Developing a Measure of General Academic Ability: An Application of Maximal Reliability and Optimal Linear Combination to High School Students' Scores

Educational and Psychological

Measurement

2015, Vol. 75(3) 475–490

© The Author(s) 2014

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164414543179

epm.sagepub.com

Dimiter M. Dimitrov^{1,2}, Tenko Raykov³, and Abdullah Ali AL-Qataee²

Abstract

This article is concerned with developing a measure of general academic ability (GAA) for high school graduates who apply to colleges, as well as with the identification of optimal weights of the GAA indicators in a linear combination that yields a composite score with maximal reliability and maximal predictive validity, employing the framework of the popular latent variable modeling methodology. The approach to achieving this goal is illustrated with data for 6,640 students with major in Science and 3,388 students with major in Art from colleges in Saudi Arabia. The indicators (observed measures) of the targeted GAA construct were selected from assessments that include the students' high school grade and their scores on two standardized tests developed by the National Center for Assessment in Higher Education in Saudi Arabia, General Aptitude Test (GAT) and Standardized Achievement Admission Test (SAAT). A unidimensional measure of GAA was developed initially, with different sets of indicators for colleges with major in Science and for colleges with major in Art. Appropriate indicators for colleges with major in Science were the high school grade, total score on GAT, and four SAAT subscales on Biology, Chemistry, Physics, and

¹George Mason University, Fairfax, VA, USA

²National Center for Assessment in Higher Education, Riyadh, Saudi Arabia

³Michigan State University, East Lansing, MI, USA

Corresponding Author:

Dimiter M. Dimitrov, Graduate School of Education, George Mason University, 4400 University Dr., MS 6D2, Fairfax, VA 22030, USA.

Email: ddimitro@gmu.edu

Math. With respect to colleges with major in Art, appropriate GAA indicators were the students' high school grade and their scores on GAT-Verbal, GAT-Quantitative, and SAAT. Although the case study is Saudi Arabia, the methods and procedures discussed in this article have broader utility and can be used in different contexts of educational and psychological assessment.

Keywords

maximal reliability, predictive validity, latent variable modeling

The development of psychometric instruments with high measurement quality has been an area of great concern and interest to methodologists as well as educational and behavioral researchers for a number of decades (e.g., Raykov & Marcoulides, 2011, and references therein). The concepts of reliability and validity have thereby played a crucial role. However, the notions of optimal linear combination (OLC) and associated maximal reliability have been receiving less attention than they deserve, especially in empirical research. Part of the reason may have been the fact that readily accessible discussions of how to apply them in the process of instrument construction and development, in particular using widely circulated software, have been lacking.

This article aims to contribute to bridging this gap by providing a treatment of the topic from the perspective of an applied researcher. Specifically, we discuss an approach to linear composite construction based on the above-mentioned optimality concepts and use it on achievement-related data from Saudi Arabia. High school graduates in Saudi Arabia who apply to colleges are required to take two tests developed and administered by the National Center for Assessment (NCA) in Higher Education, Riyadh, namely, General Aptitude Test (GAT) and Standardized Achievement Admission Test (SAAT). A key factor in the decision on acceptance to college is a composite score of the applicants based on their high school grade (HSG) and test scores on GAT and SAAT. This composite is intended to represent a linear combination of those scores with optimal weights. The predictive validity of such a composite score is oftentimes evaluated by its correlation with the first-year grade point average (FYGPA) of the students in college.

From a measurement perspective, there are several issues that arise with such composite scores, which need to be addressed. First, it is important to examine whether the composite scores are essentially unidimensional, that is, whether there is a general latent ability that dominates the students' scores on the HSG, GAT, and SAAT measures participating in the composite score. Second, one needs also to evaluate the reliability of the composite scores and to compare it to that of appropriately constructed alternatives, in particular with the (maximal possible) reliability of scores furnished by an OLC of the HSG, GAT, and SAAT components.

In this context, the purpose of the empirical study underlying this article was two-fold. The first goal was to develop a measure of general academic ability (GAA) that underlies the HSG, GAT, and SAAT scores of students who apply to colleges with major in Science and colleges with major in Art. The next goal was to identify optimal weights for HSG, GAT, and SAAT measures in a linear combination of them, which exhibits maximal reliability (e.g., Bartholomew, 1996; Li, 1997). As has been shown earlier, the composite score produced by the OLC of a given set of unidimensional measures with uncorrelated errors has also maximal predictive validity with respect to any external criterion (Penev & Raykov, 2006). An important practical implication resulting from the motivation behind this empirical study was that the composite scores of HSG, GAT, and SAAT measures with maximal reliability and maximal predictive validity can thus be expected to improve the accuracy and validity of assessments of high school graduates who apply to colleges in Saudi Arabia.

The plan of this article is as follows. In the next section, we provide the theoretical framework for examining the dimensionality of the targeted construct of GAA, the concept of OLC of measures and related maximal reliability, and statistical considerations with focus on approaches to dealing with data analytic and modeling issues that stem from the lack of multivariate normality and missing data. Subsequently, we apply the methodology and procedures outlined in that framework on data of HSGs and scores on the GAT and SAAT scales of high school graduates who apply to colleges with major in Science and colleges with major in Art in Saudi Arabia. Finally, the discussion and conclusion section summarizes the findings in this empirical study and the utility of the discussed methodology on OLC and maximal reliability for educational and behavioral research aimed at measuring instrument construction and development. The Mplus source codes for estimating the pertinent OLC and maximal reliability coefficient as well as its comparison to the reliability of the conventional composite score (overall sum of measures) are provided in Appendixes A and B.

Theoretical Framework

Dimensionality of the Proposed GAA Measure

Scale construction and development using the concepts of OLC and maximal reliability, as available currently for applied research, rests on the assumption of unidimensionality of the instrument of interest with uncorrelated component errors (e.g., Bartholomew, 1996). In this connection, and to address the first goal of the empirical study underlying this article related to the development of a measure of GAA of high school graduates in Saudi Arabia, it is hypothesized that the set of GAA indicators is (essentially) unidimensional, that is, there is one dominant GAA dimension that underlies the students' scores on HSG, GAT, and SAAT (with their verbal and quantitative subscales). To examine this hypothesis, a unidimensional model of GAA is tested for data fit using a confirmatory factor analysis (CFA) within the framework of the popular latent variable modeling (LVM) methodology (e.g., Muthén, 2002; see also Dimitrov, 2012; Raykov & Marcoulides, 2011).

The observed variables that were used as indicators of the hypothesized GAA as a latent variable (construct) were selected on subject-matter grounds from the following set of available measures: (a) HSG; (b) GAT measures on GAT-Verbal, GAT-Quantitative, GAT-total score; and (c) SAAT total score and five SAAT subscales on Biology, Chemistry, Physics, Math, and English. Different subsets of indicators were selected as indicators of the GAA for high school graduates who apply to colleges with major in Science and to colleges with major in Art, with the purpose to take into account differences in educational content, curricula, goals, and so on, between these two types of colleges. Specifically, the following six indicators were used for GAA of applicants to colleges with major in Science: HSG, GAT-total score, and SAAT subscale scores on Biology, Chemistry, Physics, and Math (the SAAT subscale on English is not included here based primarily on substantive considerations). In contrast, the following four indicators were used for GAA of applicants to colleges with major in Art: HSG, GAT-Verbal, GAT-Quantitative, and SAAT-total score. The SAAT subscales were not used with data from Art colleges because of considerations related to the nature of their educational content and curricula.

Optimal Linear Combination of GAA Indicators

The use of the LVM methodology is appropriate here due to several basic results in reliability and validity estimation, which are of special importance to the purpose of this article. A main part of the substantive concern of the empirical study used in the article is the identification of a linear combination of the GAA measures, which is associated with optimal psychometric features, specifically high measurement quality. Accordingly, it was desirable to find a linear combination of these measures that is associated with maximal possible reliability among all linear combinations of them. Such a linear combination, referred to as OLC below, not only possesses maximal reliability, but according to Penev and Raykov (2006) is also associated with maximal criterion validity, that is, maximal correlation with a predefined criterion. In the context of NCA assessment of applicants to colleges, the OLC property of furnishing maximal criterion validity is particularly important for the achievement of high predictive validity of the HSG, GAT, and SAAT measures with respect to an external criterion.

Optimal Linear Combination. The topic of maximal reliability, which is of special importance when developing a multicomponent measuring instrument with high measurement quality, has attracted a great deal of interest among psychometricians and applied researchers since the 1940s (see, e.g., Raykov, 2012, for some related references and details). Specifically, if the single factor model with uncorrelated errors is appropriate for a given data set stemming from p observed measures ($p > 3$), denoted $X_1, \dots, X_p$, with loadings $a_1, \dots, a_p$ and error variances $v_1, \dots, v_p$, respectively, then the OLC, designated Z , is defined as follows:

$$Z = (a_1/v_1)X_1 + (a_2/v_2)X_2 + \dots + (a_p/v_p)X_p \quad (1)$$

(e.g., Bartholomew, 1996).

On fitting a single factor model to GAA indicators and finding it plausible for the analyzed data set, plugging its associated estimated factor loadings and error variances into the right-hand side of Equation 1 renders the OLC with maximal reliability and maximal criterion validity. This OLC is the sought combination of the GAA indicators, which optimally combines the information in them in relation to the underlying construct and as such can be used in subsequent analyses or decisions.

From Equation 1, one obtains after some algebra the maximal reliability coefficient, denoted ρ^* , for the set of measures $X_1, \dots, X_p$ as follows:

$$\rho^* = \left(a_1^2/v_1 + \dots + a_p^2/v_p \right) / \left[1 + \left(a_1^2/v_1 + \dots + a_p^2/v_p \right) \right] \quad (2)$$

(e.g., Li, 1997).

Statistical Considerations

Lack of Multivariate Normality. Fitting of the single-factor model of relevance to the data from the empirical study used in this article requires also attending to several important statistical matters. First, the data on the available observed measures are not multivariate normal. This is because the assumption of multivariate normality does not hold for the set of GAA measures with data from colleges with major in Science or Art. Evidence for this violation is provided by the lack of univariate normality of the HSGs for applicants to either type of colleges. Specifically, although the GAT and SAAT measures are close to normal, the HSG distribution is negatively skewed and deviates markedly from normality (histograms, normality tests, and Q-Q plots were used for all GAT and SAAT measures, but they are not provided here for space consideration).

In general, the multivariate normality assumption is likely to be violated to some extent in empirical educational and behavioral research, and thus an alternative to the popular maximum likelihood (ML) method is needed for the purposes of fitting the model under consideration here. This alternative is provided by the so-called robust maximum likelihood (RML) method of parameter estimation and model testing. The RML method, which has been developed over the past 30 years or so, is readily available in widely circulated software, such as the highly popular LVM program Mplus (Muthén & Muthén, 2012). The essence of this method is the correction of the standard errors for all model parameters (in general) as well as of the chi-square goodness of fit statistic and related indexes of model fit. Details of this correction can be found, for instance, at www.statmodel.com (the website of the software Mplus).

Missing Data. Missing data are almost inevitable in contemporary educational, behavioral, and social research, as well as beyond its boundaries. Missing data implies irreversible loss of information, a loss that an analyst cannot recover from. The

available data set for the empirical study underlying this article is similarly incomplete—a number of students have not provided data on all administered measures. Hence, a method of model fitting and parameter estimation should be used, which “deals” with the missing data aspect of the available data set.

Over the past 40 years or so, statisticians and substantive researchers have been particularly interested in developing methods of “handling” incomplete data sets, like the one on which the following analyses are based (e.g., Little & Rubin, 2002). As a result of these efforts, it has been found that traditional methods, such as listwise deletion (complete case analysis), are not trustworthy in general and unless some very strong assumptions are fulfilled should better be avoided in empirical research. One of the main assumptions in this regard—missing completely at random (MCAR)—is discussed in detail and shown to be in general not testable statistically in Raykov (2011; implications of it, i.e., necessary conditions for MCAR, are testable however; see, e.g., Enders, 2010). Other traditional methods of “dealing” with missing data are not principled (i.e., based on a generally applicable principle or concept), while being essentially ad hoc, developed to only deal with specific situations, and effectively not applicable in the empirical study used in this article.

At the same time, research over the past several decades in the field of missing data analysis has developed two state-of-the-art and principled methods. These are the full information maximum likelihood (FIML) method, which represents ML in the presence of missing data, and multiple imputation (MI). Unlike the traditional approaches to incomplete data analysis, FIML and MI are generally applicable, principled, based on the fundamental concept of likelihood, and associated with optimal statistical properties in general settings. The FIML and MI methods (and specifically the present-day software implementation of MI) assume that data are missing at random (MAR; the latter assumption is not required for the theory of MI, however). This is tantamount to assuming that the propensity (probability) of missingness on any of the GAA measures used in this study is unrelated to the actually missing (and thus unobserved/unavailable) individual student values. The MAR assumption is not testable statistically because data that are needed for its testing are missing, but is effectively made when using FIML or (current software implementations of) MI on an incomplete data set in contemporary educational and behavioral research. In addition, in order to use MI with the data in the empirical study underlying this article, one needs a rather elaborate imputation model that typically includes a larger than the available number of observed measures (indicators) of the targeted GAA construct, which is not the case here.

The alternative method of dealing with missing data, FIML, also requires the assumption of MAR. This assumption, while not testable, could not be thought of as being fulfilled in the available data set. The reason is that it is not unlikely that students with low actual (unobserved/unavailable/missing) scores on any of the collected GAA measures have in fact higher probability of missing on any of them. Thus, the MAR assumption, which is basic for FIML (as well as MI in its

contemporary and widely circulated software implementations), cannot be viewed as holding in the data of this empirical study.

To counteract possible violations of the MAR assumption, a method has been developed over the past 25 years that is based on so-called auxiliary variables (Little & Rubin, 2002; Muthén & Muthén, 2012). This method includes in the model fitting process a set of other variables measured on the studied subjects, which are related to (a) the propensity of missingness and (b) the dependent variables of relevance to this study. (These variables are referred to as “auxiliary variables,” and if there were no missing values the researcher would typically not be interested in including them in a model under consideration, like the single-factor model of interest for this empirical study.) As described earlier, the dependent variables for the data from colleges with major in Science are the HSG, GAT-total score, and four SAAT subscale scores on Biology, Chemistry, Physics, and Math, so the “auxiliary variables” for analyses with these data (from the set of measures available to us) were SAAT-total score and FYGPA, which are closely related to the used GAA indicators and on substantive grounds to the propensity of missingness and actually missing values on them as well. Regarding analyses with data from colleges with major in Art, the dependent variables are the HSG, GAT-Verbal, GAT-Quantitative, and SAAT-total scores, and the “auxiliary variables” are the GAT-total score, FYGPA, and SAAT subscales on Biology, Chemistry, Physics, English, and Math. The effect of including “auxiliary variables” is the enhancement of the plausibility of the important MAR assumption (e.g., Enders, 2010).

Model Fitting. The evaluation of data fit under the CFA employed below is based on a commonly used chi-square test statistic in combination with several other goodness-of-fit indices, most notably the root mean square error of approximation (RMSEA). Given that the chi-square value increases with the increase of the sample size, which results in an artificial tendency to reject model fit, the evaluation of data fit is based on a joint examination of other goodness-of-fit indices such as the comparative fit index (CFI), the Tucker–Lewis index (TLI), and particularly the RMSEA with its 90% confidence interval (CI; see, e.g., Raykov & Marcoulides, 2006).

Results for Colleges With Major in Science

Data

The data on the variables involved in this study for colleges with major in Science came from 6,640 students at universities with major in Science. The means and standard deviations of their six GAA measures are provided in Table 1.

Testing for a Unidimensional GAA for Colleges With Major in Science

The unidimensional model of the targeted construct of general academic ability of high school graduates who apply to colleges with major in Science includes a single

Table 1. Descriptive Statistics of GAA Measures for Colleges With Major in Science ($N = 6,640$).

GAA measures	Min	Max	Mean	SD
High school grade	67.85	100.00	95.96	3.55
GAT-Total score	50.00	97.00	77.37	7.54
SAAT-Biology	1.00	20.00	10.76	3.61
SAAT-Chemistry	1.00	20.00	11.32	4.17
SAAT-Physics	1.00	20.00	11.12	3.54
SAAT-Math	1.00	20.00	8.28	3.43

Note. GAA = general academic ability; GAT = General Aptitude Test; SAAT = Standardized Achievement Admission Test.

latent factor, GAA, with six indicators as noted earlier: HSG, GAT-total score, and four SAAT subscale scores on Biology, Chemistry, Physics, and Math. The path diagram of this model is depicted in Figure 1. Testing of this model was performed using Mplus (Muthén & Muthén, 2012), with the SAAT-total score and FYGPA employed in the role of “auxiliary variables” as indicated in the preceding section, as a means of “dealing” with missing data and specifically likely violations of the MAR assumption (e.g., Enders, 2010).

The values of the goodness-of-fit indices demonstrated tenable data fit of the single-factor GAA model. Specifically, although the chi-square value was statistically significant, $\chi^2(9) = 226.40, p < .001$, which is not a surprise given the large sample size ($N = 6,640$), the values of the other fit indices were in the tenable range and as follows: CFI = .980, TLI = .967, and RMSEA = .060, with a 90% confidence interval (0.054, 0.067) (e.g., Hu & Bentler, 1999). The unstandardized factor loadings and residual variances, with their standard errors, are provided in Table 2. Based on these results, it is plausible that the general academic ability under the CFA model in Figure 1 is essentially unidimensional, that is, there is one dominant dimension, GAA, that underlies the student scores on six observed measures selected to serve as indicators of the construct for high school graduates who apply to colleges with major in Science. This finding is critical for scaling, scoring, and other psychometric procedures with the GAA and its measures, including the identification of an OLC that furnishes maximal reliability and criterion validity of the respective composite score of the GAA measures used.

Optimal Linear Combination of GAA Measures for Colleges With Major in Science

With the six GAA measures under the unidimensional model in Figure 1, as elaborated in the preceding subsection, their OLC is obtained via Equation 1. That is, with Z denoting the sought OLC, Equation 1 becomes the following:

Figure 1. CFA model of GAA of high school applicants to colleges with major in Science.
Note. CFA = confirmatory factor analysis; GAA = general academic ability; GAT = General Aptitude Test; SAAT = Standardized Achievement Admission Test.

$$Z = (a_1/v_1)X_1 + (a_2/v_2)X_2 + (a_3/v_3)X_3 + (a_4/v_4)X_4 + (a_5/v_5)X_5 + (a_6/v_6)X_6, \quad (3)$$

where X_1 = HSG, X_2 = GAT-total score, X_3 = SAAT-Biology, X_4 = SAAT-Chemistry, X_5 = SAAT-Physics, and X_6 = SAAT-Math. By replacing the factor loadings $a_1, \dots, a_6$ and error variances $v_1, \dots, v_6$ in Equation 3 with their CFA estimates (see Table 2), we obtain the following OLC of the six measures of GAA for applicants to colleges with major in Science:

$$Z = 0.112(\text{HSG}) + 0.071(\text{GAT}) + 0.523(\text{SAAT-Biology}) + 0.598(\text{SAAT-Chemistry}) + 0.693(\text{SAAT-Physics}) + 0.240(\text{SAAT-Math}). \quad (4)$$

As discussed earlier, the composite score obtained with the OLC in Equation 4 is associated with maximal reliability and maximal criterion validity (for any criterion variable used). The Mplus code in Appendix A, which was used to obtain the OLC weights in Equation 4 and the maximal reliability of the OLC composite score (see below), can also be used to obtain the reliability of the common composite (overall sum) score:

$$Z^* = \text{HSG} + \text{GAT} + \text{SAAT-Biology} + \text{SAAT-Chemistry} + \text{SAAT-Physics} + \text{SAAT-Math}. \quad (5)$$

Table 2. Estimates (With Standard Errors) of Factor Loadings and Error Variances Under the CFA Model of GAA for Applicants to Colleges With Major in Science.

GAA indicators	Loadings, α	SE(α)	Error variances, ν	SE(ν)
High school grade	1.239	0.046	11.069	0.334
GAT-Total Score	3.281	0.102	46.214	0.822
SAAT-Biology	2.787	0.038	5.332	0.126
SAAT-Chemistry	3.421	0.038	5.721	0.157
SAAT-Physics	2.896	0.037	4.179	0.111
SAAT-Math	1.929	0.045	8.032	0.199

Note. CFA = confirmatory factor analysis; GAA = general academic ability; GAT = General Aptitude Test; SAAT = Standardized Achievement Admission Test. All estimates of factor loadings and error variances are statistically significant ($p < .001$).

Accordingly, the resulting estimates of reliability for the composite score under the OLC in Equation 4 and the common composite score under Equation 5 were found to be .864 and .750, respectively. In addition, 90%, 95%, and 99% CIs for the reliability estimates under Equations 4 and 5, as well as for their difference (denoted *delta* below), are also obtained with the same analysis (see later part of the Mplus source code in Appendix A). For example, the difference between the two reliability estimates under Equations 4 and 5, respectively, was $\text{delta} = 0.114$, with 95% CI being (0.106, 0.120). This result can be interpreted as indicating that, at a 95% confidence level, the maximal reliability under the OLC in Equation 4 exceeds the reliability of the composite score under Equation 5 by a magnitude that varies from 0.106 to 0.120 in the population of students applying to colleges with major in Science. The theoretical justification and technical details on constructing a confidence interval of the difference (*delta*) between the OLC reliability (i.e., the maximal reliability) and the popular composite reliability coefficient, are provided with another study (Raykov, Gabler, & Dimitrov, 2014).

Results for Colleges With Major in Art

Data

The data on the variables involved in this study for colleges with major in Art came from 3,388 students at universities with major in Art. The range, means, and standard deviations of the four GAA measures used here are provided in Table 3.

The assumption of multivariate normality was similarly not met with these data. As was the case with data from students at colleges with major in Science, the univariate distribution of the HSG measure is negatively skewed, and so we similarly use below the RML method of parameter estimation and model testing.

Table 3. Descriptive Statistics of GAA Measures for Colleges With Major in Art (*N* = 3,388).

GAA measures	Min	Max	Mean	SD
High school grade	58	100	88.19	7.95
GAT-Verbal	43.40	94.50	69.30	8.75
GAT-Quantitative	45.70	100	68.45	8.52
SAAT-Total Score	52	90	66.66	5.91

Note. GAA = general academic ability; GAT = General Aptitude Test; SAAT = Standardized Achievement Admission Test.

Figure 2. CFA model of GAA of high school applicants to colleges with major in Art.

Note. CFA = confirmatory factor analysis; GAA = general academic ability; GAT = General Aptitude Test; SAAT = Standardized Achievement Admission Test.

Testing for Unidimensional GAA for Colleges With Major in Art

The unidimensional model for general academic ability for applicants to colleges with major in Art is also based on a single latent factor, GAA, with four indicators as noted earlier: HSG, GAT-Verbal score, GAT-Quantitative score, and SAAT-total score. The path diagram of this model is depicted in Figure 2. Fitting the model to the pertinent incomplete data set used similarly seven “auxiliary variables,” namely, GAT-total score, SAAT-Biology, SAAT-Chemistry, SAAT-Physics, SAAT-English, SAAT-Math, and FYGPA.

The values of the goodness-of-fit indices demonstrated tenable fit of this single-factor GAA model. Specifically, $\chi^2(2) = 8.00$, $p = .018$, CFI = .997, TLI = .992, and RMSEA = .030, with 90% CI of (0.010, 0.053). The unstandardized factor loadings and residual variances, with their standard errors, are provided in Table 4. Thus, the hypothesized GAA is essentially unidimensional, that is, the construct of GAA is

Table 4. Estimates (With Standard Errors) of Factor Loadings and Error Variances Under the CFA Model of GAA for Applicants to Colleges With Major in Art.

GAA indicators	Loadings, a	$SE(a)$	Error variances, ν	$SE(\nu)$
High school grade	4.187	0.213	47.959	1.535
GAT-Verbal	7.241	0.171	25.564	1.356
GAT-Quantitative	6.637	0.174	29.955	1.282
SAAT-Total Score	3.789	0.183	23.227	1.124

Note. CFA = confirmatory factor analysis; GAA = general academic ability; GAT = General Aptitude Test; SAAT = Standardized Achievement Admission Test. All estimates of factor loadings and error variances are statistically significant ($p < .001$).

dominated by one dimension that underlies the student scores on the four observed measures selected to serve as indicators of the construct for high school graduates who apply to colleges with major in Art. This finding is also critical for scaling, scoring, and other psychometric procedures with the GAA and its measures, including the identification of an OLC that furnishes maximal reliability and criterion validity of the respective composite score of the four GAA measures for colleges with major in Art.

Optimal Linear Combination of GAA Measures for Colleges With Major in Art

For the four GAA measures under the unidimensional model (see Figure 2), the OLC is obtained again via Equation 1. That is, with Z denoting the pertinent OLC, Equation 1 becomes the following:

$$Z = (a_1/\nu_1)(HSG) + (a_2/\nu_2)(GATV) + (a_3/\nu_3)(GATQ) + (a_4/\nu_4)(SAAT), \tag{6}$$

where HSG = high school grade, GATV = GAT-Verbal, GATQ = GAT-Quantitative, and SAAT = SAAT-total score. By replacing the factor loadings $a_1, \dots, a_4$ and error variances $\nu_1, \dots, \nu_4$ in Equation 6 with their estimates in Table 4, we obtain the following OLC of the four measures of GAA for applicants to colleges with major in Art:

$$Z = (0.087)(HSG) + (0.283)(GATV) + (0.222)(GATQ) + (0.163)(SAAT). \tag{7}$$

As noted earlier, the composite score obtained with the OLC in Equation 7 is associated with maximal reliability and maximal criterion validity (for any criterion variable used). Similarly, along the maximal reliability of the OLC composite score, we can also estimate the reliability of the common (overall sum) composite score

$$Z^* = HSG + GATV + GATQ + SAAT. \tag{8}$$

Using the Mplus code in Appendix B, the estimates of maximal reliability and composite reliability, that is, for the composite score under the OLC in Equation 7 and for the composite score under Equation 8, were found to be 0.818 and 0.790, respectively. In addition, 90%, 95%, and 99% CIs for the reliability estimates under Equations 7 and 8, as well as for their difference, *delta*, were also obtained. For example, the difference between the two reliability estimates under Equations 7 and 8, respectively, was $\delta = 0.028$, with 95% CI of (0.022, 0.033). This result can be interpreted as indicating that, at a 95% CI confidence level, the maximal reliability under the OLC in Equation 7 exceeds the reliability of the conventional composite score under Equation 8 by a magnitude that varies from 0.022 to 0.033 in the population of students applying to colleges with major in Art.

Discussion and Conclusion

This article was concerned with an application of the concepts of maximal reliability and OLC to the development of a measure of general academic ability for high school graduates who apply to colleges in Saudi Arabia. The indicators of the GAA construct were selected from measures that included the students' HSG and their scores on two tests developed and administered by the National Center for Assessment in Higher Education in Saudi Arabia, GAT and SAAT. For colleges with major in Science, appropriate indicators of GAA were the students' HSG, total score on GAT, and four SAAT subscales on Biology, Chemistry, Physics, and Math. For colleges with major in Art, appropriate measures of GAA were the students' HSG and their scores on GAT-Verbal, GAT-Quantitative, and SAAT-total score. The selection of GAA indicators was based on substantive considerations related to the educational content and curricula of colleges with major in Science and colleges with major in Art.

The OLC of GAA indicators that furnishes maximal reliability and maximal predictive validity was identified separately for colleges with major in Science and for colleges with major in Art. Specifically, the OLC of the six GAA indicators for colleges with major in Science is provided in Equation 4, whereas the OLC of the four GAA indicators for colleges with major in Art is provided in Equation 7. In either case, the composite score associated with the respective OLC has maximal reliability and maximal predictive validity in regard to any external criterion. This feature of the resulting optimal composite score is particularly important with respect to the validity of GAT and SAAT scores used for admission of high school graduates to colleges. The source codes with the popular LVM software Mplus (Muthén & Muthén, 2012), which is employed for the estimation of the coefficients (weights) in the OLC of GAA indicators, are provided in Appendixes A and B, respectively, and can be straightforwardly used in empirical educational and behavioral research aimed at construction and development of a measurement instrument.

In conclusion, this article discussed a readily and widely applicable methodology that can significantly help researchers and practitioners in improving the efficiency and measurement quality of assessments based on multiple-component measuring

instruments, in particular scales employed for the purpose of admission of high school graduates to colleges and universities. Although the case study is Saudi Arabia, the methods and procedures discussed in this article have broader utility and can be used in different contexts of educational and psychological assessment.

Appendix A

Mplus Code for the Computation of Weights of the Optimal Linear Combination of GAA Indicators (Equation 4), the Respective Maximal Reliability, the Reliability of the Conventional Composite Score (Equation 5), and the Difference (Delta) Between These Two Reliabilities, With 90%, 95%, and 99% Confidence Intervals for Colleges With Major in Science.

```

TITLE: Assessment of maximal reliability for optimal linear
       combination (OLC) of GAA indicators, reliability of
       the sum of these indicators, and the difference (delta)
       between the two reliabilities for high school graduates
       who apply to colleges with major in Science in KSA

DATA: FILE IS "C:\GAA.dat";

VARIABLE: NAMES ARE id univ gender HSGPA SAAT GPA1 GATV GATQ GAT
           SAATbio SAATchem SAATphys SAATEngl SAATmath;
           USEVARIABLES ARE GAT HSGPA SAATbio SAATchem SAATphys SAATmath;
           AUXILIARY = (m) SAAT GPA1;
           MISSING = ALL(-999);

ANALYSIS: ESTIMATOR = MLR;

MODEL:   GAA BY GAT* (L1)
         HSGPA (L2)
         SAATbio SAATchem SAATphys SAATmath (L3-L6);
         GAA@1;
         GAT (EV1); HSGPA (EV2);
         SAATbio(EV3); SAATchem (EV4); SAATphys (EV5);
         SAATmath (EV6);

MODEL CONSTRAINT: NEW(W1 W2 W3 W4 W5 W6 OCLREL CREL delta);
                  W1 = L1/EV1; W2 = L2/EV2; W3 = L3/EV3; W4 = L4/EV4;
                  W5 = L5/EV5; W6 = L6/EV6;
                  OCLREL = (L1^2/EV1+L2^2/EV2+L3^2/EV3+L4^2/EV4+L5^2/EV5+L6^2/EV6)
                           / (1+L1^2/EV1+L2^2/EV2+L3^2/EV3+L4^2/EV4+L5^2/EV5+L6^2/EV6);
                  CREL = (L1+L2+L3+L4+L5+L6)^2/
                           ((L1+L2+L3+L4+L5+L6)^2 + EV1+EV2+EV3+EV4+EV5+EV6);
                  delta = OCLREL - CREL;

OUTPUT: CINTERVAL;
```

Note. The input ASCII file (GAA.dat) contains the following variables: HSG = High School Grade, GAT = GAT-total score, SAAT = SAAT-total score, GPA1 = First-Year GPA in college, GATV = GAT-Verbal, GATQ = GAT-Quantitative, SAATbio = SAAT-Biology, SAATchem = SAAT-Chemistry, SAATphys = SAAT-Physics, SAATEngl = SAAT-English, and SAATmath = SAAT-Math.

Appendix B

Mplus Code for the Computation of Weights of the Optimal Linear Combination of GAA Indicators (Equation 7), the Respective Maximal Reliability, the Reliability of the Conventional Composite Score (Equation 8), and the Difference (Delta) Between These Two Reliabilities, With 90%, 95%, and 99% Confidence Intervals for Colleges With Major in Art.

```
=====
TITLE: Assessment of maximal reliability for optimal linear
       combinaion(OLC) of GAA indicators, reliabiliy of
       the sum of these indicators, and the difference (delta)
       between the two reliabillites for high school graduates
       who apply to colleges with major in Art in KSA

DATA: FILE IS "C:\GAA_ART.dat";

VARIABLE: NAMES ARE id univ gender HStype HSGPA GAT SAAT GPAPrep GPA1
            GATV GATQ SAATbio SAATchem SAATphys SAATEngl SAATmath;
            USEVARIABLES ARE GATV GATQ HSGPA SAAT;
            AUXILIARY = (m) GAT SAATbio SAATchem SAATphys SAATEngl SAATmath GPA1;
            MISSING = ALL(-999);

ANALYSIS: ESTIMATOR = MLR;

MODEL: GAA BY GATV* (L1)
        GATQ (L2)
        HSGPA (L3)
        SAAT (L4);
        GAA@1;
        GATV (EV1); GATQ (EV2); HSGPA (EV3); SAAT (EV4);

MODEL CONSTRAINT: NEW(W1-W4 OCLREL SUMREL delta);
                  W1 = L1/EV1; W2 = L2/EV2; W3 = L3/EV3; W4 = L4/EV4;
                  OCLREL = (L1^2/EV1+L2^2/EV2+L3^2/EV3+L4^2/EV4)
                  / (1+L1^2/EV1+L2^2/EV2+L3^2/EV3+L4^2/EV4);
                  SUMREL = (L1+L2+L3+L4)^2 / ((L1+L2+L3+L4)^2+EV1+EV2+EV3+EV4);
                  delta = OCLREL - SUMREL;

OUTPUT: CINTERVAL;
=====
```

Note. The input ASCII file (GAA.dat) contains the following variables: HSG = High School Grade, GAT = GAT-total score, SAAT = SAAT-total score, GPA1 = First-Year GPA in college, GATV = GAT-Verbal, GATQ = GAT-Quantitative, SAATbio = SAAT-Biology, SAATchem = SAAT-Chemistry, SAATphys = SAAT-Physics, SAATEngl = SAAT-English, and SAATmath = SAAT-Math.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Bartholomew, D. J. (1996). *The statistical approach to social measurement*. New York, NY: Academic Press.
- Dimitrov, D. M. (2012). *Statistical methods for validation of assessment scale data in counseling and related fields*. Alexandria, VA: American Counseling Association.
- Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Li, H. (1997). A unifying expression for the maximal reliability of a linear composite. *Psychometrika*, 62, 245-249.
- Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data*. New York, NY: Wiley.
- Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, 29, 87-117.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
- Penev, S., & Raykov, T. (2006). On the relationship between maximal reliability and maximal validity for linear composites. *Multivariate Behavioral Research*, 41, 105-126.
- Raykov, T. (2011). On testability of missing data mechanisms in incomplete data sets. *Structural Equation Modeling*, 18, 419-430.
- Raykov, T. (2012). Scale development using structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 472-492). New York, NY: Guilford Press.
- Raykov, T., Gabler, S., & Dimitrov, D. M. (2014). Maximal reliability and composite reliability: Estimating their difference for multi-component measuring instrument using latent variable modeling. Manuscript submitted for publication.
- Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling*. Mahwah, NJ: Erlbaum.
- Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York, NY: Taylor & Francis.