The Total Score With Maximal Reliability and Maximal Criterion Validity: An Illustration Using a Career Satisfaction Measure

Educational and Psychological
Measurement

2018, Vol. 78(6) 1108–1122

© The Author(s) 2017

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164417738564

journals.sagepub.com/home/epm

Yuanshu Fu¹, Zhonglin Wen¹, and Yang Wang¹

Abstract

The maximal reliability of a congeneric measure is achieved by weighting item scores to form the optimal linear combination as the total score; it is never lower than the composite reliability of the measure when measurement errors are uncorrelated. The statistical method that renders maximal reliability would also lead to maximal criterion validity. Using a career satisfaction measure as an example, the present article calculated the maximal reliability and maximal criterion validity and compared them with the composite reliability and the scale criterion validity, respectively. The improvement of reliability and validity indicated that the optimal linear combination is preferred when forming a total score of a measure. The Mplus codes for analyzing maximal reliability, maximal criterion validity, and related parameters are provided.

Keywords

optimal linear combination, maximal reliability, maximal criterion validity

Introduction

The development of psychometric instruments with high measurement quality always concerns methodologists as well as empirical researchers (Dimitrov, Raykov, & Alqataee, 2015). Since reliability and validity are two basic concepts related to the quality of measuring instruments (also referred to as measure or scale in this article),

¹Guangdong Key Laboratory of Mental Health and Cognitive Science/Center for Studies of Psychological Application/School of Psychology, South China Normal University, Guangzhou, China

Corresponding Author:

Zhonglin Wen, School of Psychology, South China Normal University, Guangzhou, 510631, China.

Email: wenzl@scnu.edu.cn

they have received an impressive amount of attention over the past century (Penev & Raykov, 2006). Reliability simply describes the “consistency” of a measure, while validity estimates how “well” a test measures what it purports to measure (Cohen & Swerdlik, 2009). Validity is the ultimate goal for examining measurement quality; meanwhile, we also require reliability to reach a satisfactory level in order to make a valid test. Sometimes the relationship between reliability and validity appear to be paradoxical, and it is likely that a measure with lower reliability is preferred because of higher validity (Guilford, 1954; Li, 2003): Validity is a more important concept than reliability. However, over the past few decades, methodological researchers have shown much more interest in the estimation of reliability rather than validity.

The discussion of reliability estimation focuses predominantly on linear composites of sets of measures that provide multiple converging sources of information about latent constructs of interest. Often, researchers have been preoccupied with the overall sum score reliability, frequently referred to as the composite reliability (unit-weighted or unweighted reliability). An alternative approach to estimate reliability is associated with the concept of the optimal linear combination (OLC), which renders maximal reliability. Maximal reliability is achieved by giving each item score an optimal weight. The linear combination formed by these weighted total scores, which result in maximal reliability, is called the OLC. The OLC furnishes maximal reliability that is never lower than the composite reliability (Li, 1997; Li, Rosenthal, & Rubin, 1996; Raykov, 2004; Raykov, Gabler, & Dimitrov, 2015).

The aim of maximal reliability is not simply to achieve reliability to a maximum value; it is also to identify a linear combination of items that minimize the effect of measurement errors on the total score. Traditionally, we obtain the linear combination of an overall sum score by assigning unit weights to each item, ignoring the fact that each item contributes differently to the common latent construct because the factor loadings of items are rarely the same. In fact, an infinite number of linear combinations can be derived from giving various weights to individual items, and the OLC is the combination that allows the reliability of the total score of a measure reach its maximum. The core value of the OLC and the associated maximal reliability is not about finding an “upper bound” for reliability; rather, it provides a total score in the shape of a linear combination that is affected by measurement error to the least possible degree.

Currently, the discussion of the OLC and the associated maximal reliability remains mostly theoretical. Only a few empirical studies have been conducted applying these concepts. For example, Willoughby, Pek, and Blair (2013) used maximal reliability in an attempt to develop an optimal short form measure of executive function. Dimitrov et al. (2015) applied maximal reliability during the process of developing a measure of general academic ability, showing that maximal reliability is a method that can significantly help researchers improve the efficacy as well as measurement quality. The lack of applied research addressing maximal reliability may be partly due to the confusing concept of weighted sums of measurement items under OLC and partly due to the lack of readily accessible approaches to the process of measurement quality evaluation. Nonetheless, given that the maximal reliability is

never lower than the composite reliability, it is a method worth introducing, as it would improve scale constructs to some extent in most situations.

Validity, which is at least as important as reliability, is a rather complex concept that is often reflected by accumulated evidence rather than a single indicator (Anastasi, 1988). Criterion validity is a major type of evidence that is frequently of interest in empirical research (McDonald, 1999). Penev and Raykov (2006) demonstrated that the weighted total score yielding maximal reliability also provided maximal criterion validity for congeneric measures if the measurement errors are uncorrelated among themselves, with the latent construct and with the criterion variable. In other words, the OLC possesses the highest validity as well as reliability if the above-mentioned assumptions are satisfied. However, there is a severe lack of empirical research on the estimation of differences between the criterion validity achieved by the OLC and the simple total score.

This article aims to help bridge 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 OLC and use it on a developed congeneric measure, *career satisfaction* (CS). Career satisfaction is an indicator of subjective career success, and it signifies the satisfaction that an employee derives from the intrinsic and extrinsic aspects of his or her career (Karatepe & Olugbade, 2016; Kong, Cheung, & Song, 2012). The career satisfaction measure was developed by Greenhaus, Parasuraman, and Wormley (1990) with five items (sample item: "I am satisfied with the success that I have achieved in my career"). Career satisfaction is an important variable in organizational psychology and has been found to be closely related to numerous factors, including work support (Armstrong-Stassen & Ursel, 2009; Karatepe, 2012; Karatepe & Olugbade, 2016), performance (Armstrong-Stassen & Ursel, 2009; Karatepe, 2012), turnover intentions (Karatepe & Olugbade, 2016), and perceived career success (Eby, Butts, & Lockwood, 2003).

Two criteria were used to examine the concurrent validity of the career satisfaction measure: (a) *Internal marketability* (IN), which is a type of perceived career success, adopted from Eby et al. (2003) that contains three items (sample item: "My company views me as an asset to the organization") and (b) *Possibilities for professional development* (PD), which is included in organizational support, developed by Bakker, Demerouti, Taris, Schaufeli, and Schreurs (2003) and contains four items (sample item: "My work offers me the opportunity to learn new things"). Responses to these three measures and their respective items were made on a 5-point scale (1 = *totally disagree*, 5 = *totally agree*).

The present study used CS as an example to show how maximal reliability and maximal criterion validity could be implemented in realistic settings. The purpose of the present study was twofold: (a) to estimate reliability and criterion validity using both unweighted and weighted total scores in a developed measure and evaluate their differences and (b) to provide readily applicable procedures (including Mplus codes) for maximal reliability and maximal criterion validity estimation. An important empirical implication resulting from this study was to demonstrate that it is possible

to improve the accuracy of a measurement by using statistical methods such as assigning weights to item scores.

Our study is based on four assumptions: (a) all scales used in this study are unidimensional and measure a single latent construct; (b) measurement errors are uncorrelated among themselves; (c) measurement errors are uncorrelated with the latent construct; (d) the error terms are uncorrelated with the criterion.

Optimal Linear Combination of Measurement Indicators

A substantive concern of this empirical study was to identify a linear combination of the career satisfaction measure that 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 the maximal possible reliability among all linear combinations. Such a linear combination, which is referred to as the OLC, not only possesses maximal reliability but is also associated with maximal criterion validity for unidimensional measures under the condition that the predefined criterion is uncorrelated with measurement errors. This condition is easily tested using latent variable modeling (Penev & Raykov, 2006).

The idea of weighting item scores in order to improve the quality of multi-item scales can be dated back to the 1940s (Mosier, 1943) and has continued to attract the attention of psychometricians (Gabler & Raykov, 2016¹; Li, 1997; Li et al., 1996; Penev & Raykov, 2006; Raykov, 2004; Raykov, Gabler, & Dimitrov, 2016). Given that the concepts of OLC and maximal reliability are rather unfamiliar for most readers, the theoretical background is explained in detail below.

Assume that a set of (approximately) continuous items $x_1, x_2, \dots, x_p$ ($p > 3$) are the indicators of a unidimensional psychometric measure (the common latent construct). Suppose these p items measure a latent variable ξ , and $\delta_1, \delta_2, \dots, \delta_p$ are measurement errors of $x_1, x_2, \dots, x_p$. The confirmatory factor analysis (CFA) model with intercepts can be expressed as follows (e.g., Jöreskog, 1971):

$$x_i = \tau_i + \lambda_i \xi + \delta_i \quad (i = 1, 2, \dots, p) \quad (1)$$

where λ_i is the factor loading of item i on the latent variable ξ , τ_i is the associated intercept, and δ_i is the error term of item i . These error terms are assumed to be uncorrelated among themselves as well as with the latent variable ξ throughout this article. To identify the measurement model, we fix the factor variance at one ($\text{var}(\xi) = 1$). For an unstandardized solution, let the item loadings be $\lambda_1, \lambda_2, \dots, \lambda_p$ and error variances be $\theta_1, \theta_2, \dots, \theta_p$ ($\theta_i = \text{var}(\delta_i)$), respectively.

It is well known that the composite score, denoted as Z , with p items for a single factor model is calculated as follows:

$$Z = X_1 + X_2 + \dots + X_p \quad (2)$$

The composite reliability coefficient for a congeneric measure is as follows (e.g., Brunner & Süb, 2005; McDonald, 1999; Raykov & Grayson, 2003):

$$\rho = \frac{(\lambda_1 + \lambda_2 + \dots + \lambda_p)^2}{[(\lambda_1 + \lambda_2 + \dots + \lambda_p)^2 + \theta_1 + \theta_2 + \dots + \theta_p]} \quad (3)$$

The OLC of the measure is defined as the combination that possesses the highest reliability among the possible linear combinations,

$$Z^* = w_1X_1 + w_2X_2 + \dots + w_pX_p \quad (4)$$

The optimal weight is identified as follows (Bartholomew, 1996; Raykov et al., 2015):

$$w_i = \frac{\lambda_i}{\theta_i} \quad (i = 1, 2, \dots, p) \quad (5)$$

After combining Equations (4) and (5), the OLC (weighted total score) with maximal reliability is as follows:

$$Z^* = \frac{\lambda_1}{\theta_1}X_1 + \frac{\lambda_2}{\theta_2}X_2 + \dots + \frac{\lambda_p}{\theta_p}X_p \quad (6)$$

and the population maximal reliability, denoted as ρ^* , is obtained (Li, 1997; Raykov et al., 2015):

$$\rho^* = \frac{\frac{\lambda_1^2}{\theta_1} + \frac{\lambda_2^2}{\theta_2} + \dots + \frac{\lambda_p^2}{\theta_p}}{\left[1 + \left(\frac{\lambda_1^2}{\theta_1} + \frac{\lambda_2^2}{\theta_2} + \dots + \frac{\lambda_p^2}{\theta_p}\right)\right]} \quad (7)$$

It is important to note that $\rho \leq \rho^*$ always holds. That is, there is no linear combination such as Equation (4) that has reliability greater than ρ^* . In particular, the reliability coefficient achieved from the unit-weighted linear combination in Equation (2) is always lower than, if not equal to, that achieved from OLC.

Furthermore, for any criterion C that is uncorrelated with the measurement errors, maximal criterion validity is also obtained under OLC if and only if the measure is congeneric (Penev & Raykov, 2006).

An Example

Data

The data in the present study came from 475 middle school teachers in China. The basic descriptive statistical information for the indicators is provided in Table 1. The absolute value of skewness is between 0.10 and 0.74 and that of kurtosis is between 0.09 and 1.11, suggesting that the multivariate normality assumption was not

Table 1. Descriptive Statistics of Career Satisfaction, Possibilities for Professional Development, and Internal Marketability Measures.

Indicators	M	SD	Skewness	Kurtosis	ω
Career satisfaction					
Item 1	3.61	1.04	−0.74	−0.09	
Item 2	3.03	1.16	−0.10	−0.90	
Item 3	3.52	1.04	−0.54	−0.50	
Item 4	3.47	0.95	−0.67	−0.22	
Item 5	2.79	1.14	−0.60	−1.11	
Professional development	14.06	3.01	−0.21	−0.31	0.72
Internal marketability	10.16	2.27	−0.58	0.66	0.70

Note. ω is omega coefficient (McDonald, 1970) used to calculate the composite reliability of measurements (Raykov & Grayson, 2003). The reliability of career satisfaction is of major interest in the present article and will be discussed further.

violated. Therefore, we adopted the popular maximum likelihood (ML) method for parameter estimation (Finney & DiStefano, 2006; West, Finch, & Curran, 1995). The basic descriptive statistical analysis (mean, *SD*, skewness, and kurtosis) was done using SPSS 23. All other analyses were done using Mplus 7.0 (Muthén & Muthén, 1998-2015).

Testing for Unidimensionality of Career Satisfaction

Career satisfaction is a congeneric scale that has been applied in empirical research for over 20 years. Since the assumption of unidimensionality is rather strong when applying the OLC, we first tested whether the measurement model fit our data using CFA.

The values of the goodness-of-fit indexes demonstrated a tenable data fit of the single-factor model: $\chi^2/df = 2.482$, comparative fit index (CFI) = 0.988, Tucker-Lewis index (TLI) = 0.976, and root mean square error of approximation (RMSEA) = 0.056, with a 90% confidence interval (0.016, 0.096), indicating that the model was acceptable (Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004). Equivalence testing was also adopted to provide additional information on the model fit. The T-size RMSEA (RMSEA_t) in equivalence testing was 0.096 and the T-size CFI (CFI_t) was 0.953, which indicates 95% confidence that the population CFI is above 0.953 and that the size of misspecification is no more than 0.096. By comparing these T-size results and the relative cutoff values, it was possible to distinguish the model fit between poor, mediocre, fair, close and excellent. The results showed that RMSEA_t achieved fair fit and CFI_t suggested close fit. Equivalence testing further proved that the data were well fitted into the one-factor model. For detailed mathematical proofs and application procedures, readers should refer to Marcoulides and Yuan (2016) and Yuan, Chan, Marcoulides, and Bentler (2015).

Table 2. Estimates of Factor Loadings and Error Variances Under the Unidimensional Confirmatory Factor Analysis Model.

Indicators	Loadings, a	$SE(a)$	Error variances	$SE(v)$
1	0.715	0.044	0.480	0.042
2	0.719	0.053	0.828	0.064
3	0.822	0.045	0.398	0.043
4	0.642	0.043	0.494	0.040
5	0.493	0.055	1.042	0.072

According to the CFA model, all of the standardized factor loadings were higher than 0.3 ($p < .001$). The unstandardized factor loadings and residual variances, along with their standard errors, are provided in Table 2. Based on these results, it is plausible that career satisfaction under the CFA model is essentially unidimensional. This finding is crucial for the identification of an OLC that furnishes maximal reliability and maximal criterion validity, two concepts that we will demonstrate in the following subsections.

OLC for the Measure

The OLC for the five indicators under the unidimensional model is obtained via Equation (6), as described in the preceding subsection. That is, with Z^* denoting the OLC, Equation (6) becomes the following:

$$Z^* = \left(\frac{\lambda_1}{\theta_1}\right)X_1 + \left(\frac{\lambda_2}{\theta_2}\right)X_2 + \left(\frac{\lambda_3}{\theta_3}\right)X_3 + \left(\frac{\lambda_4}{\theta_4}\right)X_4 + \left(\frac{\lambda_5}{\theta_5}\right)X_5 \quad (8)$$

By replacing the factor loadings $\lambda_1, \dots, \lambda_5$ and error variances $\theta_1, \dots, \theta_5$ in Equation (8) with the values in Table 2, we obtain the following OLC of the five indicators of career satisfaction:

$$Z^* = 1.490X_1 + 0.868X_2 + 2.065X_3 + 1.300X_4 + 0.473X_5 \quad (9)$$

It is worth noting that the ranking of the 5 weights is $w_3 > w_1 > w_4 > w_2 > w_5$. The item weights take factor loadings and error variances of the indicators into account. They are ordered according to their relative consistency and contribution to the total score.

Testing for Maximal Reliability

As discussed earlier, the weighted score obtained with the OLC in Equation (9) is associated with maximal reliability and maximal criterion validity (for any criterion variable used). The Mplus code is provided in Appendix A; this produced the OLC

Figure 1. CFA models for criterion correlation.
Note. CS = career satisfaction; IN = internal marketability; INs = total score of internal marketability; PD = possibilities for professional development; PDs = total score of possibilities for professional development.

weights in Equation (9), the maximal reliability, and the reliability of the common composite (unit-weight) score:

$$Z = X_1 + X_2 + X_3 + X_4 + X_5 \tag{10}$$

Accordingly, the resulting estimates of reliability for the weighted score under the OLC in Equation (9) and the common composite in Equation (10) were found to be 0.817 and 0.780, respectively. The difference was *delta* = 0.037, with the 95% confidence interval being (0.020, 0.050). Therefore, the maximal reliability was higher than the composite reliability in this case. The theoretical justification of their differences is provided in other studies (Li, 1997; Li et al., 1996; Raykov et al., 2015).

Testing for Maximal Criterion Validity

To test the maximal criterion validity of two criterion measures, we needed to know that the criterion was uncorrelated with measurement errors. We examined the CFA model, which assumed that the measurement errors were uncorrelated with the latent variable in the first place. The tested CFA models are shown in Figure 1. Then, we examined the difference between maximal criterion validity and scale criterion validity. The Mplus code for the maximal criterion validity estimation is given in Appendix B.

When using *internal marketability* as a criterion, as Figure 1a, the model fit indexes provided reasonable results: $\chi^2/df = 2.69$, CFI = 0.982, TLI = 0.970, and RMSEA = 0.06, with a 95% CI (0.032, 0.089). Equivalence testing showed that $RMSEA_t = 0.089$ and $CFI_t = 0.948$, which reflected fair fit and close fit, respectively. We concluded from the moderate fit indexes that the criterion was not correlated with measurement errors, thereby justifying the following analysis of the maximal validity coefficient. The maximal and scale validity coefficients were 0.591 and 0.581, respectively, and the difference was $\delta = 0.010$, with the 95% CI being (0.006, 0.015).

When using *possibilities for professional development* as the criterion, the same procedure was applied. The CFA result for Figure 1b also showed moderate fit, $\chi^2/df = 2.179$, CFI = 0.987, TLI = 0.979, and RMSEA = 0.05, with a 95% CI (0.019, 0.080). Equivalence testing results indicated that both T-size CFI and RMSEA reached close fit, with $RMSEA_t = 0.080$ and $CFI_t = 0.957$. The maximal criterion validity and scale criterion validity were 0.589 and 0.576, respectively, with $\delta = 0.013$. The 95% CI was (0.007, 0.018). It is worth noting that the criterion validity referred to the correlation between the latent variable (CS) and the specified criterion (PD/IN; Raykov et al., 2016). Therefore, the validity coefficients could reflect the correlation without the interference of measurement errors.

Discussion and Conclusion

The present article was primarily concerned with the application of maximal reliability and maximal criterion validity in realistic settings. With this in mind, we identified the OLC of a measure and estimated the differences between maximal reliability and composite reliability and between maximal criterion validity and scale criterion validity. Little empirical research has been conducted on maximal reliability, and none has been conducted on maximal criterion validity. Although the difference between composite reliability and maximal reliability and between scale criterion validity and maximal criterion validity may be rather small numerically (0.010-0.037 in our study), it is worth noting that we used a measure that is established and has been tested repeatedly. The present study provided evidence that the OLC can improve reliability and validity even when composite reliability and scale criterion validity are already sufficiently high. It is not difficult to imagine that the OLC would be more useful when it is necessary to improve the quality of a measure that has unsatisfactory composite reliability, especially when the composite reliability is just below the criterion (e.g., when composite reliability is slightly lower than the acceptable threshold, 0.7).

Moreover, we identified the OLC of CS indicators in order to furnish maximal reliability and maximal criterion validity. We emphasized that the linear combination of the total score obtained from Equation (7) has maximal reliability with regard to any other linear combinations and maximal criterion validity with regard to any external criterion. This feature of the resulting optimal composite score is particularly

important in empirical research, especially when the reliability and criterion-related validity are lower than expected during measure development and structure confirmation. Until proven otherwise, the OLC and the resulting maximal reliability and maximal criterion validity provide an effective way to improve the quality of a measure. This improved estimation with the OLC is due to its feature of accounting for individual variance from item specificity, which is treated as part of the true score variance rather than as part of the error variance, as done in routine factor analysis applications for reliability estimation (Raykov, Marcoulides, & Gabler, 2017).

The use of weighted item scores can minimize the effect of measurement errors on the total score and render total scores more accurate. During the scale construction process, it is common to find that the factor loadings are not the same for each item. Some items may have high loading with small error variance, while other items may have low loading with large error variance. It is reasonable to wonder why we should simply add these item scores even though they contribute differently to the measure under construction. Under these circumstances, weighting item scores make perfect sense, as it fairly reflects the contribution of each item when adding them together.

The accuracy of the weighted total score is reflected in the improvement of the reliability coefficient. The homogeneity reliability and composite reliability are the same for congeneric measures, and item scores can be added together only when the homogeneity reliability, or the composite reliability for congeneric measures, is sufficiently high (Gu, Wen, & Fan, 2017; Rodriguez, Reise, & Haviland, 2015). The fact that the maximal reliability is higher than the composite reliability indicates that the items are more converged, according to the latent construct, by weighting each item. In the meantime, the maximal criterion validity indicates that the OLC exhibits the strongest relationship with an external criterion. If we use the OLC in the context of linear regression, the effect size (R^2) will reach its maximum whether the criterion is considered to be the independent or the dependent variable. Therefore, the weighted total score is more accurate compared to the unweighted total score, if not equal to it.

Although the concepts of the OLC and weighted scores may seem unconventional, the parameters can easily be obtained using the popular software Mplus (Muthén & Muthén, 1998-2015). The Mplus codes employed in the present article for the estimation of the weights in the OLC of the CS indicators, maximal reliability, maximal criterion validity, and their confidence intervals, are provided in Appendixes A and B. These codes can be used in a straightforward manner in empirical research aimed at improving the construct of a measure.

In conclusion, this article provided empirical evidence that maximal reliability and maximal criterion validity are useful for improving measurement quality. The OLC, which results in the maximal reliability and maximal criterion validity, is a readily and widely applicable methodology that can significantly help researchers, especially empirical researchers, to improve the efficacy of assessments based on congeneric measuring instruments.

Appendix A

Mplus Code for the Computation of Weights of the Optimal Linear Combination of Career Satisfaction Indicators, the Maximal Reliability, the Composite Reliability and Their Difference, With 95% Confidence Intervals

TITLE: The optimal linear combination, maximal reliability, composite reliability, and their difference;

DATA:

FILE IS <enter file name > ;

VARIABLE:

NAMES ARE CS1,CS2,CS3,CS4,CS5;

MISSING=ALL(99);

ANALYSIS:

TYPE=GENERAL;

ESTIMATOR=ML;

MODEL:

SAT BY CS1* (L1)

CS2 (L2)

CS3 (L3)

CS4 (L4)

CS5 (L5);

SAT@1;

CS1 (EV1); CS2 (EV2); CS3 (EV3); CS4 (EV4); CS5 (EV5);

MODEL CONSTRAINT:

NEW(W1 W2 W3 W4 W5 max_rel com_rel delta);

W1 = L1/EV1; W2 = L2/EV2; W3 = L3/EV3; W4 = L4/EV4; W5 = L5/EV5;

max_rel = (L1^2/EV1 + L2^2/EV2 + L3^2/EV3 + L4^2/EV4 + L5^2/EV5)

/(1 + L1^2/EV1 + L2^2/EV2 + L3^2/EV3 + L4^2/EV4 + L5^2/EV5);

com_rel = (L1 + L2 + L3 + L4 + L5)^2/

((L1 + L2 + L3 + L4 + L5)^2 + EV1 + EV2 + EV3 + EV4 + EV5);

delta = max_rel - com_rel;

OUTPUT:

CINTERVAL;

Note. This Mplus code is modified from Raykov et al. (2015). Readers may refer to Appendix A in their article for the original version.

Appendix B

Mplus Code for the Computation of Weights of the Optimal Linear Combination of Career Satisfaction Indicators, the Maximal Criterion Validity, the Scale Criterion Validity and Their Difference, With 95% Confidence Intervals, With Internal Marketability and Possibilities for Professional Development as Criteria

TITLE: Assessment of criterion-related model fit, maximal criterion validity for optimal linear combination (OLC) of career satisfaction indicators, the scale criterion validity of the sum of these indicators, and the difference (delta) between the two validity coefficients when using internal marketability as the criterion;

DATA:

FILE IS <enter file name > ;

VARIABLE:

NAMES ARE CS1 CS2 CS3 CS4 CS5 INS;

MISSING=ALL(99);

MODEL:

SAT BY CS1* (L1)

CS2 (L2)

CS3 (L3)

CS4 (L4)

CS5 (L5);

SAT@1;

IN BY INS*;

IN@1;

INS@0;

SAT WITH IN* (corr);

CS1 (EV1); CS2 (EV2); CS3 (EV3); CS4 (EV4); CS5 (EV5);

MODEL CONSTRAINT:

NEW(max_rel com_rel max_cv sca_cv delta_rel delta_cv);

max_rel = (L1^2/EV1 + L2^2/EV2 + L3^2/EV3 + L4^2/EV4 + L5^2/EV5)
/(1 + L1^2/EV1 + L2^2/EV2 + L3^2/EV3 + L4^2/EV4 + L5^2/EV5);

com_rel = (L1 + L2 + L3 + L4 + L5)^2/
((L1 + L2 + L3 + L4 + L5)^2 + EV1 + EV2 + EV3 + EV4 + EV5);

max_cv= SQRT(max_rel)*corr;

sca_cv= SQRT(com_rel)*corr;

delta_rel = max_rel-com_rel;

delta_cv = max_cv-sca_cv;

OUTPUT:

CINTERVAL;

Note. This Mplus code is modified from Raykov et al. (2016). Readers may refer to the appendix in their article for the original version.

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) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by grants from the National Natural Science Foundation of China (31771245, 31400909).

Note

1. Readers are advised to pay special attention to this reference because it discussed maximal reliability with correlated measurement errors, which is a condition beyond the concern of the present study.

References

- Anastasi, A. (1988). *Psychological testing* (6th ed.). New York, NY: Macmillan.
- Armstrong-Stassen, D. M., & Ursel, N. D. (2009). Perceived organizational support, career satisfaction, and the retention of older workers. *Journal of Occupational and Organizational Psychology*, 82, 201-220.
- Bakker, A., Demerouti, E., Taris, T. W., Schaufeli, W. B., & Schreurs, P. J. (2003). A multigroup analysis of the job demands-resources model in four home care organizations. *International Journal of Stress Management*, 10, 16-38.
- Bartholomew, D. J. (1996). *The statistical approach to social measurement*. San Diego, CA: Academic Press.
- Brunner, M., & Süb, H.-M. (2005). Analyzing the reliability of multidimensional measures: An example from intelligence research. *Educational and Psychological Measurement*, 65, 227-240.
- Cohen, R. J., & Swerdlik, M. E. (2009). *Psychological testing and assessment: An introduction to tests and measurement* (7th ed.). New York, NY: McGraw-Hill Primis.
- Dimitrov, D. M., Raykov, T., & Alqataee, A. A. (2015). 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*, 75, 475-490.
- Eby, L. T., Butts, M., & Lockwood, A. (2003). Predictors of success in the era of the boundaryless career. *Journal of Organizational Behavior*, 24, 689-708.
- Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269-314). Charlotte, NC: Information Age.
- Gabler, S., & Raykov, T. (2016). Evaluation of maximal reliability for unidimensional measuring instruments with correlated errors. *Structural Equation Modeling*, 24, 104-111. doi:10.1080/10705511.2016.1159916
- Greenhaus, J. H., Parasuraman, S., & Wormley, W. M. (1990). Effects of race on organizational experiences, job performance evaluations, and career outcomes. *Academy of Management Journal*, 33, 64-86.

- Gu, H., Wen, Z., & Fan, X. (2017). Examining and controlling for wording effect in a self-report measure: A Monte Carlo simulation study. *Structural Equation Modeling*, 24, 545-555. doi:10.1080/10705511.2017.1286228
- Guilford, J. P. (1954). *Psychometric methods*. New York, NY: McGraw-Hill.
- 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. doi:10.1080/10705519909540118
- Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109-133. doi:10.1007/BF02291393
- Karatepe, O. M. (2012). Perceived organizational support, career satisfaction, and performance outcomes. *International Journal of Contemporary Hospitality Management*, 24, 735-752.
- Karatepe, O. M., & Olugbade, O. A. (2016). The effects of work social support and career adaptability on career satisfaction and turnover intentions. *Journal of Management & Organization*, 23, 337-355.
- Kong, H. Y., Cheung, C., & Song, H. Y. (2012). From hotel career management to employees' career satisfaction: The mediating effect of career competency. *International Journal of Hospitality Management*, 31(1), 76-85.
- Li, H. (1997). A unifying expression for the maximal reliability of a linear composite. *Psychometrika*, 62, 245-249.
- Li, H. (2003). The resolution of some paradoxes related to reliability and validity. *Journal of Educational and Behavioral Statistics*, 28, 89-95.
- Li, H., Rosenthal, R., & Rubin, D. B. (1996). Reliability of measurement in psychology: From Spearman-Brown to maximal reliability. *Psychological Methods*, 1, 98-107.
- Marcoulides, K. M., & Yuan, K. H. (2016). New ways to evaluate goodness of fit: A note on using equivalence testing to assess structural equation models. *Structural Equation Modeling*, 24, 148-153.
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, 11, 320-341.
- McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, 23(1), 1-21.
- McDonald, R. P. (1999). *Test theory. A unified treatment*. Mahwah, NJ: Erlbaum.
- Mosier, C. I. (1943). On the reliability of a weighted composite. *Psychometrika*, 8, 161-168.
- Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Penev, S., & Raykov, T. (2006). On the relationship between maximal reliability and maximal validity of linear composites. *Multivariate Behavioral Research*, 41, 105-126.
- Raykov, T. (2004). Estimation of maximal reliability: a note on a covariance structure modelling approach. *British Journal of Mathematical and Statistical Psychology*, 57, 21-27.
- Raykov, T., Gabler, S., & Dimitrov, D. M. (2015). Maximal reliability and composite reliability: Examining their difference for multicomponent measuring instruments using latent variable modeling. *Structural Equation Modeling*, 23, 384-391. doi:10.1080/10705511.2014.966369
- Raykov, T., Gabler, S., & Dimitrov, D. M. (2016). Maximal criterion validity and scale criterion validity: A latent variable modeling approach for examining their difference. *Structural Equation Modeling*, 23, 544-554. doi:10.1080/10705511.2016.1155414

- Raykov, T., & Grayson, D. (2003). A test for change of composite reliability in scale development. *Multivariate Behavioral Research*, 38, 143-159. doi:10.1207/s15327906mbr3802_1
- Raykov, T., Marcoulides, G. A., & Gabler, S. (2017). Improved estimation of maximal reliability for unidimensional multicomponent measuring instruments in repeated measure studies. *Structural Equation Modeling*, 24, 755-767. doi:10.1080/10705511.2016.1183491
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2015). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21, 137-150.
- West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA: Sage.
- Willoughby, M. T., Pek, J., & Blair, C. B. (2013). Measuring executive function in early childhood: A focus on maximal reliability and the derivation of short forms. *Psychological Assessment*, 25, 664-670.
- Yuan, K. H., Chan, W., Marcoulides, G. A., & Bentler, P. M. (2015). Assessing structural equation models by equivalence testing with adjusted fit indexes. *Structural Equation Modeling*, 23, 319-330.