INVESTIGATING MULTIDIMENSIONALITY IN THE COMPREHENSIVE ASSESSMENT OF OUTCOMES IN STATISTICS (CAOS)

V.N. Vimal Rao

COLLEGE OF EDUCATION + HUMAN DEVELOPMENT

Department of Educational Psychology

UNIVERSITY OF MINNESOTA
Driven to Discover®

"The job of validation is not to support an interpretation, but to find out what might be wrong with it."
(Cronbach, 1980, p. 103)

CAOS sub-scores by topic do not meet standards for distinctness and reliability.

Their use is not recommended for statistics education research.

Only CAOS total scores should be calculated and interpreted.

BACKGROUND

CAOS developmental goals

An assessment with:

(1) questions understandable by all students finishing any introductory statistics course

(2) reliable scores

(3) the ability to identify areas where students do or do not improve in their understanding

Initial validity evidence

- Experts agree that CAOS is well suited for students completing the *consensus curriculum* (Cobb, 2007)
- Internal consistency = 0.82 (delMas et al., 2007)
- CFA confirms unidimensionality (delMas, 2014)

Unforeseen score interpretations

- Some studies calculated and analyzed CAOS subscores defined by topics (e.g., sampling variability, confidence intervals).
- Standard 1.14 of the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014) states that subscores should be distinct and reliable.

Research questions

(1) Are CAOS subscores of items grouped by topic area distinct and reliable?

(2) Is there any partition of CAOS items that creates distinct and reliable subscores?

Methods

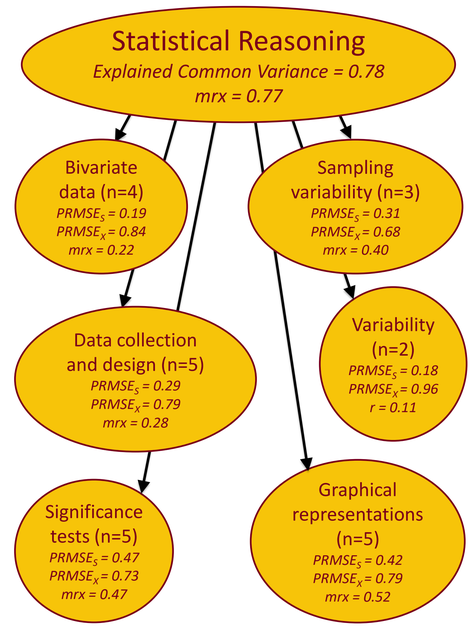
- Proportional reduction in mean squared error (PRMSE) and marginal reliability estimates (mrx) to assess reliability
- 6-topic bifactor testlet model and exploratory bifactor modelling to assess distinctiveness
- Exploratory bifactor modelling to identify any potentially distinct and reliable subscores

EXAMPLE INTERPRETATIONS

"The consensus cohort showed a substantial loss in knowledge about data collection and design (12.5 percentage point decline), whereas the randomization cohort exhibited a minor loss (1.97 percentage point decline)."
(Tintle et al., 2012, p. 27)

"The data suggested improvement in students' overall understanding of the interpretation of confidence intervals."
(Wang et al., 2019, p. 11)

SIX-FACTOR MODEL



EXPLORATORY MODEL

Model Identification

- The Schmid-Leiman orthogonalization (Schmid & Leiman, 1957; Waller, 2018)
- Between two and six factors
- Two factor model had best fit statistics

Model Characteristics

- Factors roughly aligned with descriptive and inferential topics
- MRX of 0.67 and 0.60
- PRMSE$_S$ of 0.62 and 0.53
- PRMSE$_X$ of 0.73 and 0.73
- ECV of 0.71

Model Interpretation

With PRMSE$_S$ < PRMSE$_X$, low MRX, and high ECV, it appears that even this best fitting model does not meet the criteria for distinctness and reliability. Thus, only CAOS total scores should be calculated.

FOR MORE INFORMATION ABOUT CAOS