Eurasia Journal of Mathematics, Science & Technology Education, 2016, **12**(9), 2587-2598
doi: 10.12973/eurasia.2016.1269a

An Application of Multivariate Generalizability in Selection of Mathematically Gifted Students

Sungyeun Kim
Incheon National University, REPUBLIC OF KOREA
Dan Berebitsky
Southern Methodist University, USA

•*Received 6 August 2015*•*Revised 3 November 2015* •*Accepted 29 December 2015*

This study investigates error sources and the effects of each error source to determine optimal weights of the composite score of teacher recommendation letters and self-introduction letters using multivariate generalizability theory. Data were collected from the science education institute for the gifted attached to the university located within in a capital city in Korea. The results were as follows. First, error sources for the students were relatively large suggesting that the score variances explained the differences in giftedness among the students. Second, based on the maximum generalizability coefficient for teacher recommendation letters and self-introduction letters, the optimal weight ratio should adjust to 0.7:0.3 from 0.5:0.5 in the original institution weights. These results are specific to the selection assessment instruments considered in this study; however, the methodology applied can be utilized in other selection instruments developed by many institutions.

Keywords: mathematically gifted students; multivariate generalizability theory; self-introduction letters; teacher recommendation letters

INTRODUCTION

In an effort to better prepare students for a creative economy, the Korean government recently announced a plan to foster student creativity and increase convergence between science technology and information technology (Ministry of Science, ICT and Future Planning, 2013). This plan emphasized these skills by promoting the education institutes for the gifted as a leading model. In addition, the Ministry of Education (2013) announced another initiative, "The 3rd Comprehensive Plans for the Promotion of Education for the Gifted" (The 3rd Master Plan) to foster future creativity through the expansion of education institutions for the gifted using teachers' observations and nominations for the selection of gifted

Correspondence: Sungyeun Kim,
Incheon National University, 119 Academi-ro, Yeonsu-gu, Incheon,
Republic of Korea.
E-mail: syk@inu.ac.kr

ISSN: 1305-8223 http://iserjournals.com/journals/ejmste

students. The Korean government also pledged to support the implementation of a student-customized diagnostic test to better identify gifted students in traditionally disadvantaged populations. Given the importance that the government is placing on these initiatives, great focus is being placed on the validity of these selection methods. Improper identification of gifted students can lead to maladjustment (Baum, 1994) and impeded development (Renzulli & Reis, 1986). Hence, one of the most important issues in gifted education is to validly and reliably identify these students. In this paper, we explore how generalizability theory (GT) can be employed to improve the identification process for gifted students.

In 2003, the Korean government introduced "The 1st Master plan (2003-2007)", which expanded educational programs for the gifted. At this time, most gifted education institutes used multi-step identification procedures for selection of gifted students, which included the following:

1. Principals' or teachers' recommendations
2. Tests for logical thinking or creative problem solving ability
3. In-depth interviews
4. Oral examinations

Even with all these steps, institutions often had to depend heavily on scores from the paper-and-pencil tests of the second step due to the subjective raters in the other steps. By focusing on the results of the tests, the institutions tended to select high academic achieving students (Kim, 2007). Kim (2007) also showed that these students tended to have higher socio-economic status, which no doubt provided increased opportunities to develop academically such as extensive private tutoring. Hence, the selection of the gifted during "The 1st Master Plan" could ignore student's potential ability because they focused on current academic achievement, which was often tied to family financial support. Accordingly, in "The 2nd Master Plan (2008-2011)", selection of gifted students focused more on potential ability by increasing emphasis on the measurement of giftedness and teachers' recommendations to minimize the influence of socio-economic status and advantages such as private tutoring. The Korean government suggested selecting the gifted by using teachers' recommendations in the first step, a giftedness test in the second step, an academic aptitude test in the third step, and an interview in the fourth step. In 2009, the Ministry of Education switched the selection methods from paper and pencil tests to teachers' observations and nominations when selecting the students. The percentage of gifted education institutions using teachers' observations and nominations increased from 3.44% in 2009 to 12.95% in 2010, 42.16% in 2011, and 48.35% in 2012.

Upon implementation of "The 3rd Master Plan (2013-2017)", the Ministry of Education announced that the percentage of the gifted education institutions using teachers' observations and nominations would be 70% in 2017. They would use checklists to define the behavioral characteristics in the students and performance

State of the literature

- Valid and reliable identification of mathematically gifted students is one of the most important issues in gifted education.
- Identification of gifted students commonly includes several instruments such as achievement tests, ability tests, recommendation letters, check lists, and so on each with their own relative weight.
- In many cases, however, the selection decision is arbitrarily made based on national policy or a priori judgment, not based on educational measurements from empirical research.

Contribution of this paper to the literature

- This study focused on finding optimal weights by analyzing teacher recommendation letters and self-introduction letters through the use of applying generalizability theory.
- This study suggested a method of investigating effects of diversifying relative weights of instruments on reliability indices considering both multivariate generalizability theory and rater effects.
- The methodology applied in this study can contribute to the determination of the optimal combination of relative weights in other selection instruments and to provide a basis for making informed decisions in constructing other test instruments for the selection of gifted students.

observation instruments to assess creative problem solving abilities. The intention was to select the students with the highest potential regardless of the opportunities provided by private tutoring or prior learning. By changing the selection methods, policymakers aimed to expand entrance opportunities to the gifted education system in a short time, and changes did occur as the competitive rate for entering university-attached gifted education institutions decreased from 6.94:1 in 2009 to 2.69:1 in 2012.

However, adapting the selection instruments without empirical proof for the measurement process can threaten to the validity of the assessment. The Ministry of Education, in the limitations of "The 2nd Master Plan (2008-2011)", admitted to the lack of validity and reliability of teachers' recommendations and nominations. In addition, many pieces in the literature have indicated that the understanding of the development and adaptation of selection instruments was incomplete since it was introduced suddenly, without enough research into the validity and reliability. Therefore, several researchers (Choi, 2012; Nam, 2012; Park, 2011; Rhyu & Jung, 2010; Youn & Park, 2012) analyzed the validity based on correlations and classical test theory inter rater-reliability (i.e., Cronbach's α) for the selection instruments for the gifted, and they consequently modified the selection instruments (Park, 2011; Shin, 2011) based on the study results.

Since the reliability coefficients based on the classical test theory (CTT) are focused on the consistency of the test results of observation subjects or observation processes, many studies have pointed out the insufficiencies of explanation of several error sources, which occurred in measurement situations (Brennan, 2001; Cronbach et al., 1972; Feldt and Brennan, 1989; Johnson and Johnson, 2009; Qingping, 2009; Webb et al, 2007). For example, in selecting the gifted based on teachers' observation and nomination, the students' scores derived from these instruments overlooked the effects of scoring methods, the leniency of raters, the difficulty of items, and scoring distributions, other than their giftedness. In addition, CTT treats error as random and cannot be used to distinguish systematic measurement error from random measurement error. Generalizability theory (GT) was developed to address these problems encountered in CTT. GT is a measurement model that can be used to investigate relative effects of multiple error sources in measurement error on test scores. A remarkable feature of GT is that the relative contributions from individual sources of measurement error to the overall error variance can be observed including the interaction between the sources based on a generalizability study (G-study). An important application of GT is that the model can be used to design an optimal test, through a decision study (D-study), for a specific purpose with minimum measurement error. As Bachman (2004) demonstrated, CTT approaches to estimating reliability can be treated as a special case of GT.

In the univariate GT, the test score that a student obtained on a test is conceived of as a sample from a universe of all possible scores that are admissible. Each characteristic of the measurement situation forms a source of error in test scores, which we call a facet. Students are objects of measurement and are not a facet. While univariate GT can be used to estimate variance components and score dependability for the overall test, multivariate GT can be developed to address the reliability of measurements, where multiple scores, which represent performance on different constructs, are used to produce a composite score. Multivariate GT separates both observed variance and covariance into components, and it is used to find the conditions that produce maximum reliability for the composite score. Obviously, a univariate GT is a special case of a multivariate GT (Brennan, 2001a).

Given the multiple measures that are being used and have been used to select students into the gifted institutes, it is imperative for the research and policy communities to understand how to optimize the process to ensure reliability. Therefore, the purpose of this study is not to generalize the results of reliability analyses on selection instruments for identification of gifted students but to illustrate the method to produce the optimal combination of relative weights to maximize the reliability coefficients. In a relative evaluation, these reliability coefficients serve as a generalizability index, and in an absolute evaluation, they function as an index of dependability. In the literature, there are a few studies related to selection of the gifted using GT. Kim and Han (2013) examined teacher recommendation letters and self-introduction letters separately based on the multivariate GT. Joni and colleagues (2011) used the multivariate GT to analyze cognitive ability tests including verbal, quantitative, and nonverbal reasoning; however, they did not consider rater effects. In addition, Cropper (1996) used traits, attributes and student behavior referrals (TABS) based on the univariate GT. However, we located no study concerning both the multivariate GT and rater effects. In this study, we consider teacher recommendation letters and self-introduction letters at the same time reflecting rater effects based on the multivariate GT. The goal is to make this methodology more intuitive by applying it to existing gifted instruments.

The research questions are as follows:

1. What effects do both teacher recommendation letters and self-introduction letters have on error variance?
2. What are the optimal weights for teacher recommendation letters and self-introduction letters?
3. What effect does changing the relative weights of teacher recommendation letters and self-introduction letters have on the effective weights?

METHODS

Data and instruments

Teacher recommendation letters and self-introduction letters were collected from 90 passed students (12 females and 78 males) applying to the gifted education program in 2011 in one science education institute for the gifted, which is attached to a university located in the Seoul metropolitan area. Students from 36 schools in the same area were selected based on a teachers' observation and nomination system. The rubric of teacher recommendation letters and self-introduction letters consists of a set of 4 domains and 16 specific items used to evaluate giftedness. The domains are cognitive, mathematical, affective and social, and each specific item is scored on 5 point scale. The cognitive domain consists of intellectual curiosity, concentration, reading, and self-regulation ability. The mathematical domain consists of mathematical understanding, mathematical creativity, mathematical problem solving, and mathematical critical thinking. The affirmative domain consists of challenge, gumption, task commitment, and mathematical aptitude. Finally the social domain consists of leadership, morality, sociability, and humor. A more specific scoring rubric is provided in Appendix.

The two raters consisted of elementary school teachers trained to recognize the gifted through the district in-service training and university course work and had teaching experience for the gifted program evaluated using analytic method. During the evaluation, raters maintained the time interval, and evaluated randomly changing order of students.

Relative, nominal, and effective weights

For both instruments, researchers wanted a composite that had a particular proportion u_t of the composite score range related to teacher recommendation letters, and a proportion $u_s = 1 - u_t$ of the composite score range related to self-introduction letters. These u_t and u_s call relative weights. The relative weights are defined a priori by researchers. In this study, for both teacher recommendation letters and self-introduction letters $u_t = u_s = 0.5$.

For obtaining composite scores, the relative weights, predetermined composite score range, and other characteristics of the instruments are used to gain nominal weights such that

$$\omega_t X_t + \omega_s X_s = C \tag{1}$$

where ω_t is the nominal weight for teacher recommendation letters, ω_s is the nominal weight for self-introduction letters, X_t is teacher recommendation letters' total score, X_s is self-introduction letters' total score, and C is the composite score. In this study, we set $C = 100$, for teacher recommendation letters and self-introduction letters $\omega_t = \omega_s = 0.5 \cdot 100 / 80 = 0.625$.

In the context considered here, X_t and X_s are raw scores in the total-score metric (TSM). However, in GT, the usual convention is to do calculations in the mean-score metric (MSM). This convention is used in the computer software program mGENOVA (Brennan, 2001b), which was used to compute the results in this study. Hence, to gain composite results using mGENOVA, it is necessary to change terms on the left side of (1) to their analogues in the MSM. To do so produces

$$v_t \bar{X}_t + v_s \bar{X}_s = C, \tag{2}$$

where v_t is the nominal weight for teacher recommendation letters, and v_s is the nominal weight for self-introduction letters based on the MSM, $\bar{X}_t$ is teacher recommendation letters' mean score, and $\bar{X}_s$ is self-introduction letters' mean score (Brennan, 2009). In this study, both teacher recommendation letters and self-introduction letters are $v_t = v_s = 0.625 \cdot 16 = 10$.

If researchers' interests are not reliabilities based on teacher recommendation letters and self-introduction letters, respectively, but based on composite score reflecting both teacher recommendation letters and self-introduction letters at the same time. We should consider effective weights.

For each instrument, the proportional contribution to composite universe score variance is

$$ew_t(p) = \frac{v_t(v_t \sigma_t^2(p) + v_s \sigma_{ts}(p))}{\sigma_C^2(p)}, \tag{3}$$

$$ew_s(p) = \frac{v_s(v_s \sigma_s^2(p) + v_t \sigma_{ts}(p))}{\sigma_C^2(p)}, \tag{4}$$

which means that $ew_t(\tau) + ew_s(\tau) = 1$. These $ew_t(\tau)$ and $ew_s(\tau)$ call effective weights. It is important to note that nominal weights express the researcher's judgment about the relative importance of teacher recommendation letters and self-introduction letters. However, effective weights based in part on nominal weights show the relative statistical contribution of each instrument (Wang & Stanley, 1970).

Representative relative effects, nominal TSM and MSM weights, and score ranges for teacher recommendation letters and self-introduction letters are provided in Table 1. In this study, we used 39 conditions based on 0.25 increment in teacher recommendation letters and self-introduction letters.

Analyses

Data was analyzed according to the multivariate GT. GT provides a framework to conceptualize and disentangle multiple sources of error in a measurement procedure(Brennan, 2001a). For teacher recommendation letters and self-

Table 1. Several conditions of relative weights and nominal weights

Relative Weights		Nominal TSM weights		Nominal MSM weights		Score Range	
T	S	T	S	T	S	T	S
0.1	0.9	0.125	1.125	2	18	0-10	0-90
0.2	0.8	0.250	1.000	4	16	0-20	0-80
0.3	0.7	0.375	0.875	6	14	0-30	0-70
0.4	0.6	0.500	0.750	8	12	0-40	0-60
0.5	0.5	0.625	0.625	10	10	0-50	0-50
0.6	0.4	0.750	0.500	12	8	0-60	0-40
0.7	0.3	0.875	0.375	14	6	0-70	0-30
0.8	0.2	1.000	0.250	16	4	0-80	0-20
0.9	0.1	1.125	0.125	18	2	0-90	0-10

Note1. T means teacher recommendation letters and S means self-introduction letters.
2. Shadowed region are the original weight using sample size.

introduction letters, with students (p) as the object of measurement, three facets contribute to the student's scores variability, i.e., types of instrument (t), items (i) and raters (r). It is frequently the case that for each rater, all students are scored with the same sets of items, whereas the instrument types (teacher recommendation letters, self-introduction letters) are considered as fixed. We treated the two instrument types as fixed because they do not change across the subsequent educational policy based on observation and nomination system to select the gifted, even though the raters and items in the rubric are replaced. Given this conceptualization, a multivariate random facet $p^{\bullet} \times r^{\bullet} \times i^{\bullet}$ G study design would be most appropriate. The notation used in this paper follows Brennan (2001a). The superscript filled circle $\bullet$ designates that the facet is crossed with the fixed domains.

The data were analyzed using mGENOVA and the Excel (Brennan, 2001b). The mGENOVA software estimates variance and covariance components, and the Excel calculates reliability coefficients, relative weights, and effective weights for the composite score in the multivariate GT.

RESULTS

What effect does teacher recommendation letters and self-introduction letters have on error variance?

The variance and covariance component estimates for the $p^{\bullet} \times r^{\bullet} \times i^{\bullet}$ design are reported in Table 2. Variance component estimates for the students in both teacher recommendation letters and self-introduction letters were relatively large. These results suggest that the score variances explained the differences in the giftedness among the students. The large magnitude of variance component estimates for the residual is not too surprising, since it incorporates variability attributable to the three-way interaction as well as any other sources of variation that are not included in the analysis. Hence, it is often associated with the highest order interaction (Brennan, 2001a, p.83). In addition, the disattenuated correlation, i.e., true score correlation adjusted measurement errors, was 0.904 ($= 0.148 / \sqrt{0.200 \cdot 0.145}$), suggesting that a student that scored high on teacher recommendation letters tended to score high on self-introduction letters as well, and vice versa.

The covariance between students and raters was also relatively large compared to the corresponding variance components. This indicates that, depending on rater leniencies, students' relative ranks were comparatively consistent on both instruments. On the contrary, all other covariance components, except for the students, were relatively small, meaning that the score covariation due to these sources of variation was negligible.

In both teacher recommendation letters and self-introduction letters, the interaction effects between students and raters were relatively large, suggesting that students were rank-ordered very differently across the raters. In addition, in teacher recommendation letters, item effects and interaction effects between students and items are relatively large. This suggests that the contents of items are a considerably greater source of variability in students' scores than raters. On the contrary, in self-introduction letters, leniency of raters were still a considerably greater source of variability in students' scores.

What are optimal weights for teacher recommendation letters and self-introduction letters?

Composite universe score variance, error variance, and reliability indices are provided in Table 3. These results depend on several relative weight conditions from Table 1 with the nominal MSM weights, and the results from Table2. The

Table 2. G-study Results for the $p^{\bullet} \times r^{\bullet} \times i^{\bullet}$ design

	Teacher Recommendation Letters		Self-introduction Letters
$\sigma^2(p)$	0.200		0.134
$\sigma_{ts}(p)$		0.148 (0.904)	
$\sigma^2(r)$	0.004		0.023
$\sigma_{ts}(r)$		0.012	
$\sigma^2(i)$	0.070		0.012
$\sigma_{ts}(i)$		0.022	
$\sigma^2(pr)$	0.077		0.091
$\sigma_{ts}(pr)$		0.045	
$\sigma^2(pi)$	0.069		0.027
$\sigma_{ts}(pi)$		0.000	
$\sigma^2(ri)$	0.024		0.004
$\sigma_{ts}(ri)$		0.006	
$\sigma^2(pri)$	0.337		0.352
$\sigma_{ts}(pri)$		0.098	

Note 1. σ^2 denotes the variance, and σ_{ts} denotes the covariance between teacher recommendation letters and self-introduction letters.

2. ( ) denotes the disattenuated correlation coefficients.

Table 3. Composite score variances and reliability indices

	Composite scores				
Relative Weights(T:S)	**Universe score variance**	**Relative error variance**	**Absolute error variance**	**Generalizability coefficient**	**Index of the dependability**
0.1: 0.9	54.760	20.898	25.538	0.724	0.682
0.2: 0.8	56.355	19.022	23.301	0.748	0.707
0.3: 0.7	58.254	17.629	21.585	0.768	0.730
0.4: 0.6	60.456	16.718	20.391	0.783	0.748
0.5: 0.5	62.962	16.290	19.717	0.794	0.762
0.6: 0.4	65.771	16.344	19.564	0.801	0.771
0.7: 0.3	68.884	16.880	19.933	0.803	0.776
0.8: 0.2	*72.301*	*17.889*	*20.822*	*0.802*	*0.776*
.0.9: 0.1	76.021	19.400	22.233	0.797	0.774

Note 1. T means teacher recommendation letters and S means self-introduction letters.

2. Shadowed region are the original weight using sample size.

shaded region denotes the original relative weights in this study; the bolded values denote the highest generalizability coefficient; and the italic values denote the highest index of the dependability. In the original data, the generalizability coefficient was 0.794, and the index of the dependability was 0.762. In this study, rater effects, item effects and interaction effects between raters and items were greater than zero. Hence the index of dependability was always smaller than the generalizability coefficient.

Clearly different instrument weights result in distinct composite universe score variance, error variance, and reliability indices. Generally, the effect of increasing the relative weight assigned to the teacher recommendation letters was an increase in universe score variance, a decrease in error variance, and an increase in composite score reliability coefficients. Conversely, increasing this weight in self-introduction letters resulted in lower universe score variance, higher error variances, and lower reliability indices. When we use the relative weights of 0.7: 0.3, comparing teacher recommendation letters to self-introduction letters, we can obtain the highest generalizability coefficient: 0.803; and with 0.8: 0.2, we can obtain the highest index of the dependability: 0.762. Based on Table 2, teacher recommendation letters are more reliable than the self-introduction letters. When we put higher weights in teacher recommendation letters, we can obtain high reliability coefficients.

Effective weights

Effective weights quantify the proportion of composite variance contributed by a given fixed facet. In this study, instrument type is the fixed facet and effective weights represent the proportion of composite score variance attributable to the teacher recommendation letters and self-introduction letters. Relative weights are based on the proportion of composite score points assigned to each instrument type. However, effective weights are based on the proportion of composite score variance attributable to each instrument–either universe score variance or error variance (Brennan, 2009).

Table 4 provides the effective weights that resulted from all conditions related to Table 1. The shaded region denotes the original data in this study, the bolded values denote the highest generalizability coefficient and the italic values denote the highest index of the dependability. In the original data, the relative weights are 0.5 : 0.5 because these instruments have the same number of items. However, effective weights based on universe score variance in the original data are 0.553 : 0.447. This suggests the relative contribution of teacher recommendation to composite variance must be larger for self-introduction letters. This is not surprising given that effective

Table 4. Effective weights

Relative weights	**Effective Weights**					
	$ew(p)$		$ew(\delta)$		$ew(\Delta)$	
T: S	**T**	**S**	**T**	**S**	**T**	**S**
0.1: 0.9	0.112	0.888	0.054	0.946	0.056	0.944
0.2: 0.8	0.225	0.775	0.131	0.869	0.132	0.868
0.3: 0.7	0.337	0.663	0.231	0.769	0.229	0.771
0.4: 0.6	0.447	0.553	0.352	0.648	0.345	0.655
0.5: 0.5	0.553	0.447	0.486	0.514	0.474	0.526
0.6 : 0.4	0.654	0.346	0.622	0.378	0.606	0.394
0.7: 0.3	0.750	0.250	0.748	0.252	0.733	0.267
0.8: 0.2	*0.839*	*0.161*	*0.856*	*0.144*	*0.844*	*0.156*
0.9: 0.1	0.923	0.077	0.940	0.060	0.934	0.066

Note 1. T means teacher recommendation letters and S means self-introduction letters.
2. Shadowed region are the original weight using sample size.

weights were based on score variability in Table 1, and the variance component, based on students for teacher recommendation letters, was larger than self-introduction letters. If the relative weights are adjusted to provide equal effective weights with 0.50 : 0.50, the generlizability coefficient and the index of the dependability of the composite score are modestly decreased to 0.789 and 0.756, respectively. To produce these results, the relative weights were adjusted to be 0.45 and 0.55 in teacher recommendation letters and self-introduction letters, respectively.

CONCLUSIONS

In recent years, the Korean government, among others, has placed a large emphasis on the identification and education of gifted students in order to improve the future workforce. Initiatives though have struggled with finding the best selection method for identifying gifted students. This study investigated the effects that teacher recommendation letters and self-introduction letters, two common selection methods, have on reliability indices and error variances. Furthermore, it illustrated a method to produce the optimal combination of relative weights to maximize the reliability coefficients of composite scores using multivariate GT, which has the potential to provide a better selection method for those attempting to identify gifted students. It should be noted that the optimal combination of relative weights obtained in this study is not generalizable to other selection instruments, but the methodology applied can be utilized in other selection instruments developed by many institutions.

The generalizability coefficient in teacher recommendation letters was 0.789 ($= 0.200/(0.200+0.077/2+0.069/16+0.337/32)$), but the generalizability coefficient was lower, 0.697 ($= 0.134/(0.134+0.091/2+0.027/16+0.352/32)$), in self-introduction letters. Composite score reliabilities were higher than the score reliability for each instrument because both of them were given the same relative weights. One method to increase composite score reliability would be to give the teacher recommendation letter instrument higher relative weight in forming composite scores. The optimal relative weights for the teacher recommendation letters must be 0.7 to maximize reliability because of the low reliability of self-introduction letters. Considering the relationship between composite score reliability and effective weights leads to the conclusion that highly weighted teacher recommendation letters are necessary to maximize the reliability. When it comes to the composite universe score variance, the effective weight of teacher recommendation letters should be 0.75 to maximize the reliability coefficient. For relative composite error variance, the higher effective weight of teacher recommendation letters also corresponds to higher composite score reliability.

The implications of these results are as follows: First, when researchers use the same rubric for teacher recommendation letters and self-introduction letters, the relative weights should be different for each instrument. This is because teachers and students think about giftedness differently; when teachers make teacher recommendation letters and students make self-introduction letters, they will emphasize different characteristics of giftedness. Hence, in evaluating both instruments, researchers should show the rubrics from each gifted institution, and provide some guidelines for how to write these letters. In addition, as suggested by Kim & Han (2013), we should use different rubrics for teacher recommendation letters and self-introduction letters, and these two rubrics should complement each other. Second, the multivariate GT can be applied to find optimal relative weights based on the rubric of each instrument. Optimal weights for these measures will provide evaluators better information for making informed decisions in using selection instruments. Third, the results of the multivariate GT showed that the assumption of summing or averaging scores, which many test developers use, can be wrong. Although nominal weights represent the test developer's desired weights, reflecting the relative importance of different domains, the nominal weights are not necessarily the same as effective weights (Wang & Stanley, 1970). For example, in this study, the two instruments are all scored on a five-point-scale with the same items. Nonetheless, the effective weight associated with teacher recommendation letters was larger than that of self-introduction letters. In situations where the instrument scores are on substantially different scales, or different item numbers, this distinction will be significant. In this latter circumstance, it will be nearly impossible for test developers to interpret descriptions of a composite score based on relative weights. Hence, we investigated the degree of match between actual and desired weighting of different instruments in a composite score as defined in test developers.

Possible limitations to this study include the following: first, we do not consider estimating variance components for domains. Because we use mGENOVA in analyzing most cases, it supports a maximum of only three facets within each instrument as fixed levels (Brennan, 2001b). Hence in this study, we use raters instead of domains, because many researchers mention their limitations related to not considering rater facets. Second, we only use teacher recommendation letters and self-introduction letters as instruments; however, there are many other instruments such as gifted behavioral checklists, parents' observation sheets, scholastic records, portfolios, and intensive interviews. Therefore, future research is necessary to understand how they interact with recommendation and introduction letters in the gifted selection process. In addition, teacher recommendation letters and self-introduction letters are balanced data composed of the same item numbers within each domain. However, many instruments vary in item numbers within each domain, score scales within each domain, and rater training and implementation. Consequently, future research needs to consider how to obtain optimal weights using unbalanced instruments. Finally, analyses reported in this study were based on relatively small samples for the purpose of illustration. Therefore, estimated variance and covariance components are subject to sampling error and should be interpreted cautiously.

ACKNOWLEDGEMENT

Preparation of this article was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2013S1A3A2055007).

REFERENCES

Bachman, L. (2004). *Statistical analysis for language assessment*, Cambridge: Cambridge University Press.

Baum, S. M. (1994). Meeting the need of gifted/learning disabled students. *The Journal of Secondary Gifted Education, 3*(5), 6-16.

Brennan, R. L. (2001a). *Generalizability Theory*. New York: Springer.

Brennan, R. L. (2001b). *Manual for mGENOVA Version 2.1*. Iowa City, IA: Center for Advanced Studies in measurement and Assessment, The University of Iowa.

Brennan, R. L. (2009). Notes about nominal weights in multivariate generlaziability theory. *CASMA Research Report, No.25*. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment. The University of Iowa.

Choi, E. (2012). *Gifted observations used in the recommended starting teacher recommendation and cover letter analysis of the relationship between educational outcomes and the actual*. Master's thesis, Ajou university.

Cronbach, L. J., Gleaser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The Dependability of Behavioral Measurement*. Wiley, New York.

Cropper, C. P. (1996). A generalizability theory study of a survey instrument to identify gifted and talented students: The looking for traits, attributes, and behaviors student referral form. Unpublished doctoral dissertation. Texas Tech University.

Feldt, L. S. & Brennan, R. L. (1989). Reliability. In: Linn, R. L. (Ed.), *Educational measurement*, 3rd ed. The American Council on Education, Macmillan, 105-146.

Johnson, S. & Johnson, R. (2009). *Conceptualising and interpreting reliability*, Coventy: Ofqual.

Joni, M. L. & Emily, R. L. (2011). Multigroup generalizability analysis of verbal, quantitative, and nonverbal ability tests for culturally and linguistically diverse students. *Educational and Psychological Measurement, 72(1)*, 139-158.

Kim, M. (2007). Identification of gifted and talented students in Korea: The current situation and the search for the new direction. *International Symposium on New Directions in Identification of the Gifted and Talented*. RM 2007-45. Ministry of Education & Human Resources Development.

Kim, S. & Han, K. (2013). An application of multivariate generlalizability theory to teacher recommendation letters and self-introduction letters used in selection of mathematically gifted students by observation and nomination. *Journal of Gifted Talented Education, 23(5),* 671-695.

Ministry of Education. (2009). The excavation the Science Gifted without testing. Seoul: Ministry of Education.

Ministry of Science, ICT and Future Planning. (2013). The plan to promote creative talents to raise creative economy. Seoul: Ministry of Science, ICT and Future Planning.

Nam, Y. (2012). *Selection method validity analysis of the GED as a basis for observations and nominations of mathematically gifted*. Master's thesis, Daegu National University of Education.

Park, H. (2011). *Analysis of inter rater reliability of self introduction letter and teacher's recommendation letter used in identification of gifted students by class observations and nominations*. Master's thesis, The University of Incheon.

Qingping, H. (2009). *Estimating the reliability of composite scores*. Coventy: Ofqual.

Renzulli, J. S. & Reis, S. M (1986). The enrichment triad/revolving door model: A school wide plan for the development of creative productivity. In Renzulli, J. S. (Ed.) *Systems And Models For Developing Programs For The Gifted And Talented*. Creative Learning Press.

Rhyu, J. & Jung, H. (2010). Identification model development for gifted students based on class observations and nominations. *Journal of Gifted Talented Education, 20(1),* 257-287.

Shin, E. (2011). *Analysis of behavioral characteristics of the scientific gifted students with teachers recommendations and students self-nominations*. Unpublished doctoral dissertation. Seoul National University.

Youn, E. & Park, Y. (2012). Analysis of the reliability and validity of selection results of the gifted student by observation and nomination system. *Journal of Gifted Talented Education, 22(4),* 929-942.

Wang, M. D. & Stanley, J. C. (1970). Differential weighting: A review of methods and empirical studies. *Review of Educational Research, 40*, 663-705.

Webb, N., Shavelson, R., & Haertel, E. (2007). 'Reliability coefficient and generalizability theory', *Handbooks of Statistics 26: Psychometrics* (C Rao and S Sinharay Eds), 81-124, the Netherlands: Elsevier.

◈◇◈

APPENDIX: Assessment rubric for teacher recommendation letters and self-introduction letters

Domains	Items	Assessment Criteria	Points
Cognitive Attribute	Intellectual Curiosity	acquires information quickly and shows high level of curiosity	5
		⋮	⋮
		acquires information very slowly and shows no signs of curiosity	1
	Concentration Ability	shows very high level of concentration ability	5
		⋮	⋮
		shows very low level of concentration ability	1
	Book Reading	enjoys reading books and reads a lot	5
		⋮	⋮
		shows no comment on book reading	1
	Self-directed Learning Ability	able to perform self-directed learning	5
		⋮	⋮
		shows no evidence of self-directed learning	1
Mathematical Attribute	Mathematics subject grade	always achieves high grade in mathematics	5
		⋮	⋮
		shows no comment on mathematics grade	1
	Mathematical Creativity	shows high levels of fluency, flexibility, variety, accuracy in solving mathematical problems	5
		⋮	⋮
		shows very low levels of fluency, flexibility, variety, accuracy in solving mathematical problems	1
	Mathematical Problem Solving Ability	has a great mathematical problem solving ability	5
		⋮	⋮
		has a very low mathematical problem solving ability	1
	Mathematical thinking skills	has great thinking skills that require to understand and solve mathematical problems	5
		⋮	⋮
		has poor thinking skills that require to understand and solve mathematical problems	1
Affective Attribute	Spirit of Challenge	shows the spirit of challenge to face and solve difficult problems	5
		⋮	⋮
		shows no comment the spirit of challenge to face and solve difficult problems	1
	Initiative	takes the initiative when given tasks	5
		⋮	⋮
		does not take the initiative when given tasks	1
	Task Commitment	puts all possible effort to solve difficult problems	5
		⋮	⋮
		shows no evidence of task commitment	1
	Mathematical Talent	has a deep interest and talent in mathematics	5
		⋮	⋮
		dislikes mathematics and shows no talent in it	1
Social Attribute	Leadership	has the ability to perform as a leader in a group	5
		⋮	⋮
		does not have the ability to perform as a leader in a group	1
	Morality	shows sharing and caring attitude to others	5
		⋮	⋮
		shows no sharing and caring attitude to others	1
	Sociability	able to adapt to the society and has good relationships	5
		⋮	⋮
		not able to adapt to the society and has poor relationships	1
	Sense of humor	has a sharp sense of humor and attempts to influence others through humor	5
		⋮	⋮
		shows no sense of humor	1