Artur Pokropek, Educational Research Institute (IBE)
Gary N. Marks, University of Melbourne
Francesca Borgonovi, University College London

Online Appendix. How much do students' scores in PISA reflect general intelligence and how much do they reflect specific abilities? *Journal of Educational Psychology*

https://doi.org/10.1037/edu0000687.

Indices Used in the Paper

This paper refers to 3 types of indices:

- Explained Common Variance (*ECV*)
- Omega (McDonald 1999) and its extensions OmegaH/HS
- Proportional Reduction in Mean Squared Error (*PRMSE*)

Standardized factor loadings of bi-factor model are used to compute ECV and Omega indices. In models factor loadings are denoted as:

λ_g – standardized factor loading on the general factor for *g=1,2, …, G* items,
λ_r – standardized factor loading on the specific reading factor for *r=1,2, …, R* items,
λ_m – standardized factor loading on the specific math literacy factor for *m=1,2, …, M* items,
λ_s – standardized factor loading on the specific science literacy factor for *s=1,2, …, S* items,
λ_v – standardized factor loading on the specific Raven factor for *v=1,2, …, V* items,

and:

h - denotes unique variances.

ECV_{Gen} is defined as the common variance explained by the general factor divided by the total common variance (Reise, Moore and Haviland 2010):

$$ECV_{Gen} = \frac{\sum \lambda_g^2}{\sum \lambda_g^2 + \sum \lambda_r^2 + \sum \lambda_m^2 + \sum \lambda_s^2 + \sum \lambda_v^2} \tag{1a}$$

ECV for specific factors is the proportion of the common variance each specific factor accounts for. For instance, ECV for the math factor can be expressed as:

$$ECV_{Math} = \frac{\sum \lambda_m^2}{\sum \lambda_g^2 + \sum \lambda_r^2 + \sum \lambda_m^2 + \sum \lambda_s^2 + \sum \lambda_v^2} \quad (1b)$$

Coefficient *Omega* (McDonald 1999) is a factor analytic model-based reliability estimate and is computed as the ratio of the reliable variance (measured by all factor loadings) to the total variance. *Omega* indicates of how much of the variance in observed total score can be attributed to all modeled common factors that is all factors related with set of items:

$$Omega = \frac{(\sum \lambda_g)^2 + (\sum \lambda_r)^2 + (\sum \lambda_m)^2 + (\sum \lambda_s)^2 + (\sum \lambda_v)^2}{(\sum \lambda_g)^2 + (\sum \lambda_r)^2 + (\sum \lambda_m)^2 + (\sum \lambda_s)^2 + (\sum \lambda_v)^2 + (1-h)^2} \quad (2a)$$

Coefficient *OmegaH* indices how much reliable variance of the total scores can be attributed to the general factor (Raykov 1997; Reise, Bonifay, & Haviland, 2013; Reise at al. 2013). Its calculation is similar to coefficient *Omega*, but the numerator comprises only one source of reliable variance in unit-weighted total scores, that is due to the general factor:

$$OmegaH = \frac{(\sum \lambda_g)^2}{(\sum \lambda_g)^2 + (\sum \lambda_r)^2 + (\sum \lambda_m)^2 + (\sum \lambda_s)^2 + (\sum \lambda_v)^2 + (1-h)^2} \quad (2b)$$

OmegaH can also be computed also for subscales (debnoted as *OmegaHS*). *OmegaHS* is the proportion of subscale score variance attributable to a group factor, after removing the reliable variance due to the general factor (Rodriguez, Reise, and Haviland, 2016). The estimate of OmegaHS for the math factor is:

$$OmegaHS_m = \frac{(\sum \lambda_m)^2}{(\sum \lambda_g)^2 + (\sum \lambda_r)^2 + (\sum \lambda_m)^2 + (\sum \lambda_s)^2 + (\sum \lambda_v)^2 + (1-h)^2} \quad (2c)$$

PRMSE indicates the relative importance of specific factors over the general factor in explaining variability in response patters to test items (Haberman 2008). PRMSE is computed on observed scores and does not assume that the factors are orthogonal or an underlying bifactor model (Haberman, Sinharay & Puhan, 2008).

The PRMSE can be computed for subscales ($PRMSE_S$) and for a total scores ($PRMSE_T$) and is based on radio of mean square errors (MSE) of subscale (MSE_{sub}) or total score (MSE_{total}) to the mean square error of the true construct score (MSE_{sub}):

$$PRMSE_S = 1 - \frac{MSE_{sub}}{MSE_{const}} \quad (3a)$$

$$PRMSE_T = 1 - \frac{MSE_{total}}{MSE_{const}} \quad (3b)$$

Where:

MSE_{const} - is a true variance of true score (construct)
MSE_{sub} - is a true variance of true score (construct) weighted by (*1-rliability of subscore*)
MSE_{total} - is a true variance of true score (construct) weighted by (*1-rliability of total score*)

According to Haberman (2008), a subscore has added value if:

$$PRMSE_s > PRMSE_T$$

All PRMSE statistics are computed within the regression framework. Detailed information on its estimation can be found in Haberman (2008), Haberman, Sinharay & Puhan, (2008) and Sinharay (2018).

References

Haberman, S. J. (2008). When can subscores have value? *Journal of Educational and Behavioral Statistics*, *33*, 204-229.

Haberman, S., Sinharay, S., & Puhan, G. (2008). Reporting subscores for institutions. *British Journal of Mathematical and Statistical Psychology, 62*, 79-95.

McDonald, R. P. (1999). Test theory: A unified approach. Mahwah, NJ: Erlbaum.

Raykov, T. (1997). Estimation of Composite Reliability for Congeneric Measures. *Applied Psychological Measurement, 21(2),* 173–184.

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment,* 95, 129–140.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of personality assessment*, *92*(6), 544-559.

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, *73*(1), 5-26.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures. *Journal of Personality Assessment, 98*(3), 223-237. doi:10.1080/00223891.2015.1089249

Sinharay, S. (2018). *Added Value of Subscores and Hypothesis Testing. Journal of Educational and Behavioral Statistics, 107699861878886.* doi:10.3102/1076998618788862