Added value of subscores: Can we accurately evaluate it?

Friday 25 July 2025 09:15 (15 minutes)

Oral presentation

Added value of subscores: Can we accurately evaluate it?

Author

Angelina Kuchina

Affiliation

Tilburg University

Keywords

subscore, total score, added value

Abstract

In this presentation, we will examine the concept of subscore added value, a critical topic in educational and psychological testing, focusing on its evaluation. Subscores are often reported to provide more detailed feedback to test-takers and educators, but their usefulness depends on whether they add value beyond the total score.

Haberman (2008) introduced a criterion, stating that a subscore has added value if the squared correlation between the subscore and the true subscore exceeds that of the total score and the true subscore. Since this method requires parameter estimates from a sample, one crucial issue that needs to be considered is sampling variability. Even if the subscores have added value, the sample estimates of the squared correlations with the true subscore may suggest they do not, and vice versa.

Sinharay (2019) proposed using hypothesis testing to address this issue. However, he restated the hypotheses in terms of correlations between the observed subscore or total score, and a parallel-form subscore. Sinharay suggested using established statistical methods for testing dependent correlations, such as William's (1959) t and Olkin's (1967) Z statistics, to determine the significance of the difference between these correlations.

Nevertheless, the properties of the traditional statistics may not fully apply to the context of the added value of subscores. Both tests assume a trivariate normal distribution, but this assumption may not hold for discrete test scores. Moreover, these tests assume all variables are observed, while only correlations between observed (subscore or total score) and unobserved (parallel-form subscore) variables are available in this context. Finally, these correlations are derived from the assumption that the correlation between the subscores on parallel test forms equals the squared correlation between an observed and a true subscore. However, their sampling distributions differ, which may bias statistical conclusions.

Sinharay's (2019) results obtained from resampling an existing empirical dataset have some limitations. To accurately evaluate the performance of the proposed statistics, it is crucial to control for the true population parameters and the sampling mechanism, which is impossible in real-data simulations.

To address these gaps, we present findings from a comprehensive simulation study evaluating the accuracy of Olkin's Z and William's t statistics within Sinharay's (2019) parallel-form approach and original Haberman's (2008) method. Furthermore, a non-parametric bootstrap procedure is employed as an alternative for testing the significance.

The results reveal that the performance of Olkin's Z and William's t statistics within Sinharay's parallel-form

approach is overly conservative, with low statistical power across all conditions. In contrast, applying these tests to Haberman's framework shows varied performance: inflation occurs at low subscore reliability, while high reliability results in conservative performance. These statistics, however, are the most powerful under all conditions. The non-parametric bootstrap procedure appears slightly conservative but shows promise for determining subscore added value. Nevertheless, it may face challenges in detecting small effects, particularly at low reliability levels.

These findings provide valuable insights into the practical application of subscore evaluation methods. Future work will investigate the underlying factors influencing the performance of these statistical methods.

Primary author: KUCHINA, Angelina (Tilburg University)

Co-authors: Dr BOLSINOVA, Maria (Tilburg University); Dr EMONS, Wilco (Tilburg University)

Session Classification: Session 18 : "Latent factors and errors in psychometric measurement"

Track Classification: Measurement: Measurement