University of British Columbia, Department of Psychology

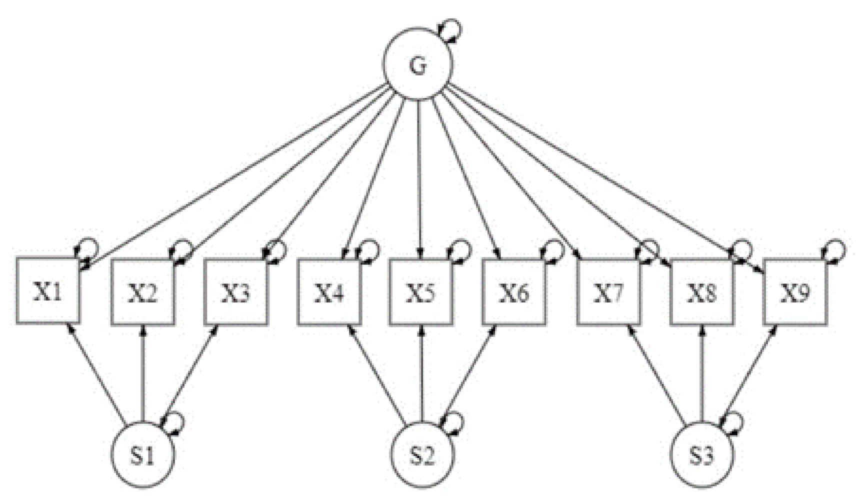
Looking Beyond Fit Indices: A Bifactor Model Best Practice Study

Sijia (Carol) Li sijia.li@ubc.ca , Victoria Savalei v.salavei@ubc.ca

Introduction

The bifactor model (see Figure 1) was developed 80 years ago[1], and received huge popularity recently. Popular applications in psychology include: general psychopathology model, general personality factor model, general mental ability model, etc.

***Figure 1*. Standard Bifactor Model**



Statistical Pitfalls of Bifactor Model

- Tendency to overfit[2]
- Propensity to produce anomalous results including small loadings, unexpected negative loadings, and negative factor variances[3]
- Instability across samples[4]
- Under-identification problems[5]

Best Practice Recommendations

- Ensuring solution interpretability; considering alternative models (i.e., hierarchical model, bifactor S-1 model)[6]
- Reliability and dimensionality indices, including omegaH, omegaHS, ECV, PUC, H, and FD[7]

Method

With keyword "bifactor", from *PsycInfo,* we found 109 articles published in 2020, in which N = 97 fit one or more confirmatory bifactor model(s). From 81 articles with model structural details reported, we coded $N_m = 107$ bifactor models:

- Model statistics: fit indices, omega reliabilities, dimensionality measures.
- Model structure: factor loadings, scale length, # of group factors, model type (i.e., S-1 model).
- Data disclosure: correlation matrix, raw dataset, factor loadings.
- Content-related information: area of application, psychological construct, study conclusion.

Results

Overview (out of 97 articles)

- Psychological constructs: health/clinical assessment (53%), social/personality assessment (25%), cognitive (9%), organizational (7%), and educational (6%).
- Model comparison: any alternative model (81%), unidimensional model (46%), multiple factor model (70%), higher-order model (37%), 20% (other bifactor model).
- At least one model fit indices (100%); no reliability or dimensionality measures (53%).
- Factor loading (66%), correlation matrix (19%), dataset (8%); retained (85%), rejected (15%).

Model Structure (out of 89 retained models)

- Total scale (M = 21.67, SD = 17.03): 10-20 items (50%), <30 items (87%). Subscale (M = 5.94, SD = 4.04): <5 items (48%), <10 items (84%).
- Group factors (M = 3.27, SD = 1.57): 2 group factors (34%), 1-4 group factors (84%).
- Standard bifactor model (78%), bifactor S-1 models (22%), post-hoc bifactor S-1 model (73%)

Factor Loading (out of 57 retained models)

- Standardized factor loading (see Figure 2): general factor (M = 0.547, SD = 0.200), group factor (M = 0.386, SD = 0.250).
- Negative loading: general factor (M = -0.088), group factor (M = -0.103); at least one negative loading (anywhere: 46%, general factor: 9%, group factor: 42%).
- Small loading: at least one $\lambda_{gen} \leq |0.5|$ (88%), at least one $\lambda_{gen} \leq |0.3|$ (51%), at least one $\lambda_{grp} \leq |0.3|$ (81%).

Omega Reliabilities (out of 49 retained models)

- Total scale: ω_H (M = 0.768, SD = 0.142), ω_{HS} (M = 0.148, SD = 0.116); $\omega_H > 0.7$ (71%), $\omega_H < 0.5$ (6%). Negative Residual Variances (13%).
- Subscale: $\omega_{H,j}$ (M = 0.507, SD = 0.239), $\omega_{HS,j}$ (M = 0.320, SD = 0.220), at least one $\omega_{HS,j} > \omega_{H,j}$ (69%), at least one $\omega_{HS,j} > 0.5$ (45%).
- Replication: omega reliabilities replicated in 13 articles (out of 22); average absolute deviation was 0.113 (SD = 0.099).

***Figure 2*. Standardized Factor Loadings: Max, Min, Mean, and Range**

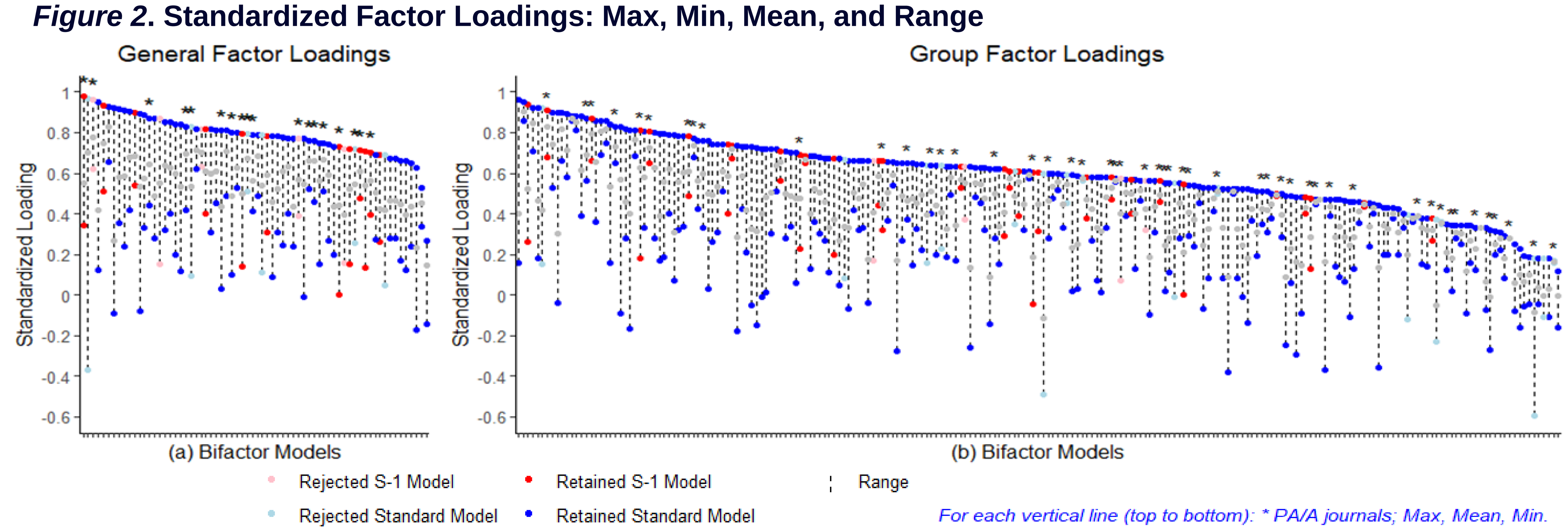
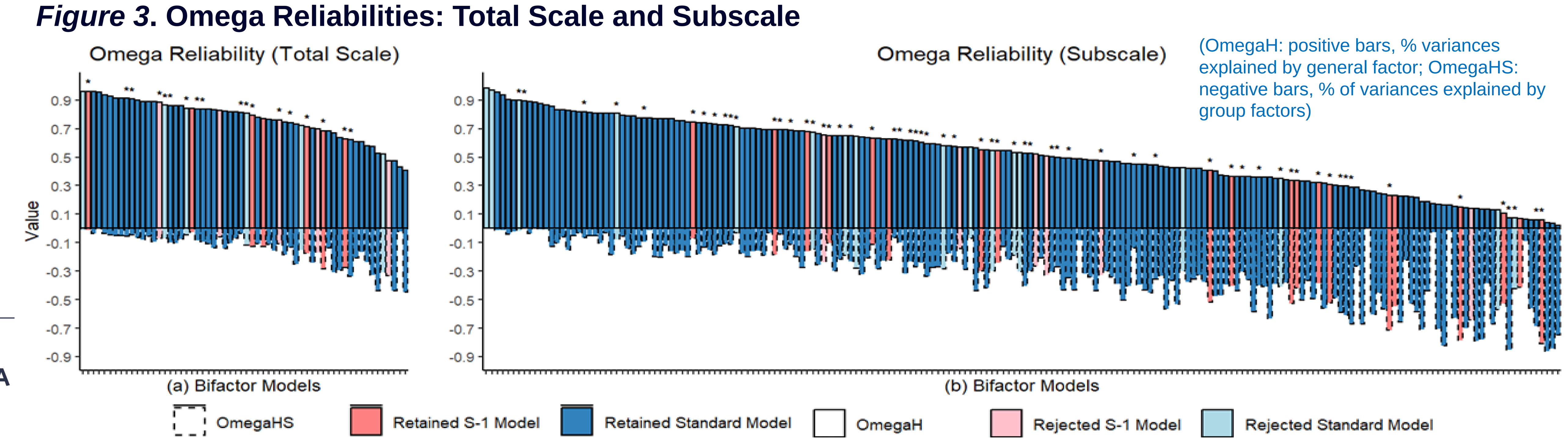
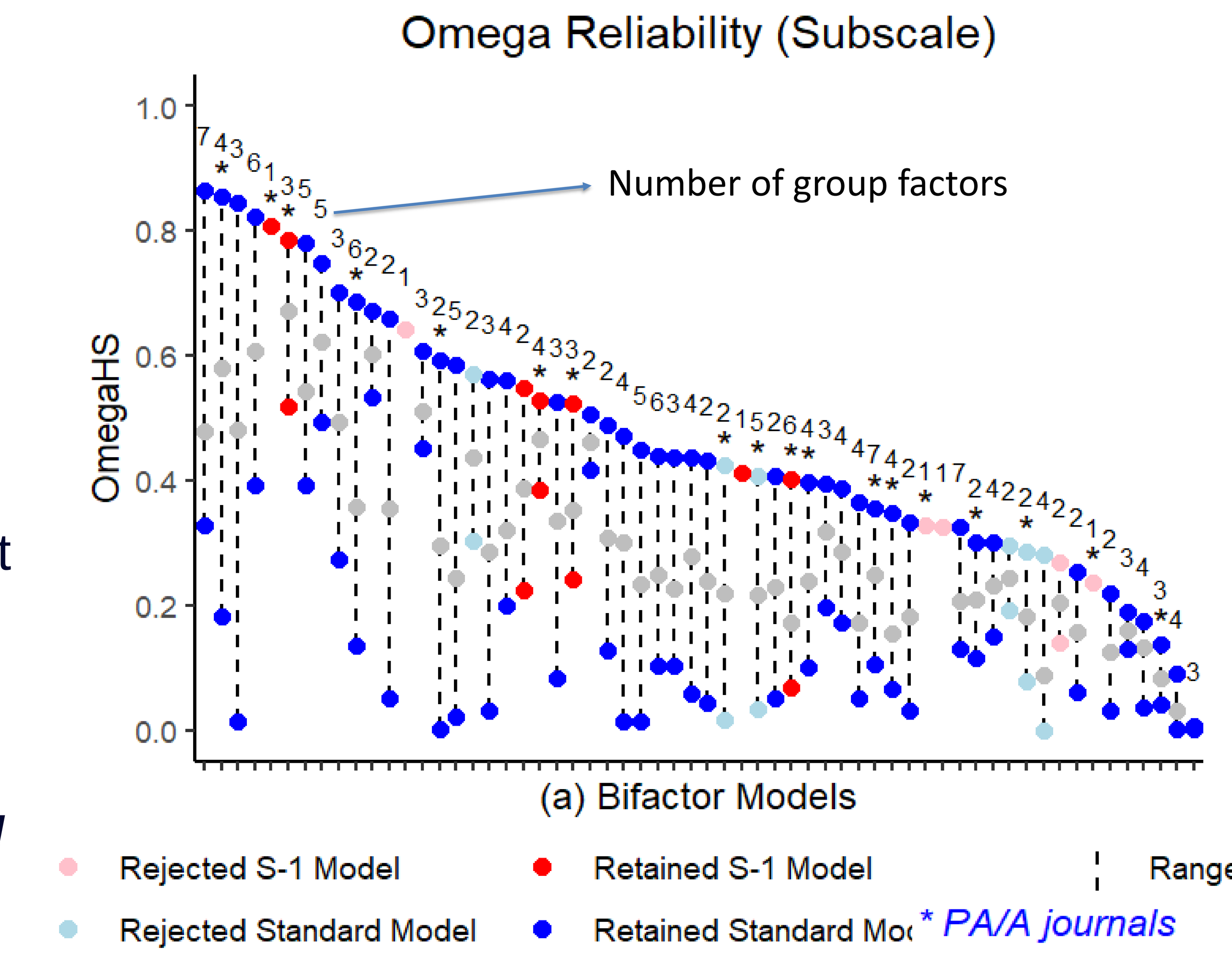


***Figure 3*. Omega Reliabilities: Total Scale and Subscale**



Discussion

***Figure 4*. OmegaHS (% of variances explained by group factors): Min, Max, Mean, Range**



- Reliability/Dimensionality measures: not considered by 53% of the articles; reported values cannot be replicated in 41% of studies.
- At least one anomalous result: retained bifactor models (100%), rejected models (100%).
- Rejected models: weaker subscales ($\omega_{HS,j}$: M = 0.241), smaller negative loadings (general factor: M = -0.345, group factor: M = -0.201). Limitation: small sample size.
- Group factors are weaker and more problematic: less reliable, smaller factor loadings, more likely to have unexpected negative loadings and small loadings.
- Group factors are expected to be less reliable than general factor. We suggest an alternative statistics "partial or conditional reliability":

$$c_{HS} = \frac{\sum_{j=1}^{k}(\sum_{t=1}^{m_j}\lambda_{g,t})^2}{\sum_{j=1}^{k}(\sum_{t=1}^{m_j}\lambda_{g,t})^2 + \sum_{i=1}^{p}U_i}$$

- (k: # of group factors, m_j: # of items in subscale j, $\lambda_{g,t}$: group factor loading, U_i: i[th] residual variance)

References

1. Holzinger, K.J., Swineford, F. (1937). The Bi-factor method. *Psychometrika*, 2, 41–54.
2. Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence, 41*(5), 407–422.
3. Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods, 22*(3), 541–562.
4. Bornovalova, M. A., Choate, A. M., Fatimah, H., Petersen, K. J., & Wiernik, B. M. (2020). Appropriate use of bifactor analysis in psychopathology research: Appreciating benefits and limitations. *Biological Psychiatry, 88*(1), 18–27.
5. Green, S., & Yang, Y. (2017). Empirical underidentification with the BIFACTOR model: A case study. Educational and *Psychological Measurement, 78*(5), 717–736.
6. Watts, A. L., Poore, H., & Waldman, I. (2019). Riskier tests of the validity of Bifactor models of Psychopathology. Clinical *Psychological Science, 7*(6), 1285–1303.
7. Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137–150.

THE UNIVERSITY OF BRITISH COLUMBIA