Exploring Estimates of Multilevel Reliability for School Based Behavioral Measures
Katie Scarlett Lane Pelton, M.A. Betsy McCoach, Ph.D.
University of Connecticut

Behavioral Assessment in Schools

- The National Survey of Children’s Health 2020 found increases in point prevalence of anxiety, depression, and conduct problems up to 9.2%, 4.0%, and 8.1%, respectively (Lebrun-Harris et al., 2022).
- In a 2017, approximately 20% of school psychologists reported being involved in universal behavior screening process at their schools – one method for detecting students’ behavioral needs (Benson et al., 2019).
  - A teacher typically rates their entire homeroom class or a given class period (Oakes et al., 2017).
  - 4-point Likert-type scales are most common (Pelton et al., 2024).
  - Data is typically heavily skewed.
  - Schools often want to aggregate scores to examine program effectiveness at class and/or school level.

Simulation Approach

- Simulated multilevel, ordinal, skewed data modeled after a free-access behavior screening instrument with the simstudy package.
- Instrument Parameters
  - 7 items with a 4-point scale
  - Inter-item correlation = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6
  - Item base probabilities from previous study (Schatschneider et al., 2014).
- Level-two variables
  - Class size: non-zero Poisson distribution centered at 20 with 100
  - Class behavior score: mean = 0, variance = 1, normally distributed
- Level-one variables
  - Student behavior variance (adjustment variable): mean = 0, variance = 0, 2, 4, 6
  - Calculated single- and multilevel-alpha and omega for each of the 500 iterations per condition

Multilevel Reliability

- Estimating score reliability at the student level is common practice, though it should also be estimated at the school or classroom level if aggregating scores (Jak & Jorgensen, 2017).
- Lai (2021) extends estimation of alpha into a multilevel framework.
- Composite two-level estimate for individual and configural constructs incorporates within- and between-cluster loadings and residual variance from multilevel CFA.
- Within-cluster and between-cluster reliability estimates include only the relevant variance component and may vary substantially in both magnitude and meaning depending on the construct being measured.

Preliminary Findings

Internal Consistency Estimates as a Function of Inter-Item Rho

- Single-Level Composite
- Multilevel Composite
- Multilevel Within
- Multilevel Between

Relative Bias in Single Level Alpha as a Function of Inter-Item Rho

- Multilevel Composite
- Multilevel Within
- Multilevel Between

Simulation Approach

- Simulated multilevel, ordinal, skewed data modeled after a free-access behavior screening instrument with the simstudy package.
- Instrument Parameters
  - 7 items with a 4-point scale
  - Inter-item correlation = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6
  - Item base probabilities from previous study (Schatschneider et al., 2014)
- Level-two variables
  - Class size: non-zero Poisson distribution centered at 20 with 100
  - Class behavior score: mean = 0, variance = 1, normally distributed
- Level-one variables
  - Student behavior variance (adjustment variable): mean = 0, variance = 0, 2, 4, 6
  - Calculated single- and multilevel-alpha and omega for each of the 500 iterations per condition

Simulation Approach

- Simulated multilevel, ordinal, skewed data modeled after a free-access behavior screening instrument with the simstudy package.
- Instrument Parameters
  - 7 items with a 4-point scale
  - Inter-item correlation = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6
  - Item base probabilities from previous study (Schatschneider et al., 2014)
- Level-two variables
  - Class size: non-zero Poisson distribution centered at 20 with 100
  - Class behavior score: mean = 0, variance = 1, normally distributed
- Level-one variables
  - Student behavior variance (adjustment variable): mean = 0, variance = 0, 2, 4, 6
  - Calculated single- and multilevel-alpha and omega for each of the 500 iterations per condition

References


