true intelligence, true mathematical abilities, and true reading abilities follow the standard multivariate normal distribution.

$$T \sim N(0, \Sigma)$$

Given this assumption, the variance-covariance matrix (Σ) of the three true abilities is identical with the correlation matrix. The correlations between true abilities consulted the results from McCoach, Yu, Gottfried, and Gottfried's study (2017) that reported the correlations among the latent intelligence, latent math achievement, and latent reading achievement. To be specific,

$$\Sigma = \begin{bmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.85 & 0.7 \\ 0.85 & 1 & 0.75 \\ 0.7 & 0.75 & 1 \end{bmatrix}$$

where $r_{12} = r_{21}$, referring to the correlation between the true intelligence and mathematical ability; $r_{13} = r_{31}$, referring to the correlation between the true intelligence and true reading ability; and $r_{23} = r_{32}$, referring to the correlation between the true mathematical and true reading abilities. The syntax used to generate true abilities is presented in Table 1. In total, 1,000,000 true abilities in intelligence, math, and reading are simultaneously generated using the “mvrnorm” function in the r package “MASS”.

Table 1

Syntax Used to Generate True Abilities in Intelligence, Math, and Reading

```
# Multivariate Normal Sample
SigmaTrue=matrix(c(1,.85,.70,.85,1,.75,.70,.75,1),3,3) #VarCov matrix of true scores
MeanTrue=rep(0, times = 3)
VarTrue=rep(1,times = 3)
set.seed(10001)
TrueScore=mvrnorm(n=1000,MeanTrue,SigmaTrue)
```

Generate observed abilities. To generate observed abilities, this study firstly generated three tests based on the 1P-IRT model. Therefore, item difficulties are known parameters in this study. Given the item difficulties and previously-generated true abilities, the item response pattern is generated for each of the 1,000,000 true abilities. Students' observed abilities were estimated abilities based on item response patterns.

Generate Tests. This study generated three tests based on the 1P-IRT models and assumed that they were the intelligence test, math test, and reading test, respectively. According

to the published technical reports, the marginal reliability of full-scale IQ scores on the *Wechsler Intelligence Scale for Children* (5th ed., 2014) is .96, the marginal reliabilities of standard scores on the 3rd-grade *Connecticut Mastery Test* (CMT, Hendrawan & Wibowo, 2011) is .94 for both math and reading. To introduce variability in the marginal reliabilities, the marginal reliabilities for observed intelligence, observed mathematical abilities, and observed reading abilities are planned to be 0.95, 0.9, and 0.8, respectively. This study adjusted the test length to achieve the desired marginal reliabilities. The syntax used to develop tests with the target reliabilities is presented in Table 2 (a). The syntax used to generate item response patterns is presented in Table 2 (b). Further, the item difficulties of the three simulated tests are presented in Table 3, Table 4, and Table 5. The final lengths and marginal reliabilities of the three simulated tests are:

$$\begin{cases} n_1 = 130, \rho_1^2 = .946 \\ n_2 = 70, \rho_2^2 = .901 \\ n_3 = 30, \rho_3^2 = .803 \end{cases}$$

Table 2

(a) Syntax Used to Develop Tests

```
rho1=0.95 # marginal reliability of the intelligence test
rho2=0.90 # marginal reliability of the math test
rho3=0.80 # marginal reliability of the intelligence test
diffTest1.final = diffTest(rho1) # item difficulties of the intelligence test
diffTest2.final = diffTest(rho2) # item difficulties of the math test
diffTest3.final = diffTest(rho3) # item difficulties of the reading test
# function of generating item difficulties to achieve a given marginal reliability (rho).
diffTest = function (rho) {
  rep=1000
  length.test=30
  set.seed(10001)
  normAb=rnorm(2000,mean = 0, sd = 1)
  diffTest=0
  diffTest.final=0
  info.test.people = 0
  for(try in 1:rep){
    diffTest = matrix(runif(length.test,-2,2),nrow=1)
    test.answers=irtgen(diffTest, normAb)
    colnames(test.answers)=paste0('itme',1:length.test)
  }
}
```

```

ab.int=mirt(test.answers,l,itemtype = 'Rasch')
diff_rho=abs(marginal_rxx(ab.int)-rho)
if(diff_rho <= 0.005) {
  diffTest.final=diffTest
  break
}
length.test=length.test+5
}
Return (diffTest.final)
}

```

Note. “mirt” is a multidimensional item response theory package for the R environment (Chalmers, 2012).

(b) Syntax Used to Generate Item Responses

```

irtgen = function(b,theta){
  nexam = length(theta)
  n = length(b)
  prob = 0
  b = matrix(b,ncol=1)
  theta = matrix(theta,ncol=1)
  d = apply(theta,1,"-",b)
  prob = t(1/(1+exp(-1.7*d)))
  data = matrix(0,nrow=nexam,ncol=n)
  try = matrix(runif(nexam*n,0,1),nrow=nexam,ncol=n)
  data = replace(data,which(try<prob),1)
  return(data)
}

```

Table 3

The Item Difficulties of the Simulated Intelligence Test

Item	b_i	Item	b_i	Item	b_i	Item	b_i	Item	b_i
1	-0.729	27	0.401	53	-1.172	79	1.048	105	-0.655
2	-0.592	28	-0.329	54	-0.863	80	-1.106	106	-0.249
3	-1.675	29	-1.077	55	0.54	81	0.537	107	1.235
4	-0.601	30	-0.552	56	-1.769	82	0.256	108	1.291
5	-1.811	31	1.124	57	-0.635	83	0.239	109	-0.516
6	-0.462	32	-1.928	58	1.469	84	1.153	110	-1.954
7	-1.113	33	-1.298	59	-0.348	85	0.398	111	1.596
8	-0.526	34	-0.45	60	1.844	86	-0.78	112	-1.133
9	0.011	35	-1.773	61	0.813	87	1.34	113	-1.373
10	0.904	36	1.439	62	1.917	88	-1.72	114	1.166
11	-1.332	37	-0.861	63	0.07	89	-1.151	115	-1.442
12	-0.441	38	-1.169	64	0.026	90	1.178	116	0.767
13	-0.978	39	0.37	65	-1.894	91	1.485	117	0.627

Item	b_i	Item	b_i	Item	b_i	Item	b_i	Item	b_i
14	-0.829	40	0.04	66	1.083	92	0.206	118	0.555
15	-0.653	41	-0.716	67	-1.16	93	-1.399	119	-1.045
16	1.586	42	-0.11	68	-1.681	94	-0.561	120	-0.355
17	-1.539	43	-0.527	69	0.873	95	-1.019	121	1.999
18	-0.476	44	-0.175	70	-1.325	96	-0.817	122	-1.142
19	-1.277	45	1.161	71	-1.189	97	1.927	123	0.425
20	0.503	46	1.785	72	1.161	98	1.915	124	1.759
21	-1.129	47	1.191	73	-0.565	99	-1.093	125	0.123
22	0.273	48	-0.303	74	1.179	100	1.507	126	-1.064
23	0.901	49	1.938	75	-1.09	101	-0.579	127	1.815
24	1.025	50	-0.651	76	1.018	102	0.225	128	-1.915
25	1.371	51	0.448	77	-0.021	103	-1.226	129	-1.36
26	-0.153	52	-1.351	78	-0.4	104	1.815	130	-1.47

Note. b_i = the difficulty of item i in the simulated intelligence test.

Table 4

The Item Difficulties of the Simulated Math Test

Item	b_i	Item	b_i	Item	b_i	Item	b_i	Item	b_i
1	-0.875	15	1.838	29	0.479	43	0.051	57	1.669
2	0.037	16	-1.395	30	1.897	44	1.813	58	-0.544
3	-1.489	17	0.121	31	-1.078	45	0.17	59	-1.518
4	1.893	18	-1.55	32	0.194	46	-1.167	60	-0.367
5	0.375	19	-0.942	33	-0.352	47	-0.98	61	-0.47
6	-0.38	20	0.3	34	-1.364	48	1.381	62	0.594
7	-0.695	21	-1.979	35	-0.959	49	0.406	63	-1.802
8	-1.451	22	0.114	36	-1.395	50	-0.835	64	-1.234
9	1.694	23	-0.673	37	-1.262	51	1.953	65	1.115
10	-1.445	24	0.839	38	1.565	52	0.569	66	1.536
11	-0.903	25	-0.702	39	1.217	53	0.56	67	-1.885
12	0.087	26	0.309	40	1.215	54	1.77	68	-1.067
13	1.323	27	-0.195	41	-1.284	55	-1.346	69	1.137
14	-1.461	28	-0.962	42	1.645	56	-1.278	70	0.591

Note. b_i = the difficulty of item i in the simulated math test.

Table 5

The Item Difficulties of the Simulated Reading Test

Item	b_i	Item	b_i	Item	b_i	Item	b_i	Item	b_i
1	0.222	7	-1.479	13	1.454	19	-0.354	25	-1.119
2	-0.763	8	0.044	14	-1.597	20	0.829	26	-0.963
3	1.709	9	0.217	15	-0.952	21	0.925	27	1.748
4	-0.414	10	1.533	16	1.954	22	-0.141	28	1.298
5	0.279	11	0.936	17	-0.277	23	-1.906	29	-1.492

Item	b_i	Item	b_i	Item	b_i	Item	b_i	Item	b_i
6	-1.149	12	0.229	18	0.529	24	0.681	30	-1.915

Note. b_i = the difficulty of item i in the simulated reading test.

For 1P-IRT model, given the known item difficulties, for students with the abilities of θ_1 in intelligence, θ_2 in math, and θ_3 in reading, the probabilities of correctly answering item i_1 in the simulated intelligence test, item i_2 in the simulated math test, and item i_3 in the simulated reading test are

$$\begin{cases} P_{i_1}(\theta_1) = P(u_{i_1} = 1 | \theta_1, b_{i_1}) = \frac{1}{1 + e^{-1.7(\theta_1 - b_{i_1})}} \\ P_{i_2}(\theta_2) = P(u_{i_2} = 1 | \theta_2, b_{i_2}) = \frac{1}{1 + e^{-1.7(\theta_2 - b_{i_2})}} \\ P_{i_3}(\theta_3) = P(u_{i_3} = 1 | \theta_3, b_{i_3}) = \frac{1}{1 + e^{-1.7(\theta_3 - b_{i_3})}} \end{cases} \quad (1)$$

$$\begin{cases} Q_{i_1}(\theta_1) = 1 - P_{i_1}(\theta_1) \\ Q_{i_2}(\theta_2) = 1 - P_{i_2}(\theta_2) \\ Q_{i_3}(\theta_3) = 1 - P_{i_3}(\theta_3) \end{cases} \quad (2)$$

Further, the information functions of the simulated tests conditional on true abilities are

$$\begin{cases} I(\theta_1) = \sum_{i_1=1}^{n_1} I_{i_1}(\theta_1) = \sum_{i_1=1}^{n_1} \frac{[P'_{i_1}(\theta_1)]^2}{P_{i_1}(\theta_1)Q_{i_1}(\theta_1)}, \quad n_1 = 130 \\ I(\theta_2) = \sum_{i_2=1}^{n_2} I_{i_2}(\theta_2) = \sum_{i_2=1}^{n_2} \frac{[P'_{i_2}(\theta_2)]^2}{P_{i_2}(\theta_2)Q_{i_2}(\theta_2)}, \quad n_2 = 70 \\ I(\theta_3) = \sum_{i_3=1}^{n_3} I_{i_3}(\theta_3) = \sum_{i_3=1}^{n_3} \frac{[P'_{i_3}(\theta_3)]^2}{P_{i_3}(\theta_3)Q_{i_3}(\theta_3)}, \quad n_3 = 30 \end{cases} \quad (3)$$

$$\text{where } P'_{i_1}(\theta_1) = \frac{\partial P'_{i_1}(\theta_1)}{\partial \theta_1} = \frac{\partial \left[\frac{1}{1 + e^{-1.7(\theta_1 - b_{i_1})}} \right]}{\partial \theta_1} = \frac{1.7 \left[e^{-1.7(\theta_1 - b_{i_1})} \right]}{\left[1 + e^{-1.7(\theta_1 - b_{i_1})} \right]^2} = 1.7 P_{i_1}(\theta_1) Q_{i_1}(\theta_1). \text{ In sum,}$$

$$\begin{cases} P'_{i_1}(\theta_1) = 1.7 P_{i_1}(\theta_1) Q_{i_1}(\theta_1) \\ P'_{i_2}(\theta_2) = 1.7 P_{i_2}(\theta_2) Q_{i_2}(\theta_2) \\ P'_{i_3}(\theta_3) = 1.7 P_{i_3}(\theta_3) Q_{i_3}(\theta_3) \end{cases} \quad (4)$$

Replacing $P'_{i_1}(\theta_1)$, $P'_{i_2}(\theta_2)$, and $P'_{i_3}(\theta_3)$ in equation (3) with the results in equation (4), then the

$$\begin{cases} I(\theta_1) = \sum_{i_1=1}^{n_1} (1.7)^2 P_{i_1}(\theta_1) Q_{i_1}(\theta_1), & n_1 = 130 \\ I(\theta_2) = \sum_{i_2=1}^{n_2} (1.7)^2 P_{i_2}(\theta_2) Q_{i_2}(\theta_2), & n_2 = 70 \\ I(\theta_3) = \sum_{i_3=1}^{n_3} (1.7)^2 P_{i_3}(\theta_3) Q_{i_3}(\theta_3), & n_3 = 30 \end{cases} \quad (5)$$

Further, the standard error conditional on true abilities is

$$\begin{cases} SE(\hat{\theta}_1) = \sqrt{\frac{1}{I(\theta_1)}}, & n_1 = 130 \\ SE(\hat{\theta}_2) = \sqrt{\frac{1}{I(\theta_2)}}, & n_2 = 70 \\ SE(\hat{\theta}_3) = \sqrt{\frac{1}{I(\theta_3)}}, & n_3 = 30 \end{cases} \quad (6)$$

Given the conditional standard error in equation (9), the conditional reliability is

$$\begin{cases} \rho_{1|\theta_1}^2 = \frac{1}{1+SE^2(\hat{\theta}_1)} = \frac{I(\theta_1)}{I(\theta_1)+1} \\ \rho_{2|\theta_2}^2 = \frac{1}{1+SE^2(\hat{\theta}_2)} = \frac{I(\theta_2)}{I(\theta_2)+1} \\ \rho_{3|\theta_3}^2 = \frac{1}{1+SE^2(\hat{\theta}_3)} = \frac{I(\theta_3)}{I(\theta_3)+1} \end{cases} \quad (7)$$

It can be assumed without loss of generality that true abilities are standardized. Therefore,

$$\begin{cases} \varphi(\theta_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\theta_1^2}{2}} \\ \varphi(\theta_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\theta_2^2}{2}} \\ \varphi(\theta_3) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\theta_3^2}{2}} \end{cases} \quad (8)$$

The average information across the entire range of abilities is

$$\begin{cases} I_1 = \int_{-\infty}^{+\infty} I(\theta_1) \varphi(\theta_1) d\theta_1 \\ I_2 = \int_{-\infty}^{+\infty} I(\theta_2) \varphi(\theta_2) d\theta_2 \\ I_3 = \int_{-\infty}^{+\infty} I(\theta_3) \varphi(\theta_3) d\theta_3 \end{cases} \quad (9)$$

The test-level reliabilities are

$$\begin{cases} \rho_1 = \frac{I_1}{I_1+1} \\ \rho_2 = \frac{I_2}{I_2+1} \\ \rho_3 = \frac{I_3}{I_3+1} \end{cases} \quad (10)$$

As reported earlier, the marginal reliabilities are .946 for observed abilities from the intelligence test with 130 items, .901 for observed abilities from the math test with 70 items, and .803 for observed abilities from the reading test with 30 items. Further, the conditional reliabilities of observed intelligence range from .56 to .98, the conditional reliabilities of observed mathematical abilities range from .47 to .97, and the conditional reliabilities of observed reading abilities range from .24 to .93 across the entire range of true abilities. The plots of test information conditional on true abilities are presented in Figure 1 and the plots of conditional standard error are presented in Figure 2.

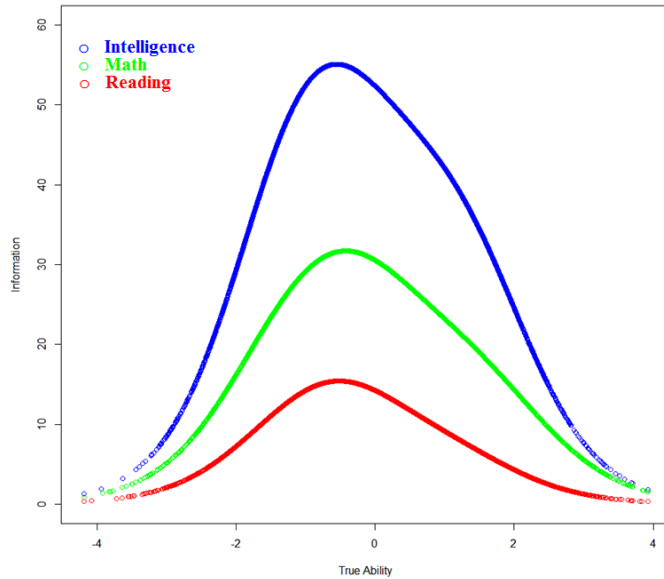


Figure 1. The plot of test information conditional on abilities.

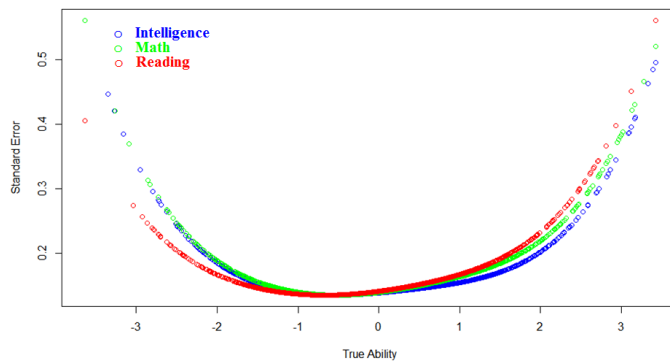


Figure 2. The plot of standard error conditional on abilities.

Given equation (6), conditional information and conditional standard error should present opposite patterns as observed in Figure 1 and Figure 2. The conditional test information achieves the peak around the population mean, which suggests that the test is most informative for students whose true abilities are around the population mean. The test is least informative for students whose true abilities locate at the two tails of the distribution of true abilities. On the contrary, the conditional standard error reaches the peak at the two tails of the distribution of true abilities, which suggests that the measurement is least accurate for students who have either extremely high or low abilities.

Generate observed abilities. The observed abilities were generated based on item response patterns. The “irtgen” function, shown in Table 2 (b), was used to generate the item response patterns based on previously generated true abilities and tests. For each of the 1,000,000 true abilities in each of the three domain areas, an item response pattern was generated. Based on the item response patterns, the “mirt” function in the “mirt” (Chalmers, 2012) package for the R environment was used to estimate the difficulty of each item in each of the three tests. The “mirt” function runs an unconditional factor analysis model to estimate the item parameters. In this study, only item difficulties were estimated by specifying “itemtype = ‘Rasch’”. Based on the estimated item difficulties and sum scores, the “fscores” function in the “mirt” package was used to estimate abilities. The “fscore” computes EAP (expect a posteriori) estimates based on the item difficulties estimated by the “mirt” function and the sum scores. EAP derives from Bayesian principle. Rather than finding the maximum value of the likelihood function as in ML method, it calculates the expected value of the posterior distribution. It divides the posterior distribution into slices according to the number of quadrature points, and computes

the weighted average value of the posterior distribution (Bock & Mislevy, 1982). The syntax is provided in Appendix A.

The criterion of gifted for true abilities and observed abilities should refer to the same percentile and result in approximately the same proportion of TG and IG students. Given that the true abilities were assumed and generated to follow the standard normal distribution, the observed abilities were rescaled to the standard distribution with a mean of 0.0 and a standard deviation of 1.0.

$$\text{Observed Mean} = \begin{cases} \overline{X'_1} = \frac{\sum_{j=1}^{1000} X'_{1j}}{1000} \\ \overline{X'_2} = \frac{\sum_{j=1}^{1000} X'_{2j}}{1000} \\ \overline{X'_3} = \frac{\sum_{j=1}^{1000} X'_{3j}}{1000} \end{cases}$$

$$\text{Observed Standard Deviation} = \begin{cases} S_1 = \sqrt{\frac{\sum_{j=1}^{1000} X_{1j}^2 - \frac{1}{n}(\sum_{j=1}^{1000} X_{1j})^2}{n-1}} \\ S_2 = \sqrt{\frac{\sum_{j=1}^{1000} X_{2j}^2 - \frac{1}{n}(\sum_{j=1}^{1000} X_{2j})^2}{n-1}} \\ S_3 = \sqrt{\frac{\sum_{j=1}^{1000} X_{3j}^2 - \frac{1}{n}(\sum_{j=1}^{1000} X_{3j})^2}{n-1}} \end{cases}$$

$$\text{Rescaled Observed Ability} = \begin{cases} X_{1j} = \frac{X'_{1j} - \overline{X'_1}}{S_1} \\ X_{2j} = \frac{X'_{2j} - \overline{X'_2}}{S_2} \\ X_{3j} = \frac{X'_{3j} - \overline{X'_3}}{S_3} \end{cases}$$

The descriptive statistics of the true abilities, observed abilities, and rescaled observed abilities are summarized in Table 6. The histograms plots of the generated true abilities, observed abilities, and rescaled observed abilities are presented in Figure 3.

Table 6

Descriptive Statistics of the Generated True Abilities, Observed Abilities, and Rescaled Observed Abilities

	Measures	N	Min	Max	Mean	SD
True Ability	Intelligence	1000000	-4.625	5.073	0.000	1.000
	Math	1000000	-4.752	5.523	0.000	1.001
	Reading	1000000	-4.682	4.699	0.000	1.000
Raw Observed Ability	Intelligence	1000000	-5.528	5.457	0.004	1.682
	Math	1000000	-5.231	5.184	0.002	1.665
	Reading	1000000	-4.593	4.584	0.000	1.620
Rescaled Observed Ability	Intelligence	1000000	-3.290	3.243	0.000	1.000
	Math	1000000	-3.143	3.112	0.000	1.000
	Reading	1000000	-2.836	2.830	0.000	1.000

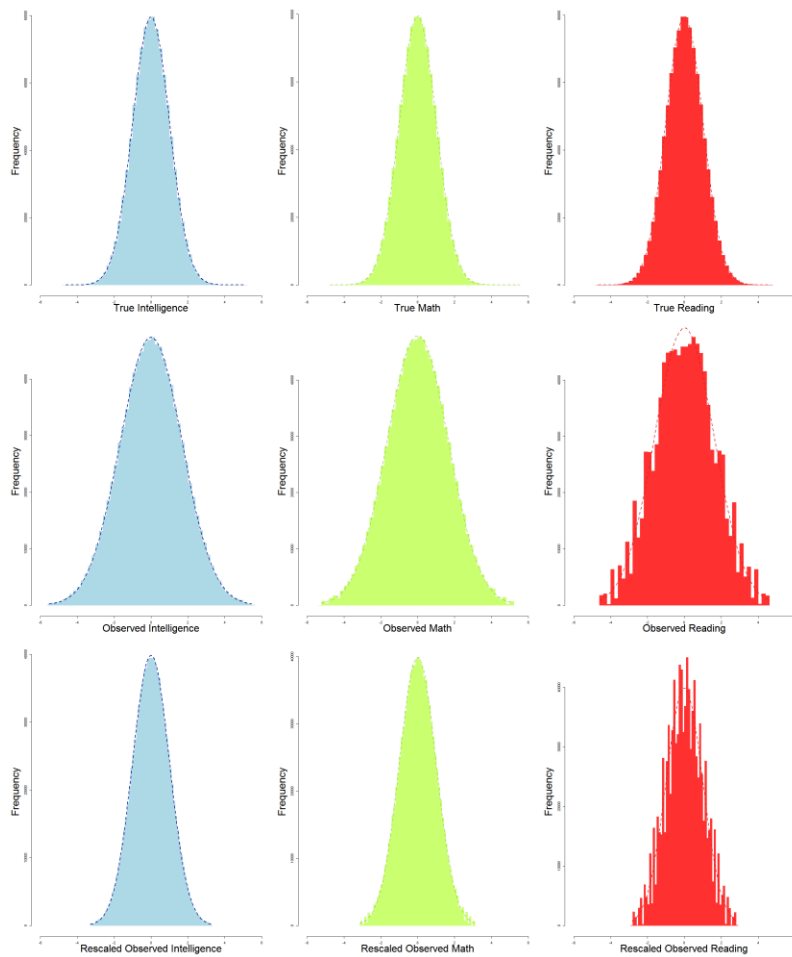


Figure 3. Histogram plots of true abilities, observed abilities, and rescaled observed abilities.

Defining Groups with Different Gifted Characteristics

As briefly mentioned in Chapter 1, to evaluate the performance of different methods of gifted identification, this study defines six groups in terms of the fulfillment of the criterion of gifted for true and observed abilities: the TG group, the IG group, the TPG group, the FNG group, the TNG group and the FPG group. The definitions of the six groups are summarized in Table 7. Given the focus on identifying gifted students, the target groups of this study are the TG group, the IG group, and the TPG group.

Table 7

Definitions of the Six Groups with Different Gifted Characteristics

	True Score (T)	Observed Score (X)	Relationship among Groups
TG	$T \geq C$	NA	TG = TPG + FNG
IG	NA	$X \geq C$	
TPG	$T \geq C$	$X \geq C$	IG = TPG + FPG
FNG	$T \geq C$	$X \leq C$	PPR = $\frac{TPG}{IG}$
TNG	$T \leq C$	$X \leq C$	Sensitivity = $\frac{TPG}{TG}$
FPG	$T \leq C$	$X \geq C$	

Note. T refers to true abilities; X refers to observed abilities; C refers to the criterion of gifted. TG = “truly” gifted; IG = identified gifted; TPG = true positive gifted; FNG = false negative gifted; TNG = true negative gifted; FPG = false positive gifted.

Table 8

Definitions of the Six Groups Based on a Single Measure.

	True Score	Observed Score	Probability	Intelligence	Math	Reading
TG	$T \geq 2.0$	NA	$P(T \geq 2.0)$	2.29%	2.31%	2.31%
IG	NA	$X \geq 2.0$	$P(X \geq 2.0)$	2.28%	2.33%	2.36%
TPG	$T \geq 2.0$	$X \geq 2.0$	$P(T \geq 2.0 \cap X \geq 2.0)$	1.85%	1.74%	1.53%

Note. T refers to true abilities; X refers to observed abilities. TG = “truly” gifted; IG = identified gifted; TPG = true positive gifted.

Using a standard score of 2.0 as the criterion of gifted, the percent of the TG, IG, and TPG groups in each of the three domain areas is summarized in Table 8. Among the one million simulated students, 2.29% is TG in intelligence and 2.31% is TG in math or reading. In total, 4.70% of the one million simulated students are TG in at least one of the three domain areas. The

4.70% gifted population can be further divided into seven mutually exclusive gifted groups, which are illustrated by a Venn diagram in Figure 4. The descriptive statistics of the three true abilities and correlations among the three correlations are presented in Table 9.

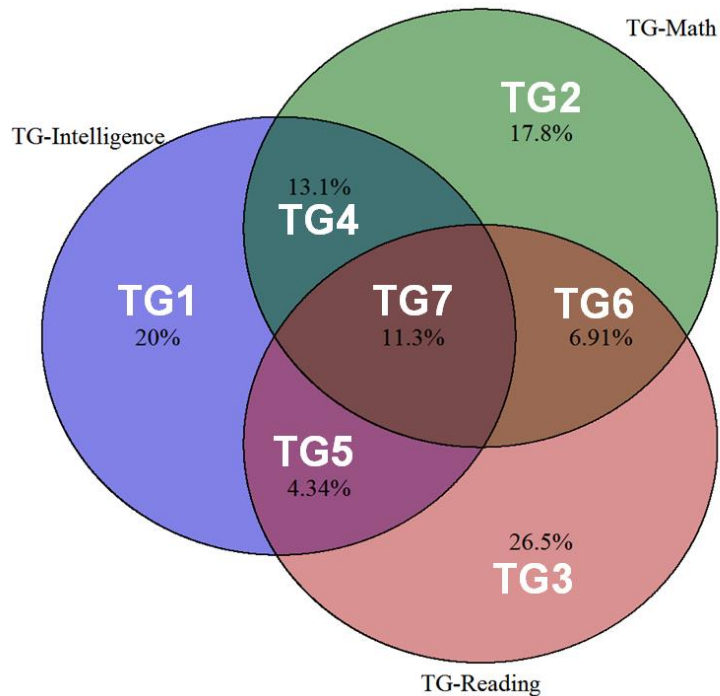


Figure 4. The Venn diagram of the proportions of students with different gifted characteristics.

Table 9

Descriptive Statistics of True Abilities and Correlations among True Abilities in Intelligence, Math, and Reading

Groups	Domain Areas	N	Descriptive Statistics				Correlations		
			Min	Max	Mean	SD	Intelligence	Math	Reading
non-TG	Intelligence	952,985	-4.625	1.995	-0.091	0.924	1.00	.83	.66
	Math	952,985	-4.752	1.995	-0.092	0.922		1.00	.72
	Reading	952,985	-4.682	1.995	-0.089	0.925			1.00
TG1	Intelligence	9,398	1.995	3.521	2.237	0.216	1.00	.20	.11
	Math	9,398	-0.236	1.995	1.503	0.351		1.00	.25
	Reading	9,398	-1.27	1.995	1.143	0.535			1.00
TG2	Intelligence	8,351	-0.133	1.995	1.511	0.353	1.00	.20	.12
	Math	8,351	1.995	3.358	2.224	0.204		1.00	.16
	Reading	8,351	-0.862	1.995	1.290	0.478			1.00
TG3	Intelligence	12,476	-1.463	1.995	1.166	0.515	1.00	.52	.13

Groups	Domain Areas	Descriptive Statistics					Correlations		
		N	Min	Max	Mean	SD	Intelligence	Math	Reading
TG4	Math	12,476	-0.985	1.995	1.296	0.462		1.00	.17
	Reading	12,476	1.995	4.193	2.279	0.258			1.00
	Intelligence	6,181	1.995	4.266	2.414	0.337	1.00	.31	.14
	Math	6,181	1.995	3.861	2.387	0.322		1.00	.18
TG5	Reading	6,181	-0.861	1.995	1.429	0.425			1.00
	Intelligence	2,046	1.995	3.689	2.301	0.260		0.16	0.05
	Math	2,046	0.585	1.995	1.67	0.256		1.00	0.08
	Reading	2,046	1.995	3.993	2.366	0.300			1.00
TG6	Intelligence	3,256	0.207	1.995	1.596	0.306	1.00	0.16	0.07
	Math	3,256	1.995	3.810	2.308	0.262		1.00	0.12
	Reading	3,256	1.995	4.699	2.413	0.340			1.00
	Intelligence	5,307	1.995	5.073	2.565	0.421	1.00	0.49	0.21
TG7	Math	5,307	1.995	5.523	2.612	0.429		1.00	0.29
	Reading	5,307	1.995	4.645	2.544	0.412			1.00

Note. TG1, TG2, and TG3 are the three groups of students who are TG in intelligence, math, and reading, respectively; TG4 is the group of students who are TG in intelligence and math but not in reading; TG5 is the group of students who are TG in intelligence and reading but not in math; TG6 is the group of students who are TG in math and reading but not in intelligence; TG7 is the group of students who are TG in intelligence, math, and reading.

The TG1, TG2, and TG3 are the three groups of students who are TG in only one of the three domain areas (intelligence, math, or reading, respectively). In total, about 65% of the gifted population ($20.0\% + 17.8\% + 26.5\% = 64.3\%$; $64.3\% * 4.70\% = 3.0\%$) is gifted in only one domain area. The TG4, TG5, and TG6 are the three groups of students who are TG in two of the three domain areas but not in the third domain area. In total, about 25% of the gifted population is gifted in two domain areas ($13.1\% + 4.34\% + 6.91\% = 24.35\%$; $24.35\% * 4.70\% = 1.14\%$). TG7 is the group of students who are TG in all three domain areas. Only about 10% of the gifted population ($11.3\% * 4.70\% = 0.53\%$) is TG in all three domain areas. The seven gifted groups are exclusive from each other and also conceptually different gifted populations. Theoretically, the identification of these seven gifted populations should apply different methods. Considering the use of combination rules in gifted identification, the select combination rule should address the characteristics of the target gifted population. For example, using the conjunctive rule to

combine a high standard of gifted for all three measures is a conceptually reasonable method to identify the gifted population TG7, but will definitely miss most of the gifted students in the other six gifted groups. I shall return to this point in more detail later.

Plan of Data Analysis for Research Question One

The first research question is aimed to understand the difference in the performance of gifted identification using different methods. In this study, seven methods of gifted identification are thoroughly investigated and summarized in Table 10.

1. A single-measure method uses observed abilities from a single measure to identify gifted students. The performance of gifted identification based on the intelligence test, the math test, and the reading test is investigated independently. The difference among the three tests is the target gifted population and the reliability of observed abilities.
2. A multi-measure method uses the conjunctive rule to combine different criteria of gifted for different measures. This method of gifted identification uses multiple measures to identify students who are gifted in a specific domain area. A standard score of 2.0 is used as the criterion of gifted for one of the three measures and then a standard score of 0.0 is used as the criterion for the other two measures.
3. A multi-measure method uses the conjunctive rule to combine the same criterion of gifted ($C = 2.0$) for all three measures.
4. A multi-measure method uses the conjunctive rule to combine the same criterion of gifted ($C = 1.28$) for all three measures.
5. A multi-measure method uses the complementary rule to combine the same criterion of gifted ($C = 2.0$) for all three measures.

6. A multi-measure method uses the compensatory rule to identify students who are gifted in all three domain areas or students who are gifted in at least one of the three domain areas.
7. A multi-measure method uses the complementary rule to combine two measures and uses the third measure as the conclusive test. This method thoroughly investigated the performance of gifted identification using the intelligence test, math test, and reading test as the conclusive test separately. The last method posits a practical situation of gifted identification, in which students who pass the screening tests need to pass the conclusive test to receive gifted services. In practice, the intelligence test commonly plays the role of the conclusive test, which makes the intelligence test primarily responsible for the misidentification of students who are truly gifted in the other fields and therefore causes a lot of controversy over the application of intelligence tests in gifted identification. To decide whether the misidentification of TG students is because of the measure used as the conclusive test or the conjunctive rule used to make final decision, this study tests three varieties of the last method: (1) using the complementary rule to combine mathematical and reading abilities in the screening process and using the conjunctive rule to combine the screening result and intelligence in the decision-making process; (2) using the complementary rule to combine intelligence and mathematical abilities in the screening process and using the conjunctive rule to combine the screening result and reading abilities in the decision-making process; and (3) using the complementary rule to combine intelligence and reading abilities in the screening process and using the conjunctive

rule to combine the screening result and math abilities in the decision-making process.

For each of the seven methods, the percent of the TG, IG, and TPG groups is examined. Further, the PPR and sensitivity are calculated. Again, the PPR is the ratio of the percent of TPG to the percent of IG, which is the probability of being TG given IG. The sensitivity is the ratio of the percent of TPG to the percent of TG, which is the probability of being IG given TG.

$$\text{PPR} = \frac{P(\text{TPG})}{P(\text{IG})}$$

$$\text{Sensitivity} = \frac{P(\text{TPG})}{P(\text{TG})}$$

Table 10

Definitions of the Three Groups Based on Multiple Measures.

No.	Combination Rule	Method Name	Criterion of Gifted				Probability of Different Groups		
			Intelligence	Math	Reading	TG	IG	TPG	
1	NA	SM-INT	$T_1 \geq 2$	NA	NA	$P(T_1 \geq 2)$	$P(X_1 \geq 2)$	$P[(T_1 \geq 2) \cap (X_1 \geq 2)]$	
		SM-MATH	NA	$T_2 \geq 2$	NA	$P(T_2 \geq 2)$	$P(X_2 \geq 2)$	$P[(T_2 \geq 2) \cap (X_2 \geq 2)]$	
		SM-READ	NA	NA	$T_3 \geq 2$	$P(T_3 \geq 2)$	$P(X_3 \geq 2)$	$P[(T_3 \geq 2) \cap (X_3 \geq 2)]$	
2	AND	AND-1-INT	$T_1 \geq 2$	$T_2 \geq 0$	$T_3 \geq 0$	$P(T_1 \geq 2 \cap T_2 \geq 0 \cap T_3 \geq 0)$	$P(X_1 \geq 2 \cap X_2 \geq 0 \cap X_3 \geq 0)$	$P[(T_1 \geq 2 \cap T_2 \geq 0 \cap T_3 \geq 0) \cap (X_1 \geq 2 \cap X_2 \geq 0 \cap X_3 \geq 0)]$	
		AND-1-MATH	$T_1 \geq 0$	$T_2 \geq 2$	$T_3 \geq 0$	$P(T_1 \geq 0 \cap T_2 \geq 2 \cap T_3 \geq 0)$	$P(X_1 \geq 0 \cap X_2 \geq 2 \cap X_3 \geq 0)$	$P[(T_1 \geq 0 \cap T_2 \geq 2 \cap T_3 \geq 0) \cap (X_1 \geq 0 \cap X_2 \geq 2 \cap X_3 \geq 0)]$	
		AND-1-READ	$T_1 \geq 0$	$T_2 \geq 0$	$T_3 \geq 2$	$P(T_1 \geq 0 \cap T_2 \geq 0 \cap T_3 \geq 2)$	$P(X_1 \geq 0 \cap X_2 \geq 0 \cap X_3 \geq 2)$	$P[(T_1 \geq 0 \cap T_2 \geq 0 \cap T_3 \geq 2) \cap (X_1 \geq 0 \cap X_2 \geq 0 \cap X_3 \geq 2)]$	
3	AND	AND-2	$T_1 \geq 2$	$T_2 \geq 2$	$T_3 \geq 2$	$P(T_1 \geq 2 \cap T_2 \geq 2 \cap T_3 \geq 2)$	$P(X_1 \geq 2 \cap X_2 \geq 2 \cap X_3 \geq 2)$	$P[(T_1 \geq 2 \cap T_2 \geq 2 \cap T_3 \geq 2) \cap (X_1 \geq 2 \cap X_2 \geq 2 \cap X_3 \geq 2)]$	
4	AND	AND-3	$T_1 \geq 1.28$	$T_2 \geq 1.28$	$T_3 \geq 1.28$	$P(T_1 \geq 1.28 \cap T_2 \geq 1.28 \cap T_3 \geq 1.28)$	$P(X_1 \geq 1.28 \cap X_2 \geq 1.28 \cap X_3 \geq 1.28)$	$P[(T_1 \geq 1.28 \cap T_2 \geq 1.28 \cap T_3 \geq 1.28) \cap (X_1 \geq 1.28 \cap X_2 \geq 1.28 \cap X_3 \geq 1.28)]$	
5	OR	OR	$T_1 \geq 2 \text{ or } T_2 \geq 2 \text{ or } T_3 \geq 2$			$P(T_1 \geq 2 \cup T_2 \geq 2 \cup T_3 \geq 2)$	$P(X_1 \geq 2 \cup X_2 \geq 2 \cup X_3 \geq 2)$	$P[(T_1 \geq 2 \cup T_2 \geq 2 \cup T_3 \geq 2) \cap (X_1 \geq 2 \cup X_2 \geq 2 \cup X_3 \geq 2)]$	
6	MEAN	MEAN-1	$T_1 \geq 2$	$T_2 \geq 2$	$T_3 \geq 2$	$P(T_1 \geq 2 \cap T_2 \geq 2 \cap T_3 \geq 2)$	$P(\frac{X_1+X_2+X_3}{3} \geq 2)$	$P[(T_1 \geq 2 \cap T_2 \geq 2 \cap T_3 \geq 2) \cap (\frac{X_1+X_2+X_3}{3} \geq 2)]$	
		MEAN-2	$T_1 \geq 2 \text{ or } T_2 \geq 2 \text{ or } T_3 \geq 2$			$P(T_1 \geq 2 \cup T_2 \geq 2 \cup T_3 \geq 2)$	$P(\frac{X_1+X_2+X_3}{3} \geq 2)$	$P[(T_1 \geq 2 \cup T_2 \geq 2 \cup T_3 \geq 2) \cap (\frac{X_1+X_2+X_3}{3} \geq 2)]$	
7	OR+AND	OR-AND-INT	$(T_2 \geq 2 \text{ or } T_3 \geq 2) \text{ and } T_1 \geq 2$			$P[(T_2 \geq 2 \cup T_3 \geq 2) \cap T_1 \geq 2]$	$P[(X_2 \geq 2 \cup X_3 \geq 2) \cap (X_1 \geq 2)]$	$P\{[(T_2 \geq 2 \cup T_3 \geq 2) \cap T_1 \geq 2] \cap [(X_2 \geq 2 \cup X_3 \geq 2) \cap (X_1 \geq 2)]\}$	
		OR-AND-MATH	$(T_1 \geq 2 \text{ or } T_2 \geq 2) \text{ and } T_3 \geq 2$			$P[(T_1 \geq 2 \cup T_2 \geq 2) \cap T_3 \geq 2]$	$P[(X_1 \geq 2 \cup X_2 \geq 2) \cap (X_3 \geq 2)]$	$P\{[(T_1 \geq 2 \cup T_2 \geq 2) \cap T_3 \geq 2] \cap [(X_1 \geq 2 \cup X_2 \geq 2) \cap (X_3 \geq 2)]\}$	
		OR-AND-READ	$(T_1 \geq 2 \text{ or } T_3 \geq 2) \text{ and } T_2 \geq 2$			$P[(T_1 \geq 2 \cup T_3 \geq 2) \cap T_2 \geq 2]$	$P[(X_1 \geq 2 \cup X_3 \geq 2) \cap (X_2 \geq 2)]$	$P\{[(T_1 \geq 2 \cup T_3 \geq 2) \cap T_2 \geq 2] \cap [(X_1 \geq 2 \cup X_3 \geq 2) \cap (X_2 \geq 2)]\}$	

Plan of Data Analysis for Research Question Two

The second research question is targeted at the IG group. Theoretically, the IG group identified by different methods should be different in the size and gifted characteristics. For example, among the seven methods presented in Table 10, using the conjunctive rule to combine a high criterion of gifted for all three measures (the AND-2-method) results in the smallest IG group, in which students are expected to be gifted in all three domain areas. Using the complementary rule to combine a high criterion of gifted for all three measures (the OR-method) results in the largest IG group, in which students are expected to be gifted in at least one of the three domain areas.

The size of the IG group is indicated by the percent of IG students within in the simulated population. The gifted characteristics of the IG groups are described by the relative PPR and sensitivity. The relative PPR is the relative percent of students within the IG group who are TG in one of the three domain areas. The relative PPR indicate what types of gifted students compose the IG group. The relative sensitivity is the relative percent of IG students within the students who are TG in one of the three domain areas. The relative sensitivity is the sensitivity of each method to a specific type of giftedness, and therefore it can be compared across methods. The relative sensitivity suggests whether a method of gifted identification favors or disfavors a certain category of gifted students. The detail of the relative PPR and sensitivity is summarized in Table 11.

Table 11

Statistical Indicators of the Gifted Characteristics of the TG, IG, and TPG Groups

Statistical Indicators	Description
P(TG)	The size of TG <i>The percentage of TG students in the entire population</i>
P(IG)	The size of IG

Statistical Indicators	Description
	<i>The percentage of IG students In the entire population</i>
$P(TG \cap IG)$	The size of TPG <i>The percentage of TPG students in the entire population</i>
$\frac{P(TG_1 \cap IG)}{P(IG)}$	The relative PPR for the TG in intelligence <i>The relative size of the intellectual TG students in the IG group</i>
$\frac{P(TG_2 \cap IG)}{P(IG)}$	The relative PPR for the TG in math <i>The relative size of the mathematical TG students in the IG group</i>
$\frac{P(TG_3 \cap IG)}{P(IG)}$	The relative PPR for the TG in reading <i>The relative size of the reading TG students in the IG group</i>
$\frac{P(TG_1 \cap IG)}{P(TG_1)}$	The relative sensitivity to the TG in intelligence <i>The relative size of the intellectual TPG students in the intellectual TG group</i>
$\frac{P(TG_2 \cap IG)}{P(TG_2)}$	The relative sensitivity to the TG in math <i>The relative size of the mathematical TPG students in the mathematical TG group</i>
$\frac{P(TG_3 \cap IG)}{P(TG_3)}$	The relative sensitivity to the TG in reading <i>The relative size of the intellectual TPG students in the intellectual TG group</i>

Note. TG_1 refers to the group of students who are TG in intelligence; TG_2 refers to the group of students who are TG in math; TG_3 refers to the group of students who are TG in reading; IG is the identified gifted group.

As mentioned earlier, the TG students (4.70% of the simulated population) who are TG in at least one of the three domain areas can be divided into seven mutually exclusive TG populations (shown in Figure 4). To better understand the gifted characteristics of the IG group, this study further examined the relative PPR and sensitivity for each of the seven mutually exclusive TG populations. Taking the gifted population TG7 (in Figure 4) as an example, the relative PPR of TG7 is the percent of students who are TG in all three domain areas within the IG group, and the relative sensitivity to TG7 is the percent of students in TG7 who are identified as gifted. The relative PPR of the seven TG populations indicates the composite gifted characteristics of the IG group. The relative sensitivity indicates how effectively a method identifies a certain TG population. On the other side, the relative sensitivity reveals how much of a TG population is misidentified by a method. Four methods are investigated in this analysis, which are the AND-2, AND-3, MEAN, and OR methods in Table 10.

Plan of Data Analysis for Research Question Three

The third research question is engaged in exploring a new method that can effectively address the measurement error issues in gifted identification. This study explored the new method under the Bayesian framework and examined whether using the posterior probability of being TG given multiple observed abilities could improve the PPR and sensitivities of gifted identification simultaneously.

To calculate posterior probability, both true and observed abilities are rounded to two decimal places. The posterior probability of true abilities ranging from $(t_i - 0.005)$ to $(t_i + 0.005)$ given observed ability ranging from $(x - 0.005)$ to $(x + 0.005)$ is written as:

$$P(t_i|x) = \frac{P(t_i, x)}{P(x)} \quad (11)$$

where $P(t_i)$ represents the frequency of true abilities ranging from $(t_i - 0.005)$ to $(t_i + 0.005)$, $P(x)$ represents the frequency of observed abilities ranging from $(x - 0.005)$ to $(x + 0.005)$, and $P(t_i, x)$ is the frequency of all the combinations of observed and true abilities within $(x \pm .005)$ and $(t_i \pm .005)$. Given all the observed ability within the range of $(x \pm 0.005)$, if there are k ($k \geq 0$) true abilities (t_1, \dots, t_k) within the range of $(t_i \pm .005)$ fulfill the criterion of gifted, the posterior probability of being TG given x is written as:

$$P(TG|x) = \frac{\sum_{i=1}^k P(x, t_i)}{P(x)}$$

The posterior probability of a true ability ranging from $(t_{1i} - 0.005)$ to $(t_{1i} + 0.005)$ given a pair of observed abilities in two domain areas within the range of $(x_1 \pm 0.005)$ and $(x_2 \pm 0.005)$ is written as:

$$P(t_{1i}|x_1, x_2) = \frac{P(t_{1i}, x_1, x_2)}{P(x_1, x_2)} \quad (12)$$

where t_{1i} and x_1 are true and observed abilities in the same domain area, x_2 are observed ability in another domain area. $P(t_{1i})$ represents the frequency of true abilities ranging from $(t_i - 0.005)$ to $(t_i + 0.005)$, $P(x_1, x_2)$ represents the frequency of all the combinations of observed abilities within $(x_1 \pm 0.005)$ and $(x_2 \pm 0.005)$, and $P(t_i, x_1, x_2)$ represents the frequency of all the combinations of true and observed abilities within $(t_i \pm 0.005)$, $(x_1 \pm 0.005)$ and $(x_2 \pm 0.005)$. Given all the combinations of observed abilities within $(x_1 \pm 0.005)$ and $(x_2 \pm 0.005)$, if there are k ($k \geq 0$) true abilities (t_{11}, \dots, t_{1k}) within the range of $(t_i \pm 0.005)$ fulfill the criterion of gifted, the posterior probability of being TG given x_1 and x_2 is written as:

$$P(TG|x_1, x_2) = \frac{\sum_{i=1}^k P(t_{1i}, x_1, x_2)}{P(x_1, x_2)}$$

The posterior probability of a pair of true ability within $(t_{1i} \pm 0.005)$ and $(t_{2i} \pm 0.005)$ given all the combinations of observed abilities within $(x_1 \pm 0.005)$ and $(x_2 \pm 0.005)$ is written as:

$$P(t_{1i}, t_{2i}|x_1, x_2) = \frac{P(t_{1i}, t_{2i}, x_1, x_2)}{P(x_1, x_2)} \quad (13)$$

where t_{1i} and x_1 are true and observed abilities in the same domain area, t_{2i} and x_2 are true and observed abilities in the same domain area. $P(t_{1i}, t_{2i})$ represents the frequency of all the combination of true abilities within $(t_{1i} \pm 0.005)$ and $(t_{2i} \pm 0.005)$, $P(x_1, x_2)$ represents the frequency of all the combination of observed abilities within $(x_1 \pm 0.005)$ and $(x_2 \pm 0.005)$, and $P(t_{1i}, t_{2i}, x_1, x_2)$ represents the frequency of all the combination of true and observed abilities within $(t_{1i} \pm 0.005)$, $(t_{2i} \pm 0.005)$, $(x_1 \pm 0.005)$ and $(x_2 \pm 0.005)$. Given all the combinations of observed abilities within $(x_1 \pm 0.005)$ and $(x_2 \pm 0.005)$, if there are k ($k \geq 0$) combinations of true abilities $(t_{11}, t_{21}), \dots, (t_{1k}, t_{2k})$ fulfill the criteria of gifted, the posterior probability of being TG given x_1 and x_2 is written as:

$$P(TG|x_1, x_2) = \frac{\sum_{i=1}^k P(t_{1i}, t_{2i}, x_1, x_2)}{P(x_1, x_2)}$$

Given the complexity of calculating the posterior probability of being TG given multiple observed measures, this study only examines the posterior probability of being TG given one and two observed measures. Considering a practical example of gifted identification that uses mathematical or reading abilities to select candidates for gifted services and uses the intelligence test as the conclusive test, this study thoroughly investigated the performance of gifted identification using the posterior probability method on identifying four types of TG students: (1) students who are TG in math, (2) students who are TG in reading, (3) students who are TG in math and intelligence, and (4) students who are TG in reading and intelligence.

Chapter 4

Results

Results for Research Question One

The first research question is focused on the difference in the performance of gifted identification using different rules to combine multiple measures. The PPR and sensitivity of the seven methods are summarized in Table 12 and plotted in Figure 5.

Except the MEAN method, the PPR and sensitivity of the other six methods are all greater than .5. The MEAN method is very sensitive to the TG students who are TG in all three domains, which can identify 94% of them; however, among the students identified as gifted by the MEAN method, only 39% of them are TG in all three domains. Among the students identified as gifted by the MEAN method, 98% of them are TG in at least one of the three domain areas; however, only 27% of the total students who are TG in at least one of the three domain are identified as gifted by the MEAN method.

Using the conjunctive rule to combine different criteria for different measures (the AND-1 method) and using a single measure perform similarly on identifying students who are gifted in a specific domain area. The PPR and sensitivity of the AND-1 method and the single measure method are both positively correlated with the reliability of observed abilities. Therefore, gifted identification based on observed abilities with higher reliabilities performs better in terms of both the PPR and sensitivity.

Using the conjunctive rule to combine a high standard of gifted ($C = 2.0$; the AND-2 method) for all three measures can identify 58% of the total students who are at or above the 97.5th percentile in all three domain areas, corresponding to 76% of the students identified as gifted by the AND-2 method. Using the conjunctive rule to combine a lower standard of gifted

($C = 1.28$; the AND-3 method) for all three measures can identify 77% of students who are at or above the 90th percentile in all three domain areas, corresponding to 84% of the students identified as gifted by the AND-3 method. The target TG population of the AND-2 rule is 0.53% of the entire population. Using the same criteria of gifted to identify gifted students actually only identify 0.41% of the entire population. The target TG population of the AND-3 rule is 3.76% of the entire population. Using the same criteria of gifted to identify gifted students identify 3.46% of the entire population. Therefore, using the conjunctive rule to combine the same criterion of gifted for all three measures may identify a smaller percent of students as gifted than the percent of the target TG population within the entire population. Increasing the criterion of gifted may result in larger difference in the percent of students identified as gifted and the percent of the target TG population within the entire population. Therefore, for the AND-2 or AND-3 method, the sensitivity, $\frac{P(TPG)}{P(TG)}$, may be always lower than the PPR, $\frac{P(TPG)}{P(IG)}$

Using the OR rule to combine a high standard of gifted ($C = 2.0$) for all three measures is aimed to identify 4.70% of the entire population who are TG in at least one of the three domain areas. However, using the OR method actually identify more students, $P(IG) = 5.03\%$, as gifted based on observed abilities. Therefore, for the OR method, the sensitivity, $\frac{P(TPG)}{P(TG)}$, may be always higher than the PPR, $\frac{P(TPG)}{P(IG)}$.

Using the OR rule to combine a high standard of gifted for two of the three measures (serving as the screening test) and then further using the AND rule to combine a high standard of gifted for the third measure (serving as the conclusive test) is aimed at the TG students who are gifted in at least two domain areas. Using the intelligence test as the conclusive test performs better than using the math test or the reading test as the conclusive test in terms of both the PPR

and sensitivity. Higher reliability of the observed abilities from the conclusive test results in the higher PPR and sensitivity.

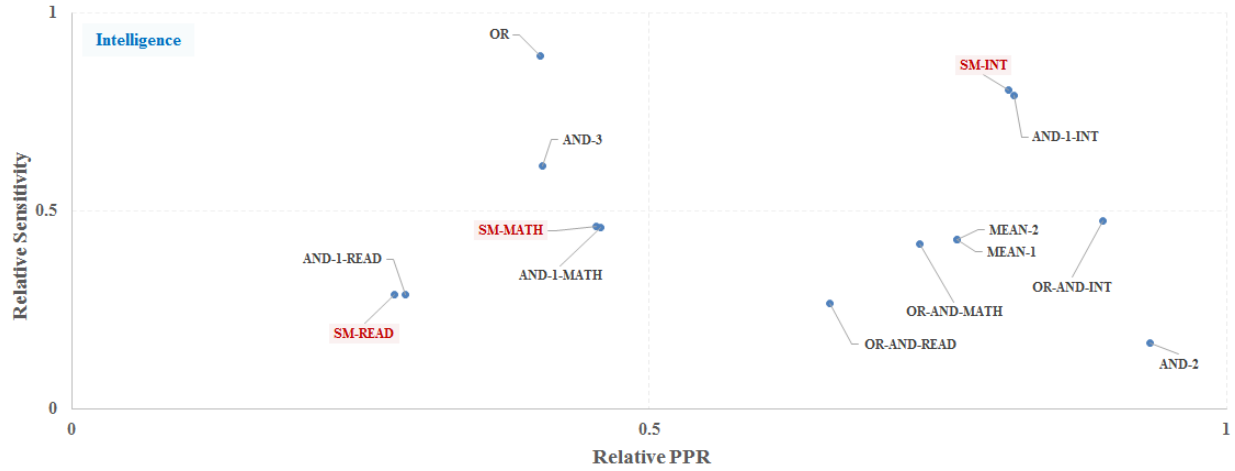


Figure 5. Plots of the PPR against sensitivity.

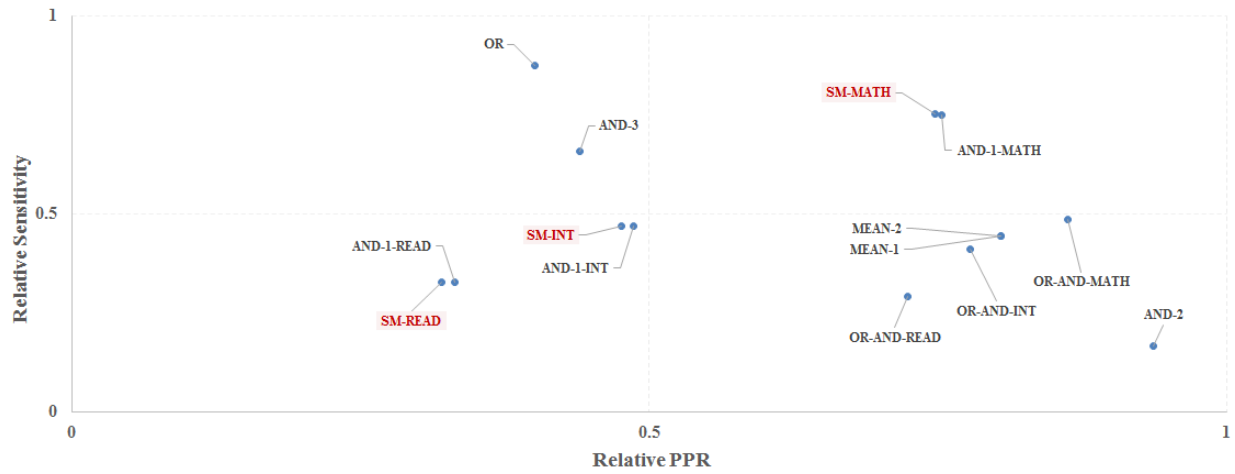
As mentioned earlier, different methods are aimed to identify different TG populations, and therefore the performance of gifted identification using different methods cannot be compared with each other. To compare the performance of gifted identification using different methods, this study calculated the relative PPR and sensitivity. The relative PPR and sensitivity indicate the performance of gifted identification using different methods on identifying students who are TG in each of the three domain areas. For the relative PPR and sensitivity, the TG populations are consistent across methods. The relative PPR is plotted against the relative sensitivity for the intellectually TG students, the mathematically TG students, and the reading TG student in Figure 6 (a), (b), (c), respectively. Considering the performance of gifted identification using the AND-2, AND-3, MEAN, and OR methods, these four methods should perform fairly across the three domain areas. The correlations among the true abilities and the reliabilities of the observed abilities in the three domain areas both affect the fairness of gifted identification using these four methods. The observed intelligence is most reliable ($r = .95$), the observed reading abilities are least reliable ($r = .80$). The true mathematically abilities are more

correlated with the true intelligence ($\rho_{12} = .85$) than the true reading abilities ($\rho_{13} = .70$) are and the true mathematically abilities are more correlated with the true reading abilities ($\rho_{23} = .75$) than the true intelligence is ($\rho_{13} = .70$). The MEAN and AND-3 methods are most sensitive to the TG students who are TG in math. The AND-2 method is more sensitive to the TG students who are TG in intelligence or math. The OR method is most sensitive to the TG students who are TG in intelligence and slightly less sensitive to the TG students who are TG in math. All four methods are least sensitive to the TG students who are TG in reading.

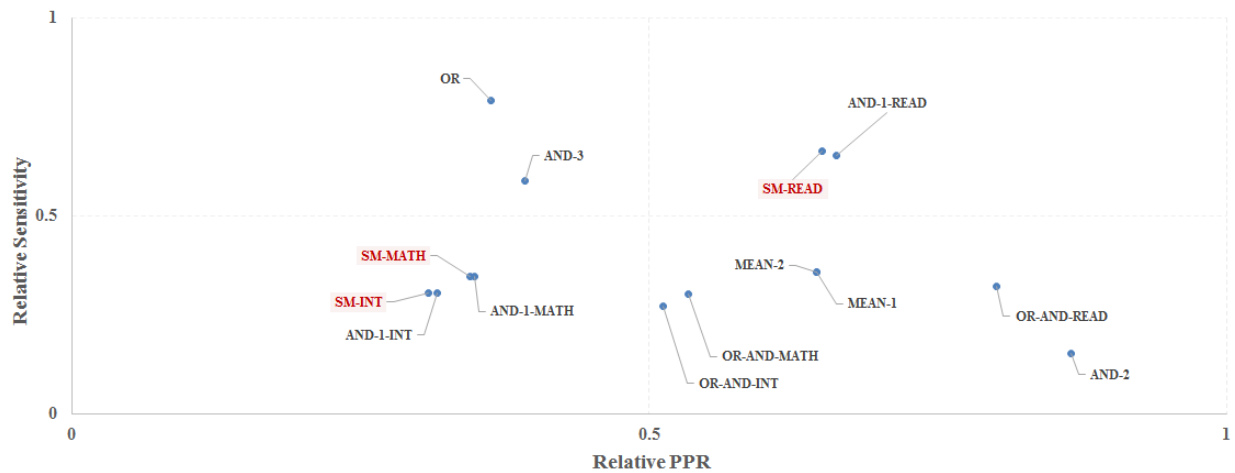
Considering the OR-AND method, gifted identification should perform fairly between the two domain areas measured by the screening tests. However, gifted identification using the intelligence test as the conclusive test is more sensitive to the TG students who are TG in math, gifted identification using the math test as the conclusive test is more sensitive to the TG students who are TG in intelligence, and gifted identification using the reading test as the conclusive test is more sensitive to the TG students who are TG in math. That is to say, the OR-AND method favors the TG students who are TG in the domain area that is more correlated with the domain area measured by the conclusive test.



(a) Plots of the relative PPR against relative sensitivity of the students TG in intelligence.



(b) Plots of the relative PPR against relative sensitivity of the students TG in math.



(c) Plots of the relative PPR against relative sensitivity of the students TG in reading.

Figure 6. Plots of the relative PPR against relative sensitivity

Table 12

The Performance of Gifted Identification Using Different Methods

Method Name	Criterion of Gifted			Percent of Different Groups			PPR	Sensitivity	Relative PPR for TG in			Relative Sensitivity to TG in		
									Intelligence	Math	Reading	Intelligence	Math	Reading
	Intelligence	Math	Reading	P(TG)	P(IG)	P(TPG)	$\frac{P(TPG)}{P(IG)}$	$\frac{P(TPG)}{P(TG)}$	$\frac{P(TG_1 \cap IG)}{P(IG)}$	$\frac{P(TG_2 \cap IG)}{P(IG)}$	$\frac{P(TG_3 \cap IG)}{P(IG)}$	$\frac{P(TG_1 \cap IG)}{P(TG_1)}$	$\frac{P(TG_2 \cap IG)}{P(TG_2)}$	$\frac{P(TG_3 \cap IG)}{P(TG_3)}$
Single Measure														
SM-INT	$T_1 \geq 2$	NA	NA	2.29%	2.28%	1.85%	0.81	0.81	81.1%	47.6%	30.9%	80.6%	47.0%	30.5%
SM-MATH	NA	$T_2 \geq 2$	NA	2.31%	2.33%	1.74%	0.75	0.75	45.4%	74.8%	34.5%	46.1%	75.4%	34.7%
SM-READ	NA	NA	$T_3 \geq 2$	2.31%	2.36%	1.53%	0.65	0.66	27.9%	32.0%	65.0%	28.7%	32.7%	66.4%
Multiple Measures														
AND-1-INT	$T_1 \geq 2$	$T_2 \geq 0$	$T_3 \geq 0$	2.26%	2.22%	1.81%	0.82	0.80	81.6%	48.6%	31.6%	79.2%	46.8%	30.5%
AND-1-MATH	$T_1 \geq 0$	$T_2 \geq 2$	$T_3 \geq 0$	2.30%	2.30%	1.73%	0.75	0.75	45.8%	75.3%	34.9%	45.9%	74.9%	34.7%
AND-1-READ	$T_1 \geq 0$	$T_2 \geq 0$	$T_3 \geq 2$	2.27%	2.28%	1.51%	0.66	0.67	28.9%	33.2%	66.2%	28.7%	32.7%	65.3%
AND-2	$T_1 \geq 2$	$T_2 \geq 2$	$T_3 \geq 2$	0.53%	0.41%	0.31%	0.76	0.58	93.4%	93.7%	86.5%	16.6%	16.5%	15.3%
AND-3	$T_1 \geq 1.28$	$T_2 \geq 1.28$	$T_3 \geq 1.28$	3.76%	3.46%	2.89%	0.84	0.77	40.8%	44.0%	39.2%	61.5%	65.9%	58.8%
OR	$T_1 \geq 2$ or $T_2 \geq 2$ or $T_3 \geq 2$			4.70%	5.03%	3.75%	0.75	0.80	40.6%	40.1%	36.3%	89.1%	87.4%	79.2%
MEAN-1	$T_1 \geq 2$	$T_2 \geq 2$	$T_3 \geq 2$	0.53%	1.28%	0.50%	0.39	0.94	76.7%	80.5%	64.5%	42.7%	44.5%	35.7%
MEAN-2	$T_1 \geq 2$ or $T_2 \geq 2$ or $T_3 \geq 2$			4.71%	1.28%	1.26%	0.98	0.27	76.7%	80.5%	64.5%	42.7%	44.5%	35.7%
OR-AND-INT	$(T_2 \geq 2$ or $T_3 \geq 2)$ and $T_1 \geq 2$			1.35%	1.22%	0.99%	0.81	0.73	89.3%	77.8%	51.2%	47.6%	41.2%	27.1%
OR-AND-MATH	$(T_1 \geq 2$ or $T_3 \geq 2)$ and $T_2 \geq 2$			1.47%	1.30%	1.04%	0.80	0.71	73.4%	86.2%	53.4%	41.7%	48.6%	30.2%
OR-AND-READ	$(T_1 \geq 2$ or $T_2 \geq 2)$ and $T_3 \geq 2$			1.06%	0.93%	0.68%	0.73	0.64	65.6%	72.4%	80.1%	26.7%	29.2%	32.3%

Note. TG stands for “Truly Gifted” students whose true abilities fulfill the criterion of gifted; IG stands for “Identified Gifted” students whose observed abilities fulfill the criterion of gifted; TPG stands for “True Positive Gifted” students whose true and observed abilities both fulfill the criterion of gifted; PPR stands for positive predictive rate.

Results for Research Question Two

The second research question is aimed to understand the target TG population and corresponding IG groups of each method of gifted identification. As mentioned earlier, the total TG students who are TG in at least one of the three domain areas can be divided into seven mutually exclusive TG populations (shown in Figure 4): three TG populations (TG1, TG2, and TG3) are gifted in only one of the three domain areas, three TG populations (TG4, TG5, and TG6) are gifted in only two of the three domain areas but not in the third domain area, and one TG population (TG7) are gifted in all three domain areas. Four methods of gifted identification are examined for the second research question, which are the AND-2, AND-3, MEAN, and OR methods. The effectiveness of identifying the seven gifted groups is summarized in Table 13 and the false negative rates of the seven gifted groups are provided in Table 14.

The target TG population of the AND-2 method is TG7 in Figure 4, which is 0.53% of the entire population. The percent of the IG group of the AND-2 method is 0.41% of the entire population. All the students identified as gifted by the AND-2 method are TG in at least one of the three domain areas. However, the probability of being identified as gifted by the AND-2 method is lower than 1% for students who are TG in only one domain area (the TG1, TG2, and TG3) and lower than 10% for students who are TG in only two domain areas (the TG4, TG5, and TG6). The probability of being identified as gifted by the AND-2 method is only 59.2% for students who are TG in all three domain areas. That is to say, almost 40% of the TG7 students are misidentified by the AND-2 method.

For research question two, this study set the total TG students who are TG ($C = 2.0$) in at least one of the three domain areas as the target TG population for the AND-3 method to test whether reducing the criterion of gifted for identification could improve the sensitivity to the TG

groups. Therefore, the percent of the target TG group is 4.70%. The percent of the IG group of the AND-3 method is 3.46%. Almost all the TG7 students (99.6%) are identified as gifted by the AND-3 method, 72.6% of the students who are TG in only two domain areas are identified as gifted by the AND-3 method, and 34.3% of the students who are TG in only one domain areas are identified as gifted by the AND-3 method. Among the students who are TG in only one domain areas, the AND-3 method is more sensitivity to the students who are TG in math. In total, 51% of the target TG population is identified as gifted by the AND-3 method.

Again, using the total TG students who are gifted in at least one of the three domain areas as the target TG population, the MEAN method identifies 94.3% of the TG7 students who are gifted in all three domain areas. The MEAN method identifies 50% of the students who are gifted in only two domain areas and 6% of the students who are gifted in only one domain areas. The MEAN method implies an idea that lower performance on one measure can be compensated by higher performance on the other measures. However, high performance on only one measure cannot compensate lower performance on the other two measures, which greatly reduce the probability of being identified as gifted by the MEAN method.

The target population of the OR method is the total TG students who are TG in at least one of the three domain areas. Almost all the students (94.5%) who are TG in at least two of the three domain areas and 70% of the students who are TG in only one of the three domain areas are identified as gifted by the OR method. However, 22.6% of the students TG in intelligence, 28.7% of the students TG in math, and 35.2% of the students TG in reading are misidentified by the OR method. Given that the TG students who are TG in reading only (TG3) is the largest TG group (1.25% of the entire population and 26.5% of the total TG students) among the seven

mutually exclusive TG groups, almost half (49.3%) of the misidentified TG students by the OR method are TG students who are TG in reading only.

Table 13

The Effectiveness of Identifying the Seven Gifted Groups

TG Groups	Criteria of Gifted for			^a Percent $P(TG_i)$	^b Relative Percent $\frac{P(TG_i)}{\sum_{i=1}^7 P(TG_i)}$	Relative PPR				Relative Sensitivity			
	Intelligence	Math	Reading			AND-2	AND-3	MEAN	OR	AND-2	AND-3	MEAN	OR
TG1	$T_1 \geq 2$	$T_2 \leq 2$	$T_3 \leq 2$	0.940%	20.0%	1.1%	9.2%	5.1%	14.5%	0.5%	33.7%	6.9%	77.4%
TG2	$T_1 \leq 2$	$T_2 \geq 2$	$T_3 \leq 2$	0.835%	17.8%	1.2%	9.7%	5.1%	11.8%	0.6%	40.0%	7.9%	71.3%
TG3	$T_1 \leq 2$	$T_2 \leq 2$	$T_3 \geq 2$	1.248%	26.5%	0.8%	11.2%	4.5%	16.1%	0.3%	31.1%	4.7%	64.8%
TG4	$T_1 \geq 2$	$T_2 \geq 2$	$T_3 \leq 2$	0.618%	13.1%	11.0%	11.3%	23.9%	11.8%	7.2%	63.4%	49.5%	96.3%
TG5	$T_1 \geq 2$	$T_2 \leq 2$	$T_3 \geq 2$	0.205%	4.4%	4.2%	5.0%	8.6%	3.8%	8.3%	85.0%	53.5%	93.9%
TG6	$T_1 \leq 2$	$T_2 \geq 2$	$T_3 \geq 2$	0.326%	6.9%	4.3%	7.7%	12.3%	5.9%	5.4%	82.1%	48.2%	91.8%
TG7	$T_1 \geq 2$	$T_2 \geq 2$	$T_3 \geq 2$	0.531%	11.3%	77.2%	15.3%	39.1%	10.5%	59.2%	99.6%	94.3%	99.5%
$\sum_{i=1}^7 P(TG_i) =$				4.71%	$\sum_{i=1}^7 P(TG_i) \times \text{Relative Sensitivity}_i =$				0.41%	2.40%	1.26%	3.75%	

Note. ^aPercent is the percent of each TG group within the one million simulated students. ^bRelative Percent is the relative percent of each TG group within the total TG group ($\frac{P(G_i)}{\sum_{i=1}^7 P(G_i)}$). T_1 , T_2 , and T_3 refer to true abilities in intelligence, math, and reading. The AND-2 method uses the conjunctive rule to combine a criterion of 2.0 for all three measures; the AND-3 method uses the conjunctive rule to combine a criterion of 1.28 for all three measures; the MEAN method uses the criterion of 2.0 for the mean of the observed intelligence, math, and reading abilities; and the OR method uses the complementary rule to combine a criterion of 2.0 for any of the three measures.

Table 14

The False Negative Rate of the Seven Gifted Groups

Groups with different Gifted Characteristics				Frequency												False Negative Rate			
				TG	True Positive Gifted				False Negative Gifted										
Name	Intelligence	Math	Reading		AND-2	AND-3	MEAN	OR	AND-2	AND-3	MEAN	OR	AND-2	AND-3	MEAN	OR			
TG1	$T_1 \geq 2$	$T_2 \leq 2$	$T_3 \leq 2$	9398	43	3170	646	7278	9355	6228	8752	2120	99.5%	66.3%	93.1%	22.6%			
TG2	$T_1 \leq 2$	$T_2 \geq 2$	$T_3 \leq 2$	8351	50	3341	656	5952	8301	5010	7695	2399	99.4%	60.0%	92.1%	28.7%			
TG3	$T_1 \leq 2$	$T_2 \leq 2$	$T_3 \geq 2$	12476	34	3875	580	8089	12442	8601	11896	4387	99.7%	68.9%	95.4%	35.2%			
TG4	$T_1 \geq 2$	$T_2 \geq 2$	$T_3 \leq 2$	6181	446	3920	3059	5953	5735	2261	3122	228	92.8%	36.6%	50.5%	3.7%			
TG5	$T_1 \geq 2$	$T_2 \leq 2$	$T_3 \geq 2$	2046	170	1738	1094	1921	1876	308	952	125	91.7%	15.1%	46.5%	6.1%			
TG6	$T_1 \leq 2$	$T_2 \geq 2$	$T_3 \geq 2$	3256	176	2674	1569	2988	3080	582	1687	268	94.6%	17.9%	51.8%	8.2%			
TG7	$T_1 \geq 2$	$T_2 \geq 2$	$T_3 \geq 2$	5307	3142	5284	5003	5281	2165	23	304	26	40.8%	0.4%	5.7%	0.5%			
SUM				47015	4061	24002	12607	37462	42954	23013	34408	9553							

Note. TG1, TG2, and TG3 are the three groups of students who are TG in intelligence, math, and reading, respectively; TG4 is the group of students who are TG in intelligence and math but not in reading; TG5 is the group of students who are TG in intelligence and reading but not in math; TG6 is the group of students who are TG in math and reading but not in intelligence; TG7 is the group of students who are TG in intelligence, math, and reading.

Results for Research Question Three

The third goal of this study is to explore new methods of gifted identification that can effectively address the measurement error issues and then improve the performance of gifted identification in terms of the PPR and sensitivity at the same time. Given this purpose, this study thoroughly investigated the performance of gifted identification based on the posterior probability of being TG given single and multiple observed abilities. To be specific, the target gifted populations of the third question are students who are TG in math ($T_2 \geq 2.0$), students who are TG in reading ($T_2 \geq 2.0$), students who are TG in math and intelligence ($T_2 \geq 2.0$ & $T_1 \geq 2.0$), and students who are TG in reading and intelligence ($T_3 \geq 2.0$ & $T_1 \geq 2.0$). For comparison, the performance of gifted identification based on observed abilities (the SM-MATH, SM-READ, and AND-2 methods in Table 10) is used as a reference, which is compared with the performance of identifying students who are TG in math or reading based on the posterior probability of being TG given the observed abilities in math or reading (the single-measure posterior probability method) and the performance of identifying the four types of TG students based on the posterior probability of being TG given observed intelligence and the observed abilities in math or reading (the multi-measure posterior probability method).

The performance of identifying students TG in math using different methods is summarized in Table 15. Using the single measure method (SM-MATH) can identify 75% of the students TG in math and the students TPG in math is also about 75% of the IG group. Given a criterion of 0.40 for the posterior probability of being TG in math given the observed mathematical abilities, $P(T_2 \geq 2.0 | X_2)$, the single-measure posterior probability method performs almost the same as the SM-MATH method does. Increasing the criterion of gifted for $P(T_2 \geq 2.0 | X_2)$ results in an increase in the PPR but an decrease in sensitivity. Actually, the single-measure

posterior probability method cannot surpass the performance of the SM-MATH method in terms of the PPR and sensitivity at the same time. Therefore, this study further tested whether gifted identification based on the posterior probability of being TG in math given observed abilities in math and observed intelligence, $P(T_2 \geq 2 \mid X_2, X_1)$, could perform better in terms of the PPR and sensitivity simultaneously. Results reveal that, for anyone of the four criteria (0.30, 0.40, 0.50, and 0.60) for $P(T_2 \geq 2 \mid X_2, X_1)$, the multi-measure posterior probability method performs better than the SM-MATH method does. Given a criterion of 0.6 for $P(T_2 \geq 2 \mid X_2, X_1)$, the multi-measure posterior probability method identifies the same amount of students TG in math (75%) as the SM-MATH method does; however, the PPR (0.92) of the multi-measure posterior probability method is much higher than the PPR of the SM-MATH method (0.75). Given a criterion of 0.30 for $P(T_2 \geq 2 \mid X_2, X_1)$, the PPR of the multi-measure posterior probability method is 0.76, which is almost the same as the PPR of the SM-MATH method; however, 92% of students who are TG in math can be successfully identified as gifted by the multi-measure posterior probability method. Further, for a criterion of 0.50, the multi-measure posterior probability method performs better than the SM-MATH method in terms of the PPR (0.84) and sensitivity (0.85) simultaneously.

Table 15

Performance of Identifying Students TG in Math ($T_2 \geq 2.0$)

Measure	Method		Sensitivity	PPR
Single	Using Observed Measures	$X_2 \geq 2$	0.75	0.75
Measure	Using Posterior Probability	$P(T_2 \geq 2 \mid X_2) \geq .05$	0.97	0.45
		$P(T_2 \geq 2 \mid X_2) \geq .10$	0.95	0.51
		$P(T_2 \geq 2 \mid X_2) \geq .20$	0.90	0.61
		$P(T_2 \geq 2 \mid X_2) \geq .30$	0.85	0.66
		$P(T_2 \geq 2 \mid X_2) \geq .40$	0.77	0.74
		$P(T_2 \geq 2 \mid X_2) \geq .50$	0.68	0.80
		$P(T_2 \geq 2 \mid X_2) \geq .60$	0.62	0.84
		$P(T_2 \geq 2 \mid X_2) \geq .70$	0.50	0.90

Measure	Method		Sensitivity	PPR
Multiple Measures	Using Posterior Probability	$P(T_2 \geq 2 \mid X_2) \geq .80$	0.44	0.92
		$P(T_2 \geq 2 \mid X_2) \geq .90$	0.31	0.96
		$P(T_2 \geq 2 \mid X_2, X_1) \geq .05$	1.00	0.59
		$P(T_2 \geq 2 \mid X_2, X_1) \geq .10$	0.99	0.61
		$P(T_2 \geq 2 \mid X_2, X_1) \geq .20$	0.96	0.68
		$P(T_2 \geq 2 \mid X_2, X_1) \geq .30$	0.92	0.76
		$P(T_2 \geq 2 \mid X_2, X_1) \geq .40$	0.87	0.81
		$P(T_2 \geq 2 \mid X_2, X_1) \geq .50$	0.85	0.84
		$P(T_2 \geq 2 \mid X_2, X_1) \geq .60$	0.75	0.92
		$P(T_2 \geq 2 \mid X_2, X_1) \geq .70$	0.67	0.97
		$P(T_2 \geq 2 \mid X_2, X_1) \geq .80$	0.62	0.99
		$P(T_2 \geq 2 \mid X_2, X_1) \geq .90$	0.58	1.00

Note. T_2 refers to true abilities in math; X_1 refers to observed intelligence; X_2 refers to observed abilities in math. PPR stands for the positive predictive rate.

The observed intelligence and math abilities of the true positive gifted (TPG) students and the false positive gifted (FPG) students identified by the multi-measure posterior probability method are plotted in Figure 7 and Figure 8. The correlation between observed intelligence and observed abilities in math is .30 within the TGP students and .14 within the FPT students. Both correlations are positive and weak but statistically significantly different from zero. For a criterion of .95 for $P(T_2 \geq 2 \mid X_2, X_1)$, the multi-measure posterior probability method identifies 58.4% of the students TG in math and almost all the IG students are TG in math. In Figure 7, within the IG group identified using the criterion of 0.95 for $P(T_2 \geq 2 \mid X_2, X_1)$, the students whose observed abilities lower than 2.0 (dots on the left side of the vertical line at 2.0) will be misidentified by the SM-MATH method. Further, comparing with the performance of gifted identification using the three rules to combine observed intelligence and abilities in math, the students whose means of the observed intelligence and abilities in math (dots under the diagonal line) will be misidentified by the MEAN method, the students whose observed intelligence and abilities in math are both lower than 2.0 (dots in the left bottom square) will be misidentified by the OR method, and only students whose observed intelligence and abilities in math are both

greater than 2.0 (dots in the upper right square) will be identified as gifted by the AND-2 method.

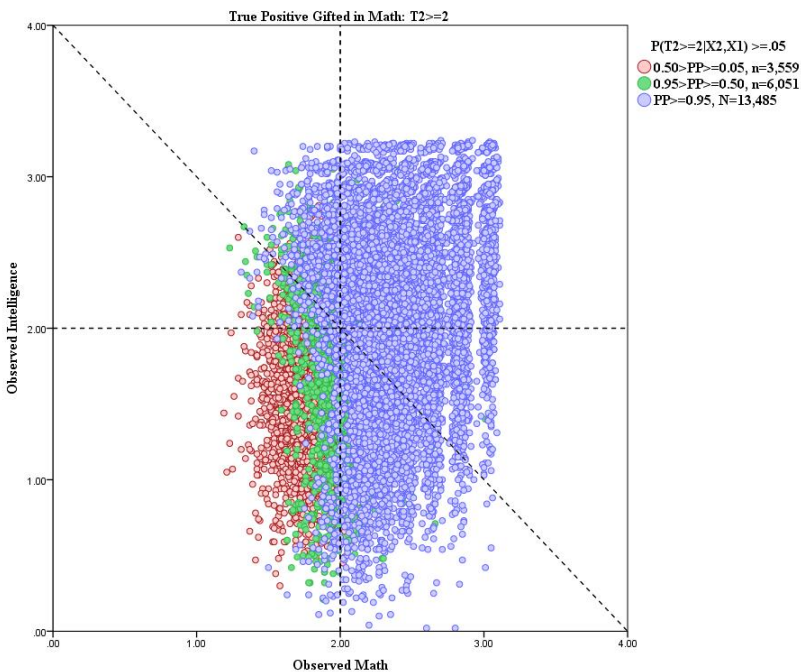


Figure 7. Students who are true positive gifted in math identified by the multi-measure posterior probability method.

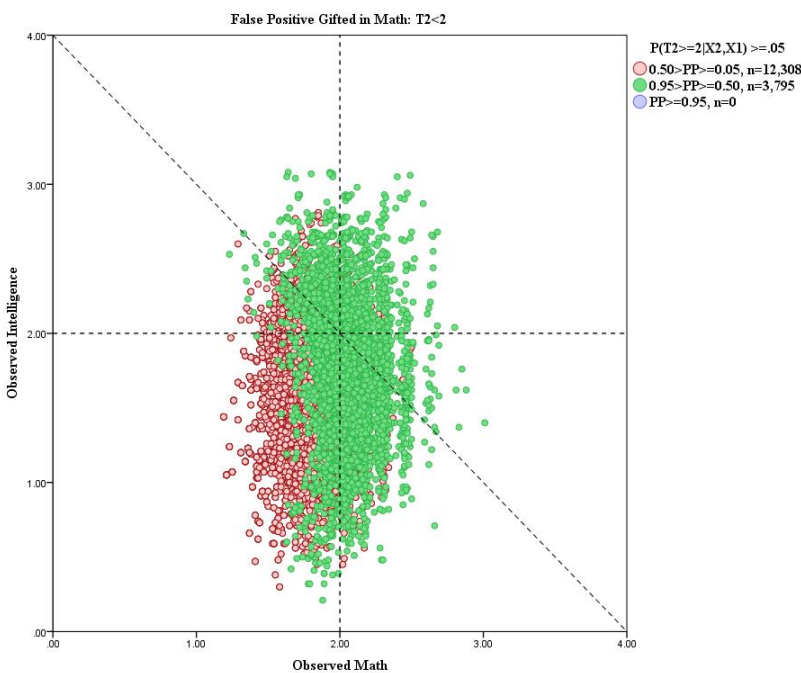


Figure 8. Students who are false positive gifted in math identified by the multi-measure posterior probability method.

The performance of identifying students TG in reading using different methods is summarized in Table 16. Using the single measure method (SM-READ) can identify 66% of the students TG in reading and the students TPG in reading is about 65% of the IG group. Given a criterion of 0.40 for the posterior probability of being TG in reading given the observed reading abilities, $P(T_3 \geq 2.0 | X_3)$, the single-measure posterior probability method performs almost the same as the SM-READ method does. Increasing the criterion of gifted for $P(T_3 \geq 2.0 | X_3)$ results in an increase in the PPR but an decrease in sensitivity. Again, the single-measure posterior probability method cannot surpass the performance of the SM-READ method in terms of the PPR and sensitivity at the same time. Therefore, this study further tested whether gifted identification based on the posterior probability of being TG in reading given observed reading abilities and observed intelligence, $P(T_3 \geq 2 | X_3, X_1)$, could perform better in terms of the PPR and sensitivity simultaneously. Results reveal that, for anyone of the three criteria (0.30, 0.40, and 0.50) for $P(T_3 \geq 2 | X_3, X_1)$, the multi-measure posterior probability method performs better than the SM-READ method does. Given a criterion of 0.5 or 0.4 for $P(T_3 \geq 2 | X_3, X_1)$, the multi-measure posterior probability method identifies more students TG in reading (75% or 78%) than the SM-READ method does; meanwhile, the PPR (0.78 or 0.75) of the multi-measure posterior probability method is also higher than the PPR of the SM-READ method (0.65). Given a criterion of 0.30 for $P(T_3 \geq 2 | X_3, X_1)$, the PPR of the multi-measure posterior probability method is 0.68, which is almost the same as the PPR of the SM-READ method; however, 85% of students who are TG in math can be successfully identified as gifted by the multi-measure posterior probability method.

Table 16

Performance of Identifying Students TG in Reading ($T_3 \geq 2.0$)

Measure	Method	Sensitivity	PPR
Single Measure	Using Observed Measures	$X_3 \geq 2$	0.66
	Using Posterior Probability	$P(T_3 \geq 2 X_3) \geq .05$	0.95
		$P(T_3 \geq 2 X_3) \geq .10$	0.95
		$P(T_3 \geq 2 X_3) \geq .20$	0.85
		$P(T_3 \geq 2 X_3) \geq .30$	0.67
		$P(T_3 \geq 2 X_3) \geq .40$	0.66
		$P(T_3 \geq 2 X_3) \geq .50$	0.54
		$P(T_3 \geq 2 X_3) \geq .60$	0.41
		$P(T_3 \geq 2 X_3) \geq .70$	0.40
		$P(T_3 \geq 2 X_3) \geq .80$	0.16
		$P(T_3 \geq 2 X_3) \geq .90$	0.09
Multiple Measures	Using Posterior Probability	$P(T_3 \geq 2 X_3, X_1) \geq .05$	1.00
		$P(T_3 \geq 2 X_3, X_1) \geq .10$	0.98
		$P(T_3 \geq 2 X_3, X_1) \geq .20$	0.93
		$P(T_3 \geq 2 X_3, X_1) \geq .30$	0.85
		$P(T_3 \geq 2 X_3, X_1) \geq .40$	0.78
		$P(T_3 \geq 2 X_3, X_1) \geq .50$	0.75
		$P(T_3 \geq 2 X_3, X_1) \geq .60$	0.61
		$P(T_3 \geq 2 X_3, X_1) \geq .70$	0.52
		$P(T_3 \geq 2 X_3, X_1) \geq .80$	0.47
		$P(T_3 \geq 2 X_3, X_1) \geq .90$	0.42

Note. T_3 refers to true abilities in reading; X_1 refers to observed intelligence; X_3 refers to observed abilities in reading. PPR stands for the positive predictive rate.

The observed intelligence and reading abilities of the true positive gifted students and the false positive gifted students identified by multi-measure posterior probability method are plotted in Figure 9 and Figure 10. The correlation between observed intelligence and observed abilities in reading is .15 within the TGP students and .09 within the FPT students. Both correlations are positive and very weak but statistically significantly different from zero. For a criterion of .95 for $P(T_3 \geq 2 | X_3, X_1)$, the multi-measure posterior probability method identifies 42.3% of the students TG in reading and almost all the IG students are TG in reading. In Figure 9, within the IG group identified using the criterion of 0.95 for $P(T_3 \geq 2 | X_3, X_1)$, the students whose observed abilities lower than 2.0 (dots on the left side of the vertical line at 2.0) will be misidentified by

the SM-READ method. Further, comparing with the performance of gifted identification using the three rules to combine observed intelligence and abilities in reading, the students whose means of the observed intelligence and abilities in reading (dots under the diagonal line) will be misidentified by the MEAN method, the students whose observed intelligence and abilities in reading are both lower than 2.0 (dots in the bottom two squares in the middle) will be misidentified by the OR method, and only students whose observed intelligence and abilities in math are both greater than 2.0 (dots in the upper right square) will be identified as gifted by the AND method. The range of observed intelligence of the TPG student in reading is wider than the range of observed intelligence of the TPG students in math. A small amount of TPG students in reading even have observed intelligence lower than the population mean ($\mu = 0$). Further, the multi-measures posterior probability method performs better on identifying students TG in math than on identifying students TG in reading. Given that the correlation between abilities in reading and intelligence is lower than the correlation between abilities in math and intelligence and the reliability of observed abilities in reading is lower than the reliability of observed abilities in math, better performance of the multi-measure posterior probability method on identifying students TG in math may suggest that the correlation between multiple measures and the reliability of observed measures are two factors that potentially affect the performance of the multi-measures posterior probability method.

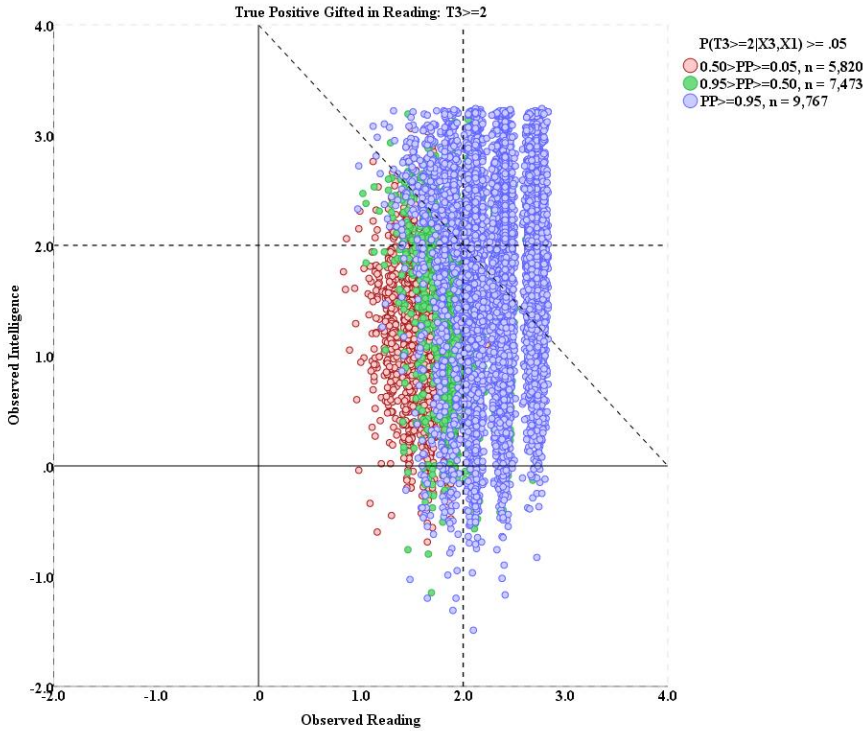


Figure 9. Students who are true positive gifted in reading identified by the multi-measure posterior probability method.

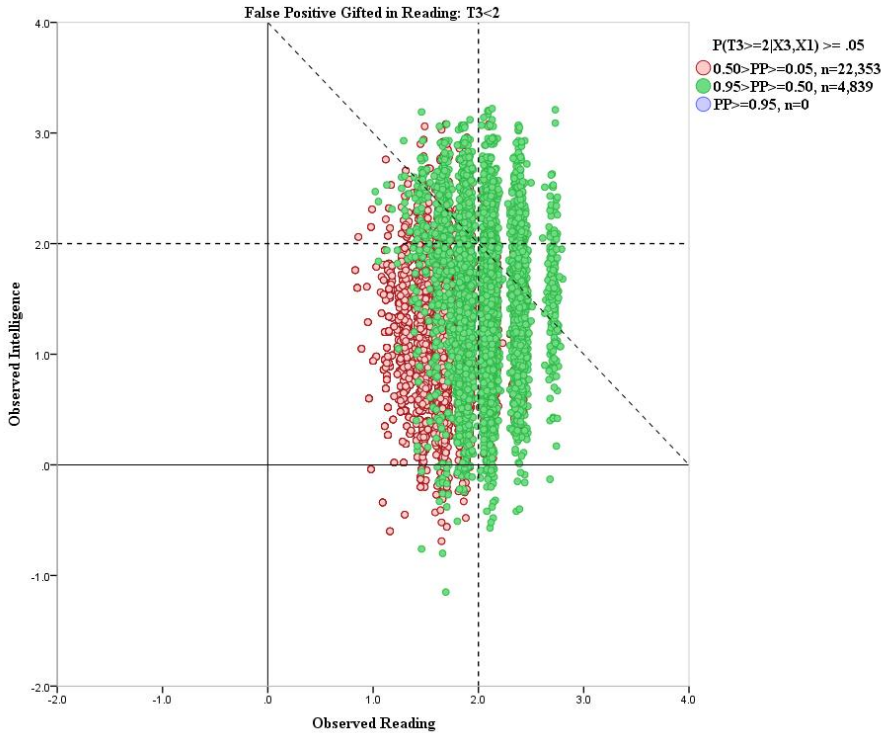


Figure 10. Students who are false positive gifted in math identified by the multi-measure posterior probability method.

The performance of identifying students TG in math and intelligence using different methods is summarized in Table 17. Compared with the performance of gifted identification based on observed intelligence and abilities in math combined by the conjunctive rule (the AND method), the gifted identification based on the posterior probability of being TG in math and intelligence given observed intelligence and abilities in math, $P(T_2 \geq 2 \ \& \ T_1 \geq 2 \mid X_2, X_1)$; the multi-measure posterior probability method can perform better in terms of the PPR and sensitivity at the same time. For anyone of the three criteria of gifted (0.40, 0.50, and 0.60) for $P(T_2 \geq 2 \ \& \ T_1 \geq 2 \mid X_2, X_1)$, the multi-measure posterior probability method performs better than the AND method does. Given a sensitivity of 0.70, the PPR of the multi-measure posterior probability method can be high up to 0.97. That is to say, almost every student identified by the multi-measure posterior method is TG in both math and intelligence. Given a sensitivity of 0.95, the PPR decreases from 0.97 to 0.78. In brief, a higher criterion for $P(T_2 \geq 2 \ \& \ T_1 \geq 2 \mid X_2, X_1)$ results in a higher PPR but a smaller IG group; a lower criterion for $P(T_2 \geq 2 \ \& \ T_1 \geq 2 \mid X_2, X_1)$ results in a higher sensitivity and a larger IG group.

Table 17

Performance of Identifying Students TG in Math and Intelligence ($T_2 \geq 2.0 \ \& \ T_1 \geq 2.0$)

Method		Sensitivity	PPR
Using Observed Measures	$X_2 \geq 2 \ \& \ X_1 \geq 2$	0.72	0.82
Using Posterior Probability	$P(T_2 \geq 2 \ \& \ T_1 \geq 2 \mid X_2, X_1) \geq .05$	1.00	0.67
	$P(T_2 \geq 2 \ \& \ T_1 \geq 2 \mid X_2, X_1) \geq .10$	1.00	0.67
	$P(T_2 \geq 2 \ \& \ T_1 \geq 2 \mid X_2, X_1) \geq .20$	0.98	0.72
	$P(T_2 \geq 2 \ \& \ T_1 \geq 2 \mid X_2, X_1) \geq .30$	0.95	0.78
	$P(T_2 \geq 2 \ \& \ T_1 \geq 2 \mid X_2, X_1) \geq .40$	0.90	0.83
	$P(T_2 \geq 2 \ \& \ T_1 \geq 2 \mid X_2, X_1) \geq .50$	0.88	0.85
	$P(T_2 \geq 2 \ \& \ T_1 \geq 2 \mid X_2, X_1) \geq .60$	0.79	0.93
	$P(T_2 \geq 2 \ \& \ T_1 \geq 2 \mid X_2, X_1) \geq .70$	0.71	0.97
	$P(T_2 \geq 2 \ \& \ T_1 \geq 2 \mid X_2, X_1) \geq .80$	0.68	0.99
	$P(T_2 \geq 2 \ \& \ T_1 \geq 2 \mid X_2, X_1) \geq .90$	0.65	1.00

Note. T_1 refers to true intelligence, T_2 refers to true abilities in math; X_1 refers to observed intelligence; X_2 refers to observed abilities in math. PPR stands for the positive predictive rate.

The observed intelligence and math abilities of the true positive gifted students and the false positive gifted students identified by multi-measure posterior probability method are plotted in Figure 11 and Figure 12. The correlation between observed intelligence and observed abilities in math is .23 within the TGP students and -.16 within the FPT students. Both correlations are weak but statistically significantly different from zero. The negative correlation between observed intelligence and abilities in math suggests that FPT students whose observed intelligence is relatively higher may have relatively lower observed abilities in math, and vice versa. For a criterion of .95 for $P(T_2 \geq 2 \ \& \ T_1 \geq 2 \mid X_2, X_1)$, the multi-measure posterior probability method identifies 64.5% of the students TG in math and intelligence and almost all the IG students are TG in math and intelligence. In Figure 11, within the IG group identified using the criterion of 0.95 for $P(T_2 \geq 2 \ \& \ T_1 \geq 2 \mid X_2, X_1)$, only students whose observed intelligence and abilities in math are both greater than 2.0 (dots in the upper right square) will be identified as gifted by the AND method. Comparing with the performance of gifted identification using the other two rules to combine the observed intelligence and abilities in math, the students whose means of the observed intelligence and abilities in math (dots under the diagonal line) will be misidentified by the MEAN method, the students whose observed intelligence and abilities in math are both lower than 2.0 (dots in the bottom left square) will be misidentified by the OR method.

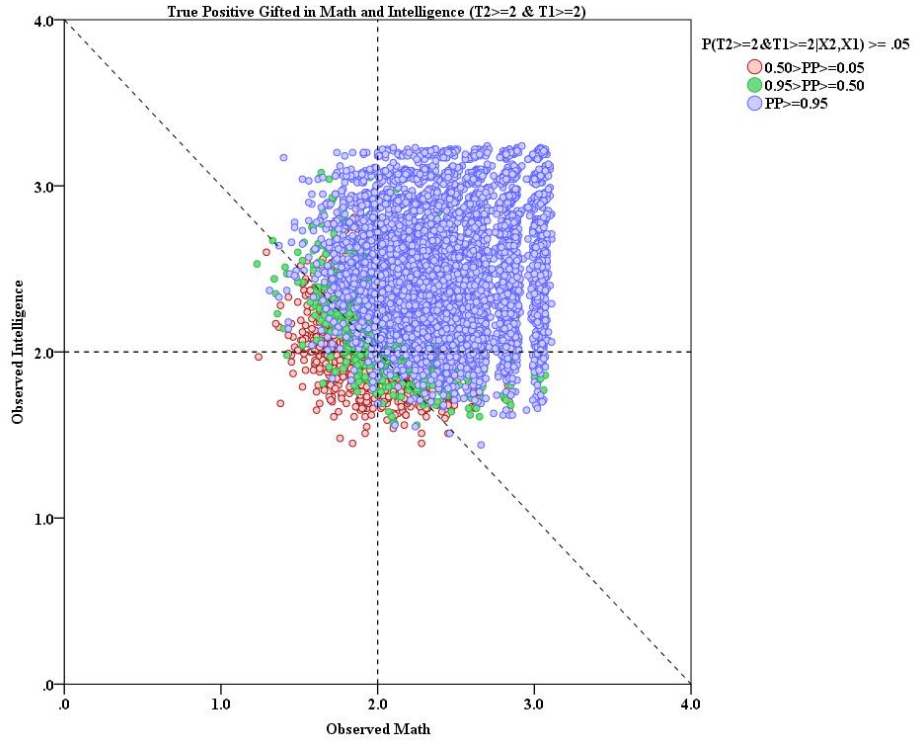


Figure 11. Students who are true positive gifted in math and intelligence identified by the multi-measure posterior probability method.

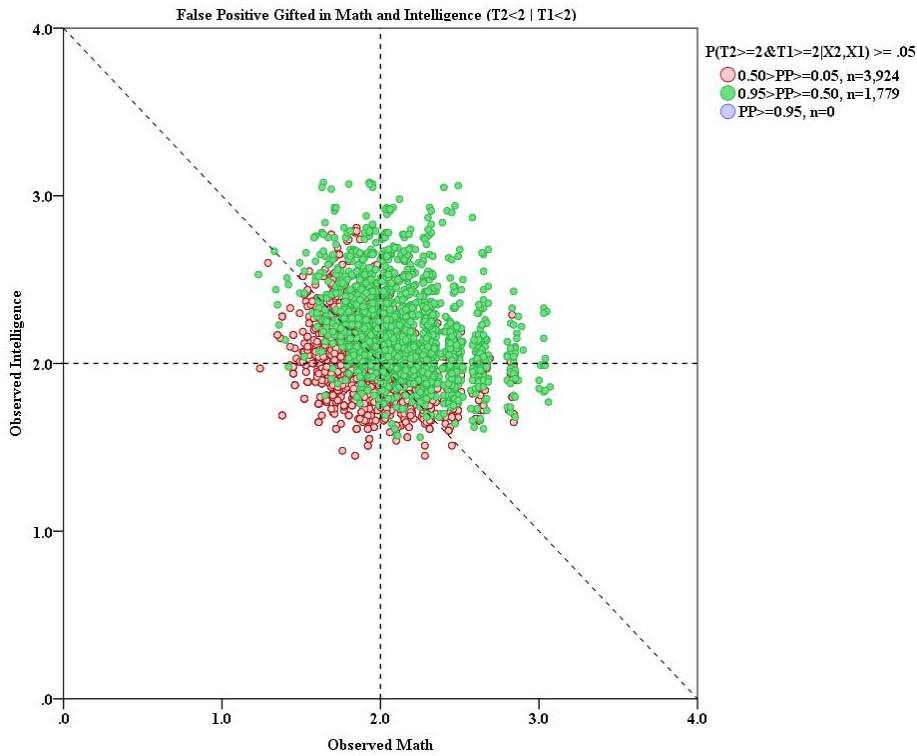


Figure 12. Students who are false positive gifted in math and intelligence identified by the multi-measure posterior probability method.

The performance of identifying students TG in reading and intelligence using different method is summarized in Table 18. Compared with the performance of gifted identification based on observed intelligence and abilities in reading combined by the conjunctive rule (the AND method), the gifted identification based on the posterior probability of being TG in reading and intelligence given observed intelligence and abilities in reading, $P(T_3 \geq 2 \ \& \ T_1 \geq 2 \mid X_3, X_1)$; the multi-measure posterior probability method can perform better in terms of the PPR and sensitivity at the same time. For anyone of the three criteria (0.40, 0.50, and 0.60) of gifted for $P(T_3 \geq 2 \ \& \ T_1 \geq 2 \mid X_3, X_1)$, the multi-measure posterior probability method performs better than the AND method does. Given a sensitivity of 0.70, the PPR of the multi-measure posterior probability method can be high up to 0.91. That is to say, most of the student identified by the multi-measure posterior method is TG in both reading and intelligence. Given a sensitivity of 0.90, the PPR decreases from 0.91 to 0.73. Again, a higher criterion for $P(T_3 \geq 2 \ \& \ T_1 \geq 2 \mid X_3, X_1)$ results in a higher PPR but a smaller IG group; a lower criterion for $P(T_3 \geq 2 \ \& \ T_1 \geq 2 \mid X_3, X_1)$ results in a higher sensitivity and a larger IG group.

Table 18

Performance of Identifying Students TG in Reading and Intelligence ($T_3 \geq 2.0 \ \& \ T_1 \geq 2.0$)

Method		Sensitivity	PPR
Using Observed Measures	$X_3 \geq 2 \ \& \ X_1 \geq 2$	0.63	0.74
Using Posterior Probability	$P(T_3 \geq 2 \ \& \ T_1 \geq 2 \mid X_3, X_1) \geq .05$	1.00	0.58
	$P(T_3 \geq 2 \ \& \ T_1 \geq 2 \mid X_3, X_1) \geq .10$	1.00	0.59
	$P(T_3 \geq 2 \ \& \ T_1 \geq 2 \mid X_3, X_1) \geq .20$	0.97	0.65
	$P(T_3 \geq 2 \ \& \ T_1 \geq 2 \mid X_3, X_1) \geq .30$	0.91	0.73
	$P(T_3 \geq 2 \ \& \ T_1 \geq 2 \mid X_3, X_1) \geq .40$	0.86	0.79
	$P(T_3 \geq 2 \ \& \ T_1 \geq 2 \mid X_3, X_1) \geq .50$	0.84	0.81
	$P(T_3 \geq 2 \ \& \ T_1 \geq 2 \mid X_3, X_1) \geq .60$	0.71	0.91
	$P(T_3 \geq 2 \ \& \ T_1 \geq 2 \mid X_3, X_1) \geq .70$	0.62	0.97
	$P(T_3 \geq 2 \ \& \ T_1 \geq 2 \mid X_3, X_1) \geq .80$	0.58	0.99
	$P(T_3 \geq 2 \ \& \ T_1 \geq 2 \mid X_3, X_1) \geq .90$	0.56	1.00

Note. T_1 refers to true intelligence, T_3 refers to true abilities in reading; X_1 refers to observed intelligence; X_3 refers to observed abilities in reading. PPR stands for the positive predictive rate.

The observed intelligence and reading abilities of the true positive gifted students and the false positive gifted students identified by multi-measure posterior probability method are plotted in Figure 13 and Figure 14. The correlation between observed intelligence and observed abilities in reading is .08 within the TGP students and -.14 within the FPT students. Both correlations are weak but statistically significantly different from zero. The negative correlation between observed intelligence and abilities in reading suggests that FPT students whose observed intelligence is relatively higher may have relatively lower observed abilities in reading, and vice versa. For a criterion of .95 for $P(T_3 \geq 2 \ \& \ T_1 \geq 2 \mid X_3, X_1)$, the multi-measure posterior probability method identifies 55.4% of the students TG in reading and intelligence and almost all the IG students are TG in reading and intelligence. In Figure 13, within the IG group identified using the criterion of 0.95 for $P(T_3 \geq 2 \ \& \ T_1 \geq 2 \mid X_3, X_1)$, only students whose observed intelligence and abilities in math are both greater than 2.0 (dots in the upper right square) will be identified as gifted by the AND method. Comparing with the performance of gifted identification using the other two rules to combine the observed intelligence and abilities in math, the students whose means of the observed intelligence and abilities in math (dots under the diagonal line) will be misidentified by the MEAN method, the students whose observed intelligence and abilities in math are both lower than 2.0 (dots in the bottom left square) will be misidentified by the OR method.

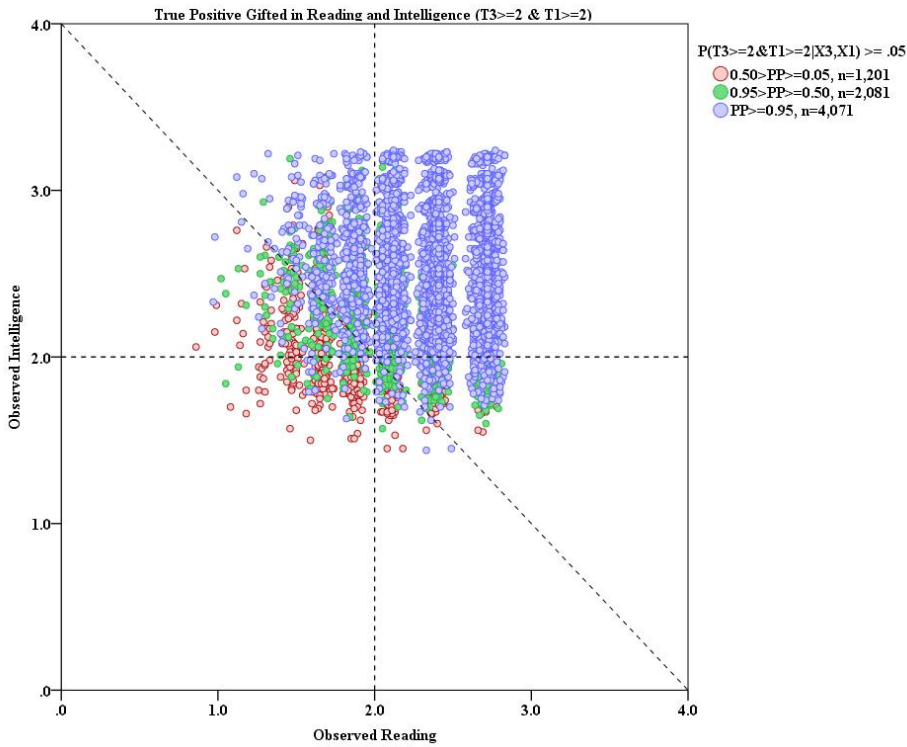


Figure 13. Students who are true positive gifted in reading and intelligence identified by the multi-measure posterior probability method.

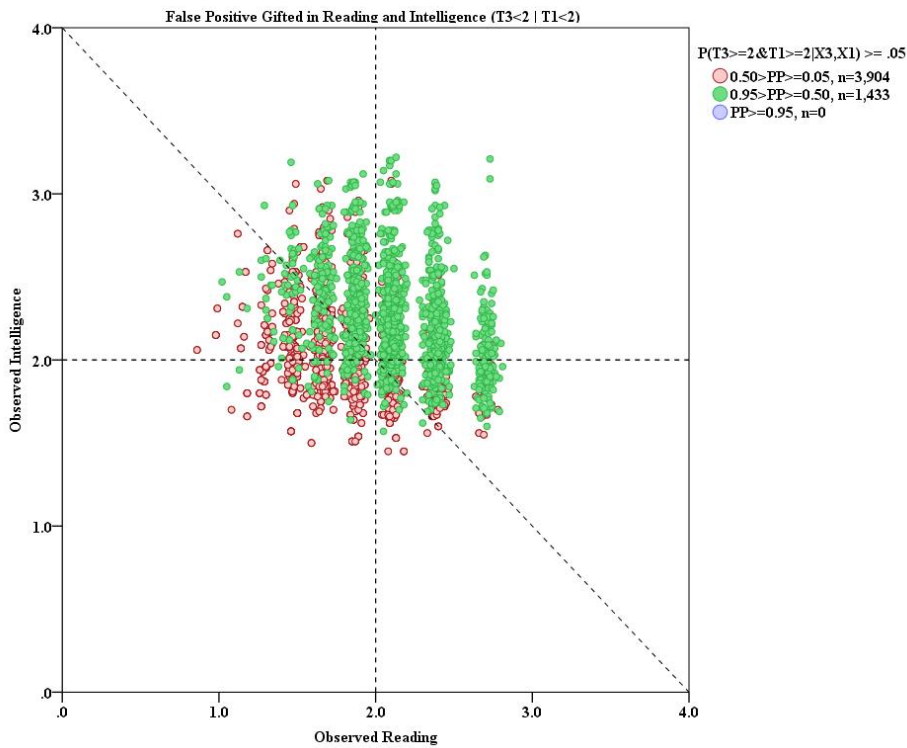


Figure 14. Students who are false positive gifted in reading and intelligence identified by the multi-measure posterior probability method.

Chapter 5

Discussion

Defining the Target Population of Gifted Identification

Gifted identification should always start with a clear definition of the target gifted population, because the target population of gifted identification directly decide the method that can and should be used. For example, if the target gifted population is students who are TG in at least one domain area, then gifted identification using the conjunctive rule to combine multiple measures of abilities in different domain areas will be very insensitive to the target gifted population and miss most of the target gifted population. According to the results in Table 12, given that gifted identification is based on the three simulated measures, the percent of students who are TG in at least one domain area but will not be identified as gifted by the AND-2 method (gifted identification using the conjunctive rule to combine the three measures) is

$$1 - \frac{P(TPG)_{AND-2}}{P(TG)_{OR}} = 1 - \frac{0.31\%}{4.70\%} = 93.4\%.$$

On the other hand, if the target gifted population is students who are TG in all involved domain areas (e.g., math, reading, and intelligence), then gifted identification using the complementary rule to combine multiple measures of abilities in different domain areas will be very inaccurate with a large amount of false positive identification. According to the results in Table 12 and Table 13, given that gifted identification is based on the three simulated measures, the percent of students who are not TG in all three domain areas but will be identified as gifted by the OR method (gifted identification using the complementary rule to combine the three measures) is

$$1 - \frac{P(TG)_{AND-2}}{P(IG)_{OR}} = 1 - \frac{0.53\%}{5.03\%} = 89.5\%.$$

It means that there is only one student who is TG in all three domain areas in every ten students who are identified as gifted by the complementary rule. Therefore, gifted identification should

clearly define the target gifted population before select the method to identify gifted students.

The performance of gifted identification directly depends on the target gifted population and the method selected correspondingly.

The target gifted population of gifted identification should be distinguished from the general gifted population defined by the NAGC (2010) or the USDE (1993). Given the diversity in giftedness, it is impossible to identified students who are gifted in very different fields (e.g., math and artistic) at the same time. Students with different gifted characteristics should be identified accordingly. In practice, gifted identification commonly has very specific purposes (e.g., qualification for a gifted program of math or science), which actually decide the target gifted population of identification. A good understanding of the target gifted population defined based on a specific purpose is essential to select the right measures and use them appropriately.

Two Qualities of “Good” Measures

Given a target gifted population, a “good” measure should possess two key qualities. The first quality is the reliability of observed scores on the measure. Results reveal that gifted identification based on observed scores with higher reliabilities performs better in terms of the PPR and sensitivity. This finding is true for gifted identification based on a single measure and gifted identification based on multiple measures. The lowest reliability tested in this study is .80, which is the reliability of the simulated observed reading abilities. Given a reliability of .80, the PPR and sensitivity of the single measures method (the SM-READ) are .65 and .66, which is lower than the PPR and sensitivity of the single measures method based on the observed intelligence (.81 and .81, respectively) and observed abilities in math (.75 and .75, respectively). A PPR of 0.65 indicates that about a third of students who are TG in reading are misidentified and a sensitivity of 0.66 indicates that about a third of students who are identified as gifted are

actually not TG in reading. Therefore, if a gifted identification based on a single measure is expected to perform with a misidentification rate lower than a third for both TG students and non-TG students, the reliability of observed scores should not be lower than .8.

Further, when multiple measures are used in gifted identification, regardless of the rules used to combine the multiple measures, the difference in the reliabilities of observed scores on different measures may result in unfairness of identification among students with different gifted characteristics. The gifted identification is least sensitive to the giftedness in the domain area that is measured with lowest reliabilities. For example, when the target gifted population is students who are gifted in at least one of the three domain areas (math, reading, and intelligence), the method of gifted identification using the complementary rule (the OR method) is least sensitive to the students who are gifted in reading but most sensitive to the students who are gifted in intelligence. The reliability of the simulated observed abilities in reading is 0.8, which is lower than the reliability of the simulated observed intelligence ($\rho_1^2 = .95$) and the simulated observed abilities in math ($\rho_2^2 = .90$). Therefore, for gifted identification, a “good” measure should produce observed scores with sufficient reliabilities. Further, when multiple measures of different abilities are involved in gifted identification, the multiple measures of abilities in different domain areas should be consistently “good” in terms of the reliabilities to guarantee the fairness of identification among students who are gifted in different domain areas.

The second quality of a “good” measure is its relevance. Given a target gifted population, a “good” measures should present adequate correlation with the target giftedness. Moreover, when multiple measures of relevant abilities in different domain areas are used in gifted identification, the imbalance in correlations among relevant abilities in different domain areas may result in unfairness of identification among students with different gifted characteristics.

After accounting for the reliabilities of observed abilities, the gifted identification that uses the conjunctive rule and the compensatory rule to combine multiple measures is more sensitive to the students who are TG in the domain area that is more correlated with the other involved domain areas. For example, in this simulation study, the correlation between the simulated true abilities in math and reading ($r_{23} = .75$) is higher than the correlation between the simulated true intelligence and true abilities in reading ($r_{13} = .70$), and the correlation between the simulated true abilities in math and true intelligence ($r_{21} = .85$) is higher than the correlation between true abilities in reading and true intelligence ($r_{31} = .70$). Results reveal that the AND-3 method and the MEAN method are both most sensitive to the students who are TG in math and least sensitive to the students who are TG in reading.

Considering the common practical situation of gifted identification that use the intelligence test as the conclusive test, if the math and reading tests are used as the screening test, the gifted identification will be more sensitive to the students who are TG in math and intelligence than the students who are TG in reading. If the reading test serves as the conclusive test and the math and intelligence tests serve as the screening test, the gifted identification is more sensitive to the students who are TG in math than the students who are TG in intelligence. However, given that true abilities in reading is less correlated with the true abilities in math and true intelligence than the true abilities in math and true intelligence are correlated with each other, the gifted identification using the reading test as the conclusive test is much less sensitive to the students who are TG in math and intelligence than the gifted identification using the intelligence test or the math test is. Therefore, if abilities in one domain area is consistently less correlated with the abilities in the other domain areas involved in a gifted identification, then the gifted identification is least sensitive to the students who are TG in the domain area least

correlated with the other relevant domain areas. In brief, a “good” measure used in gifted identification should present adequate relevance to the abilities in the target gifted areas. Further, the relationship among the abilities in different domain areas involved in gifted identification should present fair balance across domain areas.

Understanding the Identified Gifted Group

After defining the target population and selecting “good” measures, the fundamental goal of gifted identification is to effectively identify the target gifted population. Therefore, it is important to understand who are identified as gifted to evaluate the effectiveness on identifying the target gifted population. It is simple when gifted identification is aimed to identify students who are gifted in a single domain area. However, it becomes much more complicated when gifted identification is aimed to identify a gifted population defined based on abilities in different domain areas. This study simulated the situations that gifted identification is based on three measures of abilities in three different domain areas. Based on giftedness in three different domain areas, seven mutually exclusive gifted populations are defined (three gifted populations who are TG in only one domain area, three gifted populations who are TG in only two domain areas, and one gifted population who is TG in all three domain areas). Increasing the number of domain areas will increase the complexity of defining the gifted populations.

Given the four multi-measure methods of gifted identification (the AND-2, AND-3, MEAN, and OR methods) tested for the second research question, results reveal that all four methods are most sensitive to the students who are TG in all three domain area (the TG7) but least sensitive to the students who are TG in only one domain area (the TG1, TG2, and TG3). However, gifted identification that uses the conjunctive rule to combine a high standard of gifted

for all three measures can only identify two thirds of the students in TG7 and miss almost all the students in the other six TG groups.

The two key qualities of “good” measures also affect the fairness of identification among different gifted populations. Given the lower reliability of observed abilities in reading and weaker correlations between reading and the other two domain areas, the students who are TG in reading only is the least favorite TG group for all four multi-measure methods. Further, as using the complementary rule to combine multiple measures to identify the entire gifted population, 80% of the entire gifted population is successfully identified as gifted; however, among the 20% false negative gifted students, almost a half of them are students who are TG in reading only.

Better Performance of Gifted Identification by Using Bayesian Methods

This study is the first study that explores the use of Bayesian methods in gifted identification. Results reveal that gifted identification using Bayesian methods to combine multiple measures has three advantages over the gifted identification using different rules to combine multiple measures. The first advantage is the performance of gifted identification using the Bayesian method to combine multiple measures can be compared across the criteria of gifted for the posterior probabilities of being TG given observed abilities. The is because the change in criteria of gifted for the posterior probabilities does not change the target gifted population but only affect the performance of identifying the target gifted population. However, using different rules to combine multiple measures is naturally aimed to identify different gifted populations, and therefore the performance of gifted identification using different rules to combine multiple measures cannot be compared with each other. For example, using the OR method to identify the target gifted population defined by the AND rule may achieve a high sensitivity to the target gifted population. However, the OR method has to achieve the high sensitivity by sacrificing the

PPR. With a great amount of false positive gifted students, the identified gifted group by the OR method cannot represent the target gifted population defined by the AND rule. This is not an issue for gifted identification based on the posterior probability because it only adjusts the criteria of gifted for the posterior probability to achieve the desired performance of identifying the same target gifted population.

The second advantage is that gifted identification based on posterior probabilities of being TG given multiple observed abilities can perform better than gifted identification using different rules to combine multiple measures in terms of the PPR and sensitivity at the same time. Both the PPR and sensitivity are important for gifted identification. However, gifted identification using different rules to combine multiple measures performs relatively well in terms of either the PPR or sensitivity but rarely both. Results reveal that gifted identification using the conjunctive rule to combine multiple measures usually performs better in the PPR than in the sensitivity but gifted identification using the complementary rule to combine multiple measures usually performs better in the sensitivity than in the PPR. A high sensitivity with a low PPR means that most of the target gifted population are successfully identified as gifted but conflated with a great amount of false positive gifted students. A low sensitivity with a high PPR means that the identified gifted group only includes a small portion of the target gifted population but most of the identified gifted group truly belongs to the target gifted population. Results reveal that, by adjusting the criteria of gifted for the posterior probability, gifted identification based on posterior probabilities can perform at the same level of the PPR or sensitivity and perform higher in the other term. Further, gifted identification based on posterior probabilities can perform higher in the PPR and sensitivity simultaneously.

The third advantage is that the performance of gifted identification based on posterior probabilities of being TG given multiple observed abilities can be quantified for each criterion for the posterior probabilities. Therefore, the decision on identifying gifted students can be made based on the desired performance. Results reveal that using a criterion of .40 for the posterior probability of being TG given multiple observed abilities results in an approximate balance of the PPR and sensitivity. For a criterion higher than .40 for the posterior probability, the PPR is higher than the sensitivity. For a criterion lower than .40 for the posterior probability, the PPR is lower than the sensitivity. In practice, if the precision of identifying students who are TG has priority over sensitivity, then gifted identification could consider using a criterion higher than .40; however, if the first priority of gifted identification is to identify as many TG student as possible, then sensitivity is favored over the PPR and a criterion lower than .40 should be considered. Increasing the criterion for the posterior probability results in an increase in the PPR but a decrease in the sensitivity. In contrast, decreasing the criterion for the posterior probability results in a decrease in the PPR but an increase in the sensitivity. By adjusting the criterion of gifted for posterior probabilities, gifted identification may achieve a relatively higher PPR at a cost of misidentifying more TG students or achieve a relatively higher sensitivity at a cost of identifying more false positive gifted students. The quantitative information about the performance of gifted identification based on posterior probabilities can be used to plan and conduct gifted identification based on a desired performance and a clear understanding of the cost.

Implications for Practice

The major findings in this study suggest that gifted identification should start with a clear definition of the target gifted population. Further, to guarantee the performance of gifted

identification, only quality and relevant measures should be used. When gifted identification is aimed to identify students who are gifted in different domain areas simultaneously or gifted in multiple domain areas, the difference in the reliability of observed abilities in different domain areas and the correlation among abilities in different domain areas should be trivial to assure the fairness of identification among students from different gifted populations.

Gifted identification based on posterior probabilities of being gifted given observed abilities present the potential to effectively improve the performance of gifted identification in terms of the PPR and sensitivity simultaneously. However, to actually use the posterior probabilities in the practice of gifted identification needs collaboration among testing companies (e.g., the Educational Testing Service), scholars and professionals in the field of gifted education, schools, educators, and policy makers.

Limitations and Future Study

This study is a simulation study. Therefore, all the assumptions made to simulate data limit the generalization of the findings in this study. The main assumption is that true abilities and observed abilities are both continuous variables and true abilities in different domains follow the multivariate normal distribution. Further, the observed abilities were generated based on the 1P-IRT model. With the increasing application of the two-parameter (2P) and three-parameter (3P) IRT model in test development, it is definitely worthy of future studies to investigate the performance of gifted identification based on observed abilities estimated based on 2P- or 3P-IRT models.

The present study only investigates the performance of gifted identification based on posterior probabilities of being TG given observed abilities in one or two domain areas. An excellent area for further research would be how to generalize the use of posterior probabilities

in the field of gifted identification. For example, gifted identification can use the posterior probabilities of being GT in a specific domain area given repeated observed measures or given two or more observed abilities in other relevant domain areas.

References

- American Educational Research Association. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Birch, J. W. (1984). Is any identification procedure necessary? *Gifted Child Quarterly*, 28(4), 157-161.
- Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied psychological measurement*, 6, 431-444.
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48, 1-29. Retrieved from <http://www.jstatsoft.org/v48/i06/>.
- Chester, M. D. (2003). Multiple measures and high-stakes decisions: A framework for combining measures. *Educational Measurement: Issues and Practice*, 22(2), 32-41.
- Colorado State Board of Education (2015). *Rules for the administration of the exceptional children's education act*. Retrieved from <https://www.sos.state.co.us/CCR/GenerateRulePdf.do?ruleVersionId=6251&fileName=1%20CCR%20301-8>
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York, NY: Harcourt Brace Jovanovich College Publishers.
- Cronbach, L. J., Linn, R. L., Brennan, R. L., & Haertel, E. H. (1997). Generalizability analysis for performance assessments of student achievement or school effectiveness. *Educational and Psychological Measurement*, 57, 373-399.
- Douglas, K. M. (2007). A general method for estimating the classification reliability of complex decisions based on configural combinations of multiple assessment scores (Doctoral dissertation, University of Maryland, College Park).

- Douglas, K. M., & Mislevy, R. J. (2010). Estimating classification accuracy for complex decision rules based on multiple scores. *Journal of Educational and Behavioral Statistics, 35*, 280-306.
- Friedman-Nimz, R. (2009). Myth 6: Cosmetic use of multiple selection criteria. *Gifted Child Quarterly, 53*, 248-250.
- Florida Department of Education (2013). *Education of gifted students in Florida*. Retrieved from <http://www.fldoe.org/core/fileparse.php/7567/urlt/stategiftedplan.pdf>
- Gardner, H. E. (1983). Multiple approaches to understanding. *Instructional-design theories and models: A new paradigm of instructional theory, 2*, 69.
- Gilbride, T. J., & Allenby, G. M. (2004). A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science, 23*, 391-406.
- Georgia Department of Education (2016). *Georgia resource manual for gifted education services*. Retrieved from <https://www.gadoe.org/Curriculum-Instruction-and-Assessment/Curriculum-and-Instruction/Documents/Gifted%20Education/Georgia-Gifted-Resource-Manual.pdf>
- Gulliksen, H. (1950). *Theory of mental tests*. New York, NY: Wiley.
- Hambleton, R., Swaminathan, H., & Rogers, H. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.
- Heubert, J. P., Ed., & Hauser, R. M., Ed. (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press.
- Hunt, E. (2010). *Human intelligence*. New York, NY: Cambridge University Press.

- Iowa Department of Education (2011). *Gifted and talented education in Iowa*. Retrieved from <https://www.educateiowa.gov/sites/files/ed/documents/IowaCodeandRule-Gifted-Talented.pdf>
- Johnsen, S. K. (2004). *Identifying gifted students: A practical guide*. Waco, TX: Prufrock Press.
- Kelley, T. L. (1947). *Fundamentals of statistics*. Cambridge, MA: Harvard University Press.
- Kim, S., & Feldt, L. S. (2010). The estimation of the IRT reliability coefficient and its lower and upper bounds, with comparisons to CTT reliability statistics. *Asia Pacific Education Review, 11*, 179-188. doi: 10.1007/s12564-009-9062-8
- Lord, F., & Novick, M. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Pub.
- McBee, M. T., Peters, S. J., & Waterman, C. (2014). Combining scores in multiple-criteria assessment systems: The impact of combination rule. *Gifted Child Quarterly, 58*, 69-89.
- McCoach, D. B., Kehle, T. J., Bray, M. A., & Siegle, D. (2001). Best practices in the identification of gifted students with learning disabilities. *Psychology in the Schools, 38*, 403-411.
- McCoach, D. B., Yu, H., Gottfried, A. W., & Gottfried, A. E. (2017). Developing talents: A longitudinal examination of intellectual ability and academic achievement. *High Ability Studies, 28*, 7-28.
- National Association for Gifted Children. (2010). *Pre-K-grade 12 gifted programming standards*. Washington, DC: Author. Retrieved from <http://www.nagc.org/sites/default/files/standards/K-12%20programming%20standards.pdf>

- National Association for Gifted Children. (2015). *State of the states in gifted education report: 2012-2013*. Washington, DC: Author.
- Oklahoma Department of Education (2016). *Oklahoma state annual report on gifted and talented education*. Retrieved from <http://sde.ok.gov/sde/sites/ok.gov.sde/files/Regulations%20and%20Program%20Approval%20Standards.pdf>
- Page, A. (2006). Three models for understanding gifted education. *Kairaranga*, 7(2), 11-15.
- Peters, G. J. Y. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist*, 16(2), 56-69.
- RStudio (2017). *RStudio: Integrated development environment for R* (Version 1.0.143) [Computer software]. Boston, MA.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: University of Chicago Press.
- Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan*, 60(3), 180-184, 261.
- Renzulli, J. S., & Reis, S. M. (1994). Research related to the schoolwide enrichment triad Model1. *Gifted Child Quarterly*, 38(1), 7-20.
- Renzulli, J. S., (2004). Introduction to identification of students for gifted and talented programs. In Renzulli, J. S. & Reis, S. M. (Eds.). *Identification of students for gifted and talented programs*. Thousand Oaks, CA: Corwin Press.
- Roid, G. H. (2003). *Stanford-Binet intelligence scale manual*. Itasca, IL: Riverside.

- Silva, F. (1993). *Psychometric foundations and behavioral assessment*. Newbury Park, CA: Sage Publications.
- Sternberg, R. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York, NY: Cambridge University Press.
- Sternberg, R. J., & Davidson, J. E. (Eds.). (2005). *Conceptions of giftedness*. New York, NY: Cambridge University Press.
- Sternberg, R. J., Grigorenko, E., & Bundy, D. A. (2001). The predictive value of IQ. *Merrill-Palmer Quarterly*, 47, 1-41.
- U.S. Department of Education. Office of Educational Research and Improvement. (1993). *National excellence: A case for developing America's talent* (report.) Washington, D.C.
- Wechsler, D. (2003). *WISC-V: Wechsler intelligence scale for children*. San Antonio, TX: Psychological Corporation.
- Worrell, F. C. (2009). Myth 4: A single test score or indicator tells us all we need to know about giftedness. *Gifted Child Quarterly*, 53, 242-244.
- Wyatt, J., Kobrin, J., Wiley, A., Camara, W. J., Proestler, N. (2011). *College Board research report 2011-15 on SAT benchmarks: Development of a college readiness benchmark and its relationship to secondary and postsecondary school performance*. Retrieved from <https://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2011-5-sat-college-readiness-benchmark-secondary-performance.pdf>

Appendix A

The Syntax of Generating Observed abilities

```
N=1000000
n=1000
m=N/n

length.test1=length(diffTest1.final)
length.test2=length(diffTest2.final)
length.test3=length(diffTest3.final)
obs.ab1=matrix(rep(0,n),ncol=1)
obs.ab2=matrix(rep(0,n),ncol=1)
obs.ab3=matrix(rep(0,n),ncol=1)

obsAbFile0=paste0(workingFolder,"observedAb_")

start=1
set.seed(10001)
for(i in 1:m){
  end=i*n
  test1.answers=irtgen(diffTest1.final,TrueScore[start:end,1])
  test2.answers=irtgen(diffTest2.final,TrueScore[start:end,2])
  test3.answers=irtgen(diffTest3.final,TrueScore[start:end,3])
  colnames(test1.answers)=paste0('itme',1:length.test1)
  colnames(test2.answers)=paste0('itme',1:length.test2)
  colnames(test3.answers)=paste0('itme',1:length.test3)
  est.diff.test1=mirt(test1.answers,1,itemtype = 'Rasch')
  est.diff.test2=mirt(test2.answers,1,itemtype = 'Rasch')
  est.diff.test3=mirt(test3.answers,1,itemtype = 'Rasch')
  obs.ab1=fscores(est.diff.test1)
  obs.ab2=fscores(est.diff.test2)
  obs.ab3=fscores(est.diff.test3)

  obs.ab.temp=cbind(obs.ab1,obs.ab2,obs.ab3)

  if(i==1){
    obs.ab=obs.ab.temp
  } else {
    obs.ab=rbind(obs.ab,obs.ab.temp)
  }
  obsAbFile=paste0(obsAbFile0,'1-',end,'.csv')
  write.csv(obs.ab,obsAbFile)
  start=end+1
}
```
