



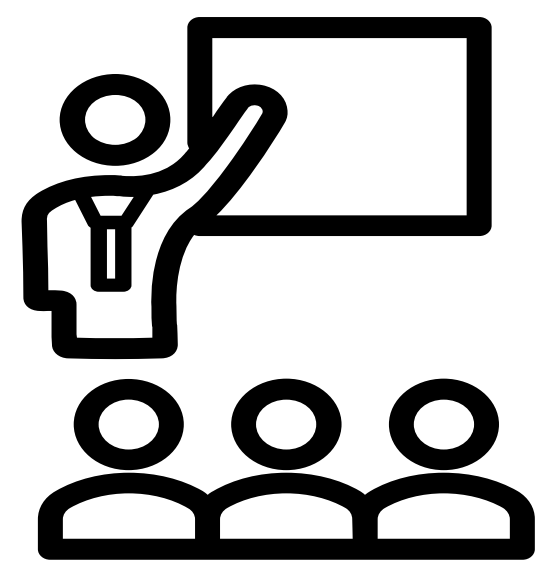
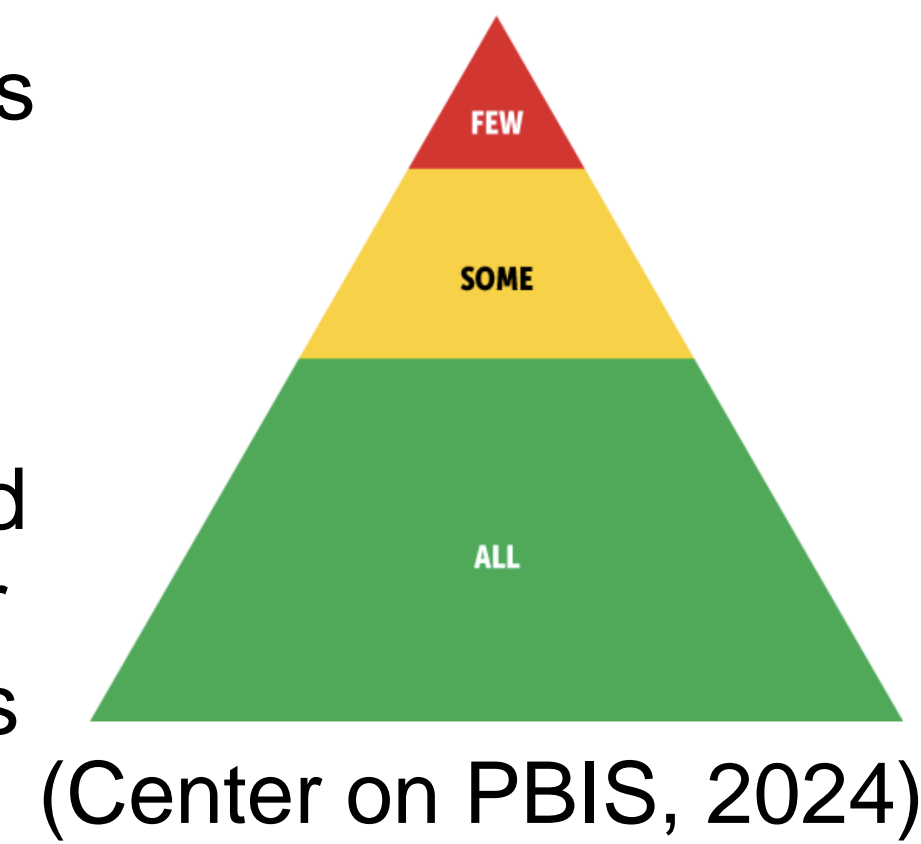
Exploring Estimates of Multilevel Reliability for School Based Behavioral Measures



Katie Scarlett Lane Pelton, M.A. Betsy McCoach, Ph.D.
University of Connecticut

Behavioral Assessment in Schools

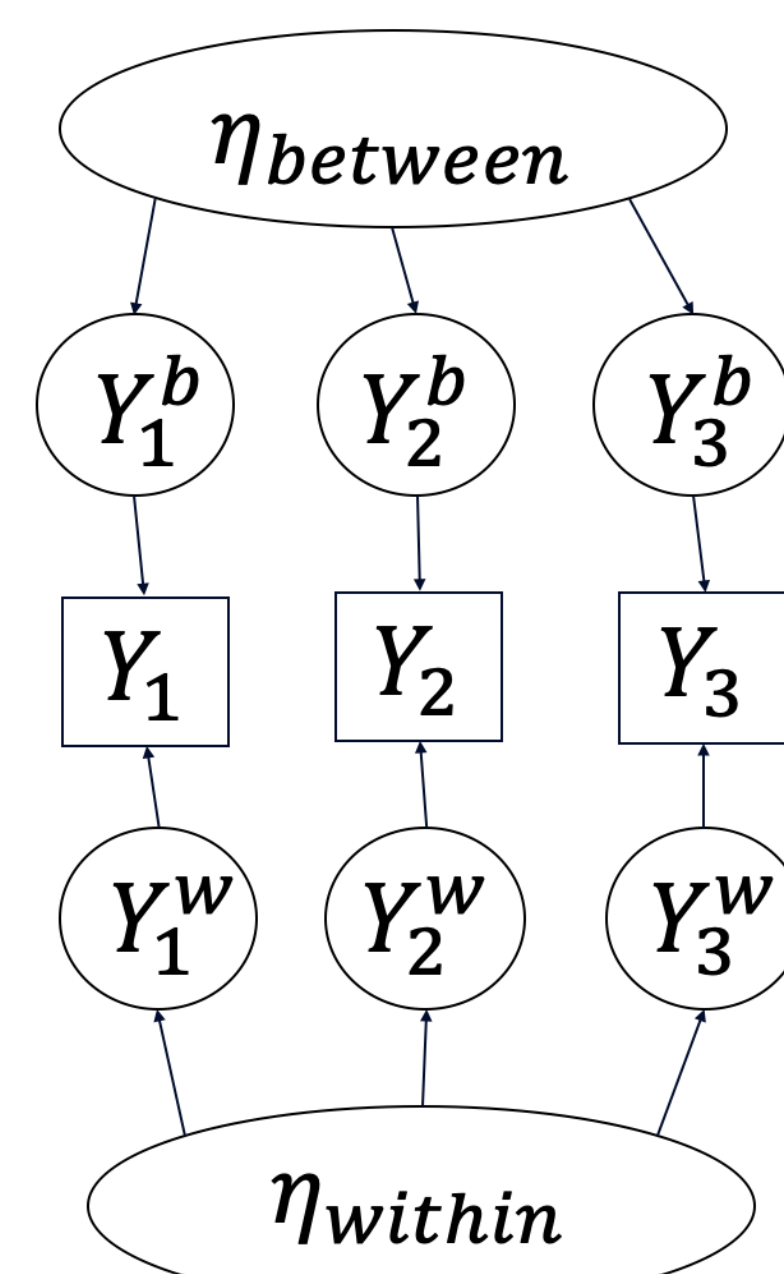
- The National Survey of Children's Health 2020 found increases in point prevalence of anxiety, depression, and conduct problems up to 9.2%, 4.0%, and 8.1%, respectively (Lebrun-Harris et al., 2022)
- In a 2017, approximately 20% of school psychologists reported being involved in universal behavior screening 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

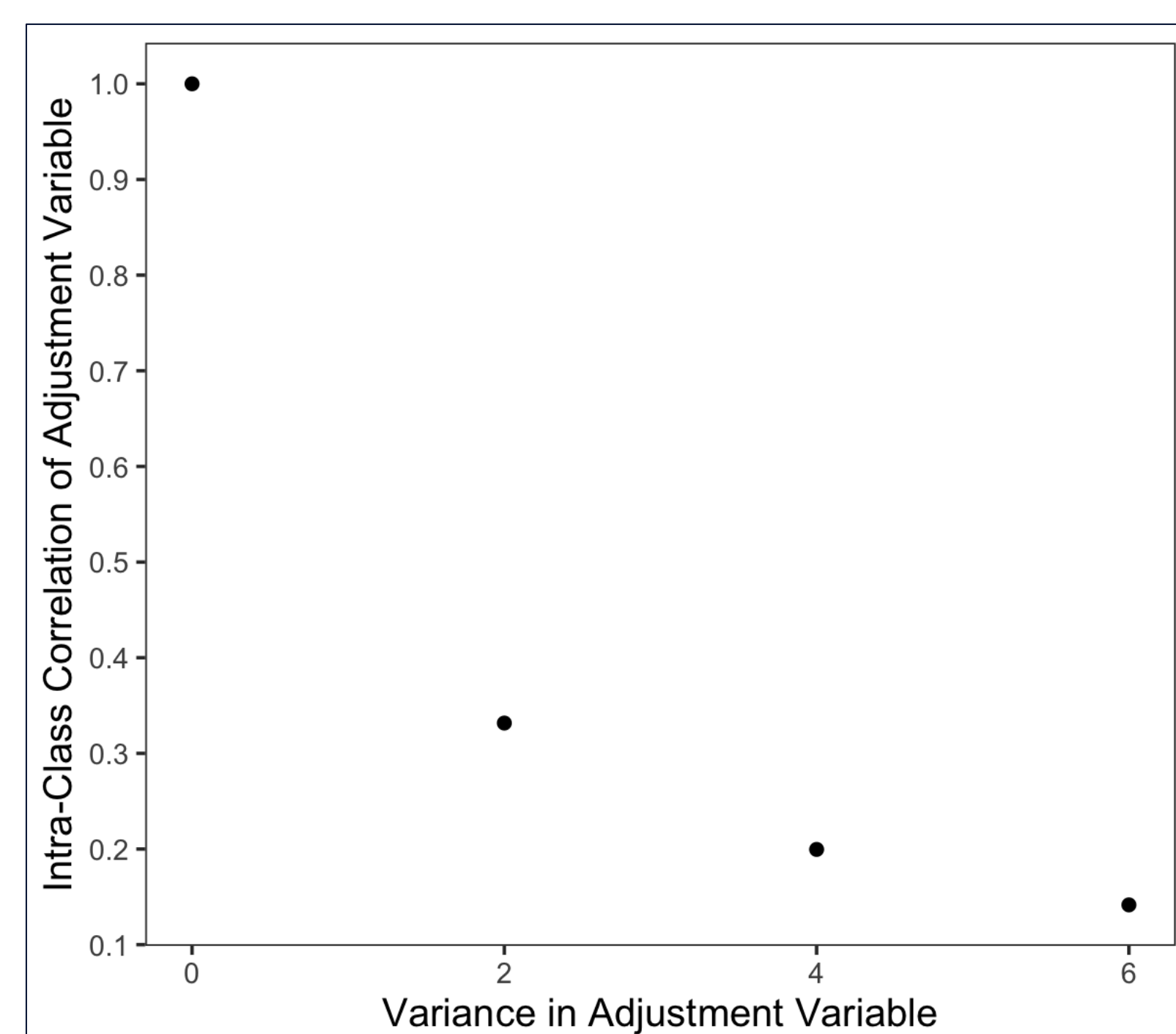
Multilevel Reliability

- 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



Simulation Approach

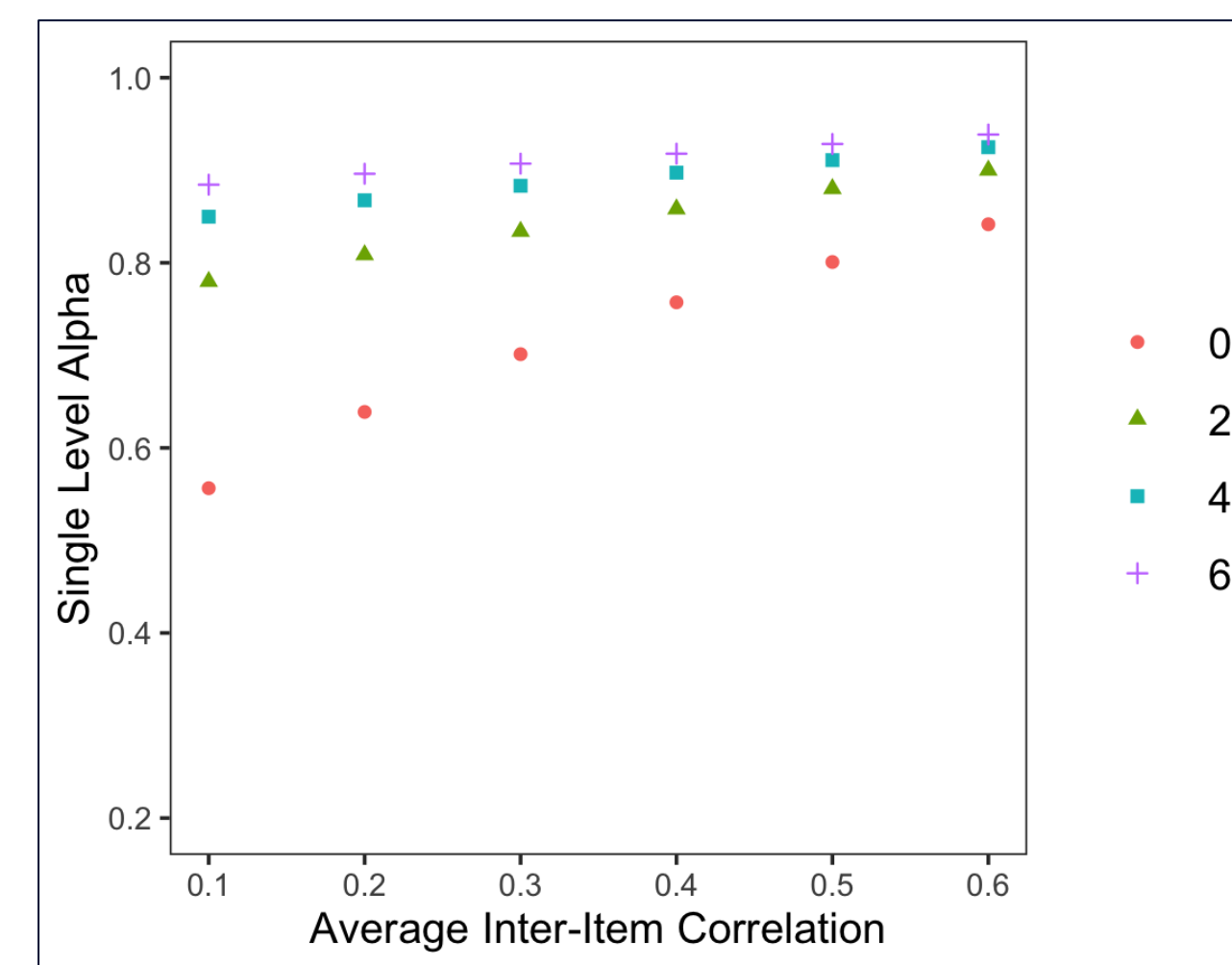
- Simulated multilevel, ordinal, skewed data modeled after a free-access behavior screening instrument with the *simstudy* package
- Instrument Parameters
 - 7 items with a 4-point scale
 - Inter-item correlation = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6**
 - Item base probabilities from previous study (Schatschneider 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



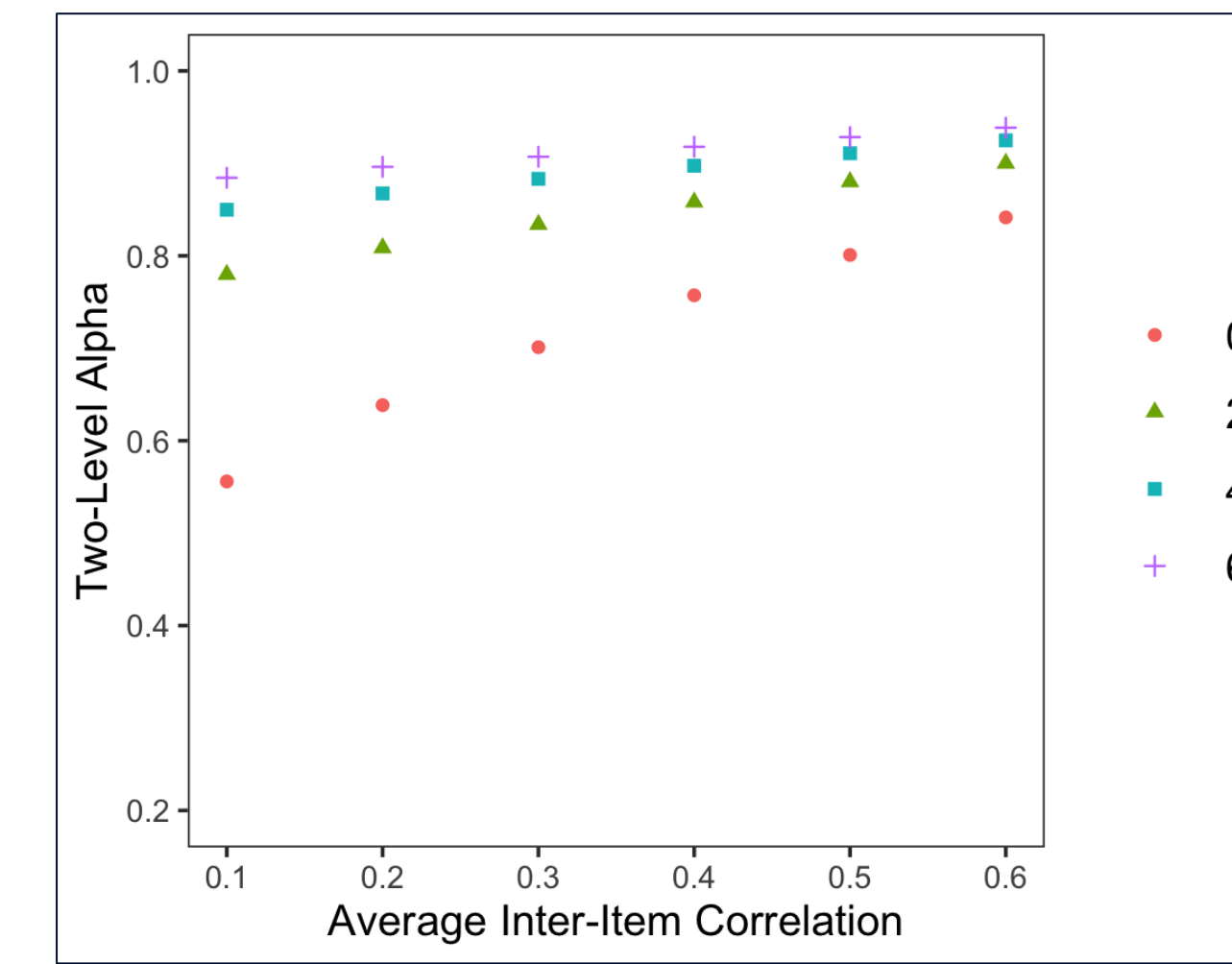
Preliminary Findings

Internal Consistency Estimates as a Function of Inter-Item Rho

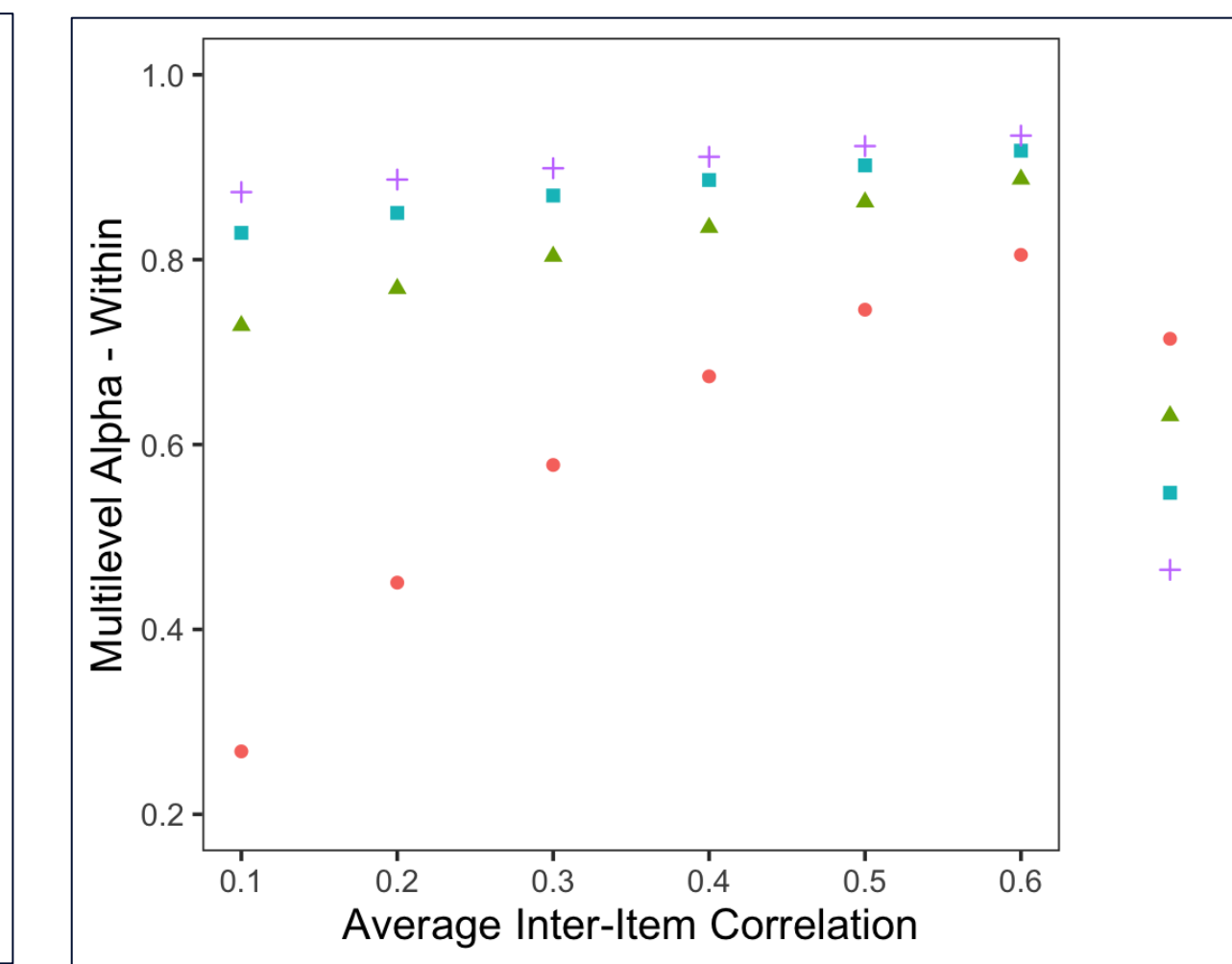
Single-Level Composite



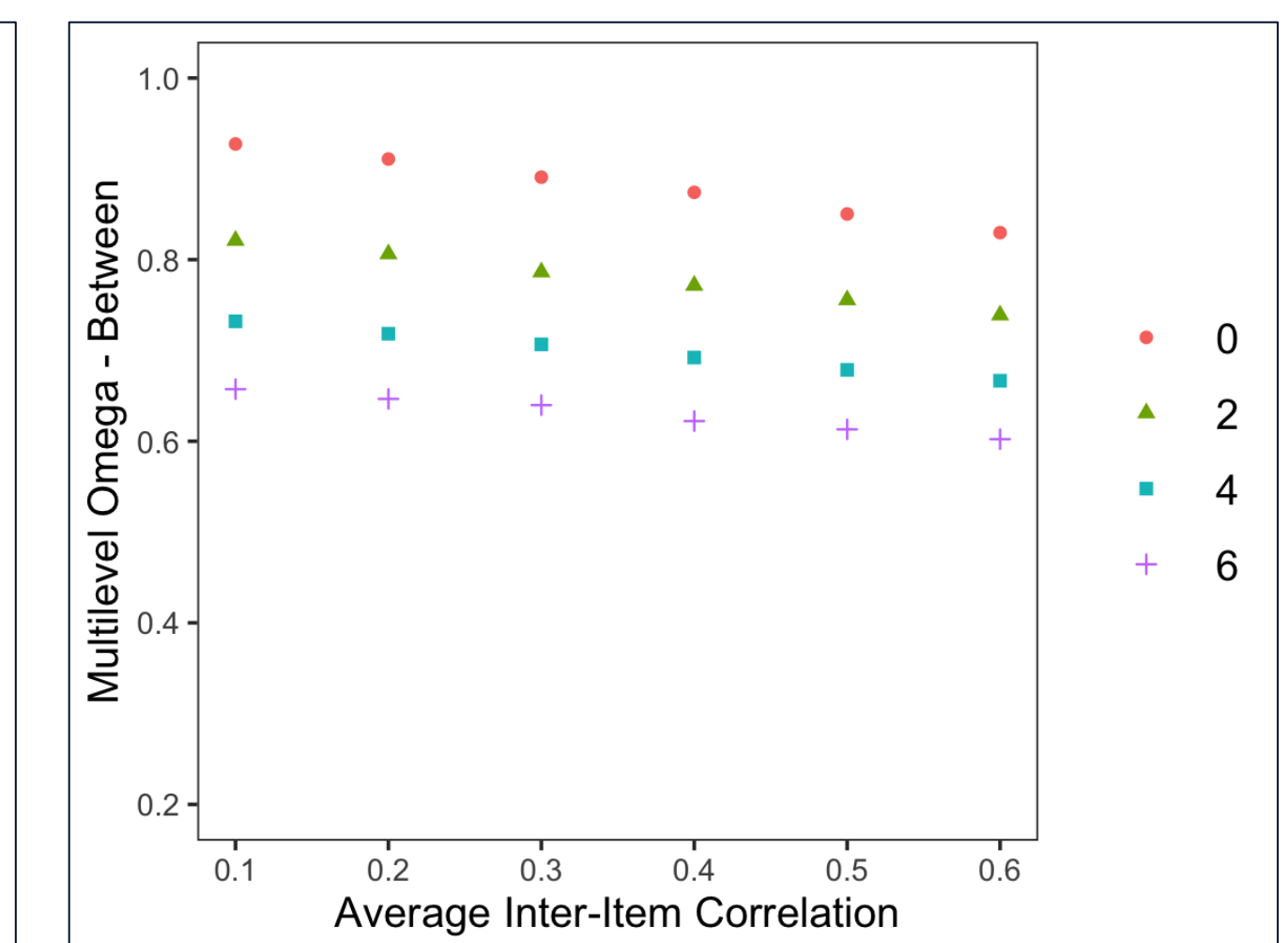
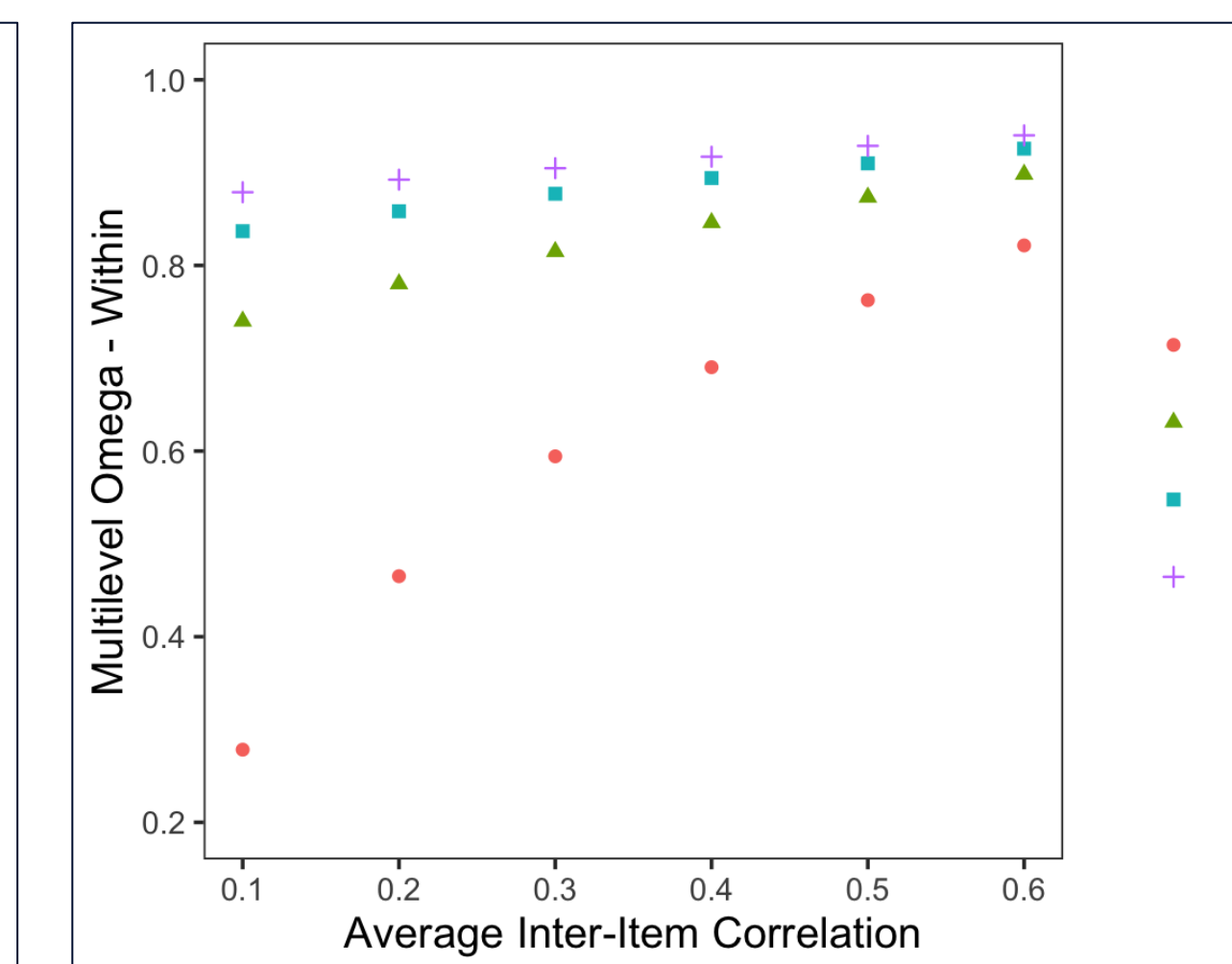
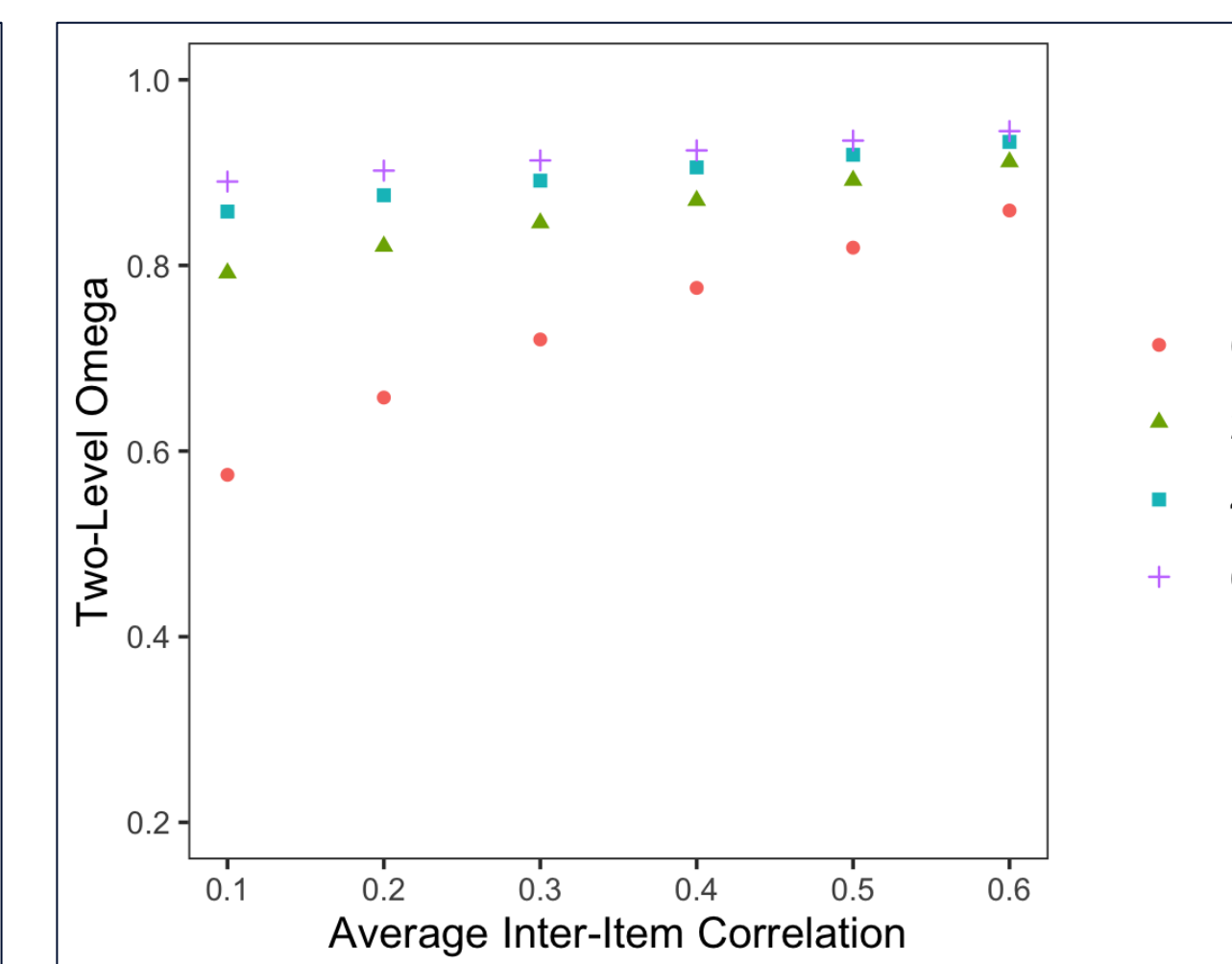
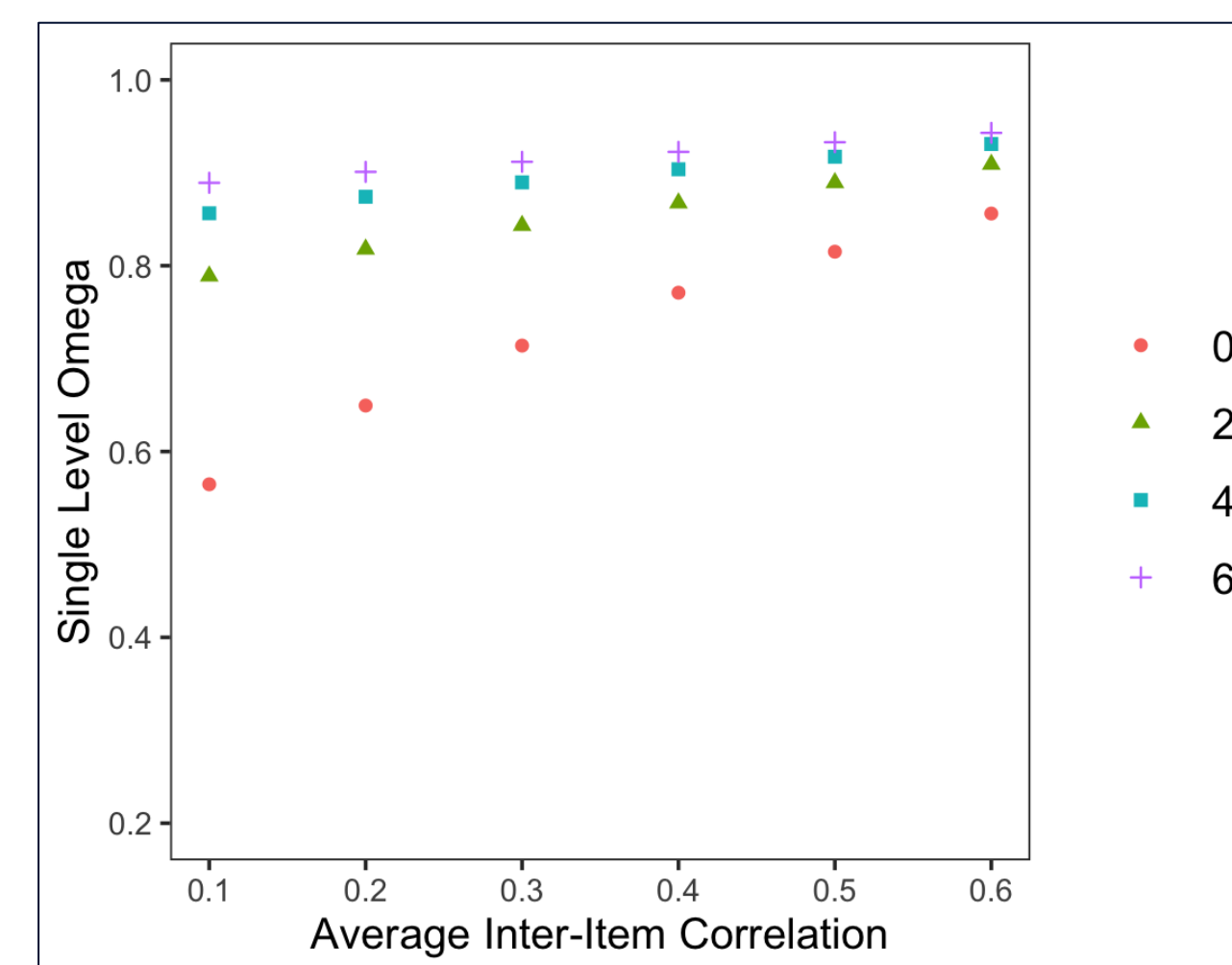
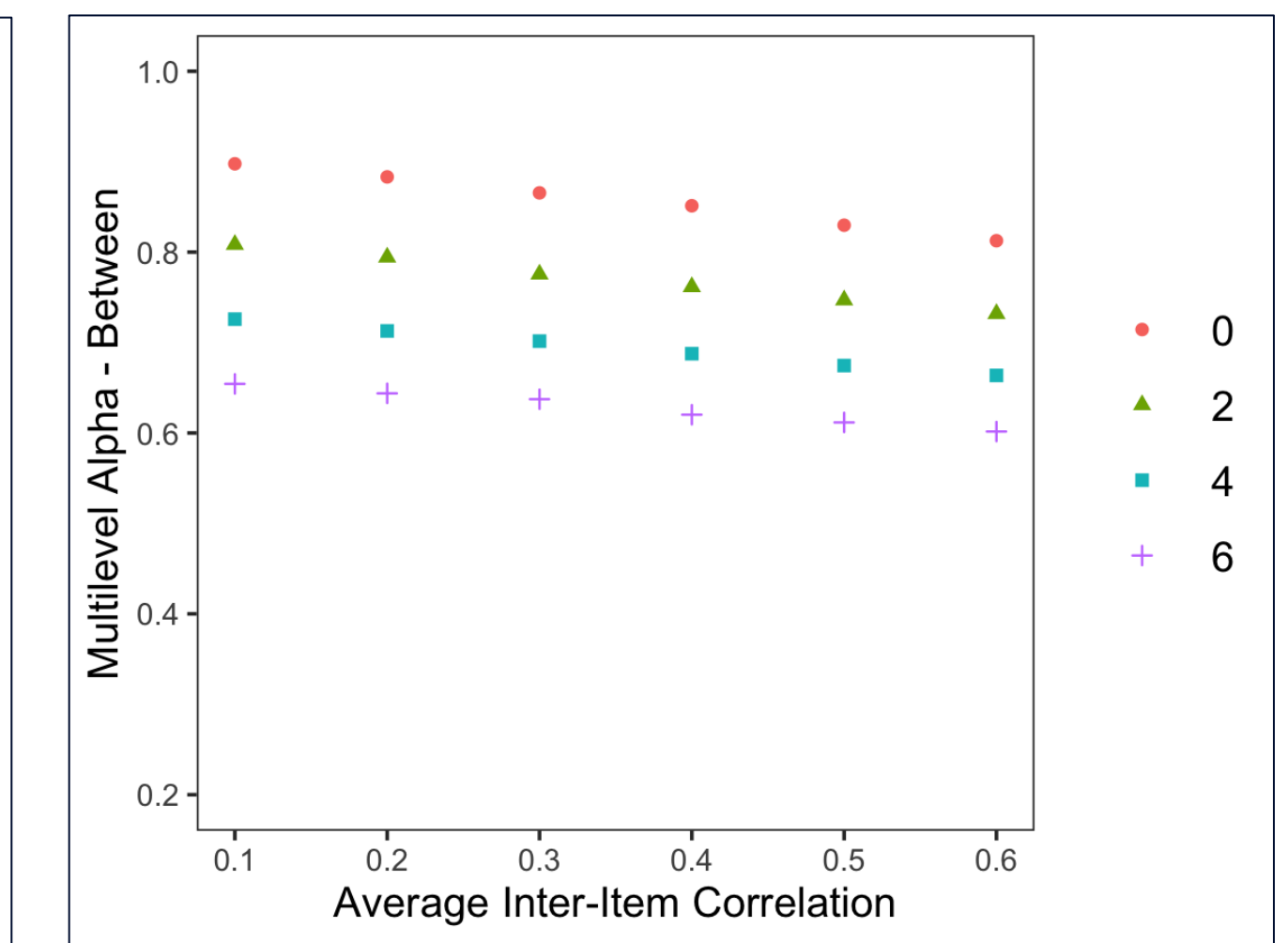
Multilevel Composite



Multilevel Within

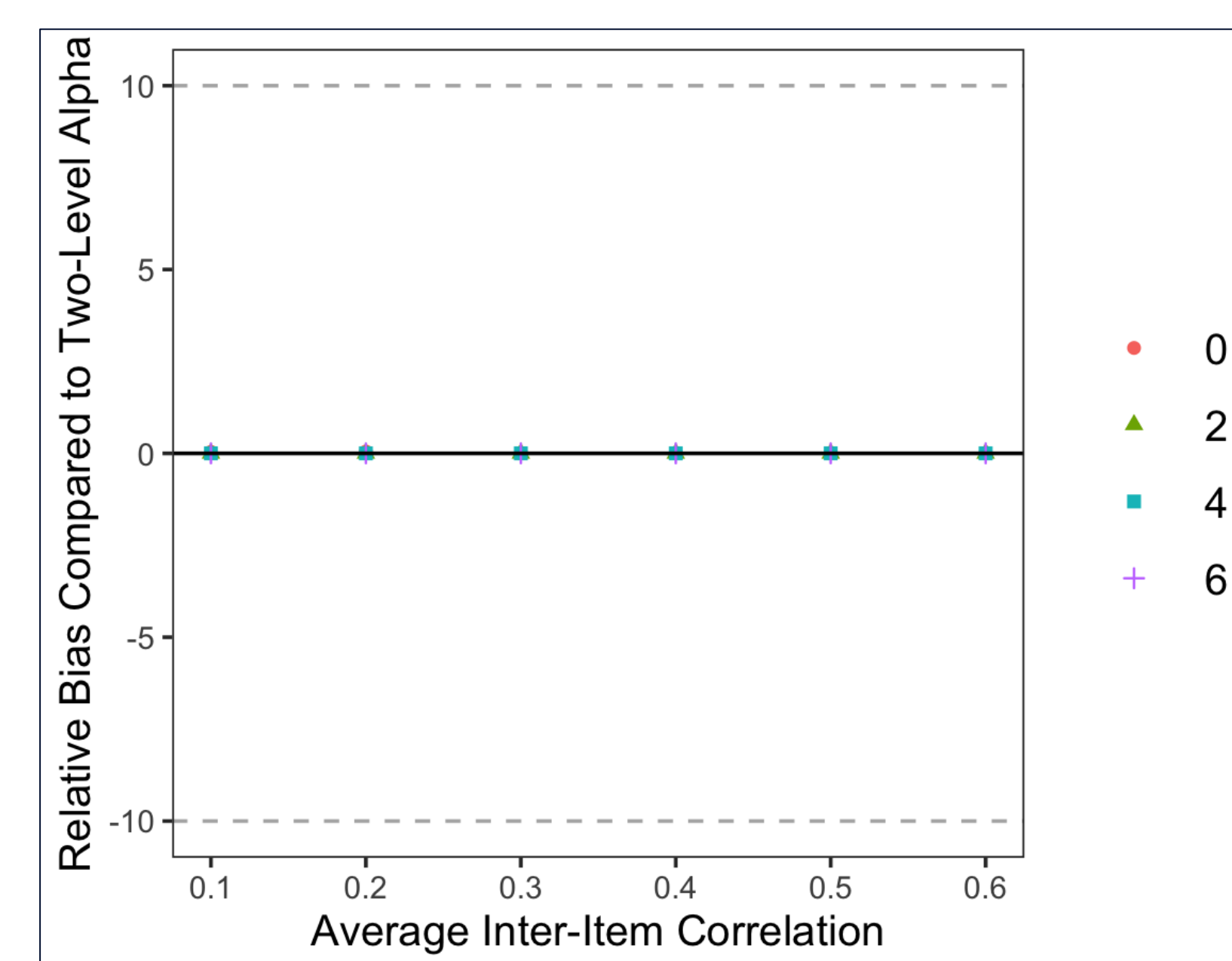


Multilevel Between

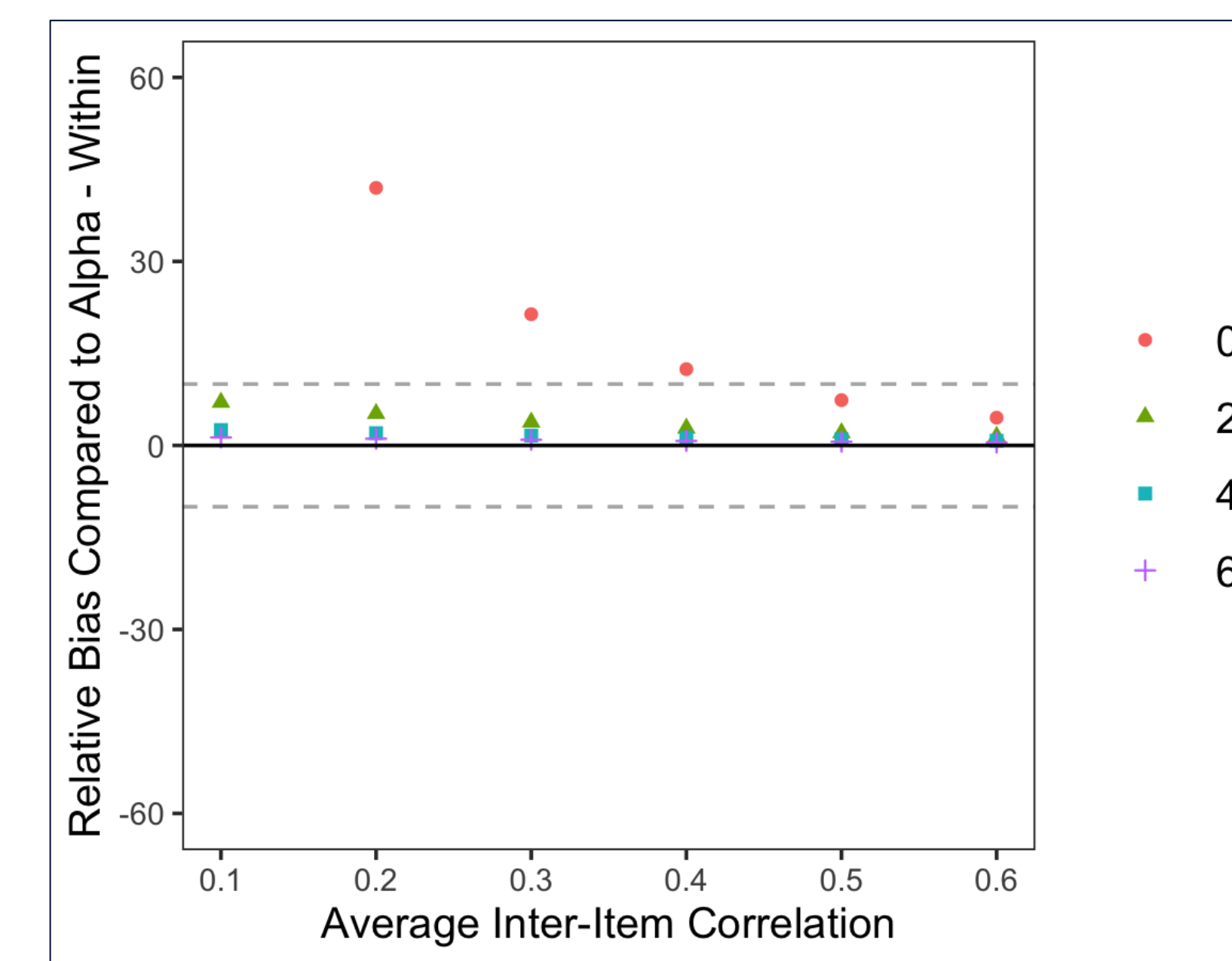


Relative Bias in Single Level Alpha as a Function of Inter-Item Rho

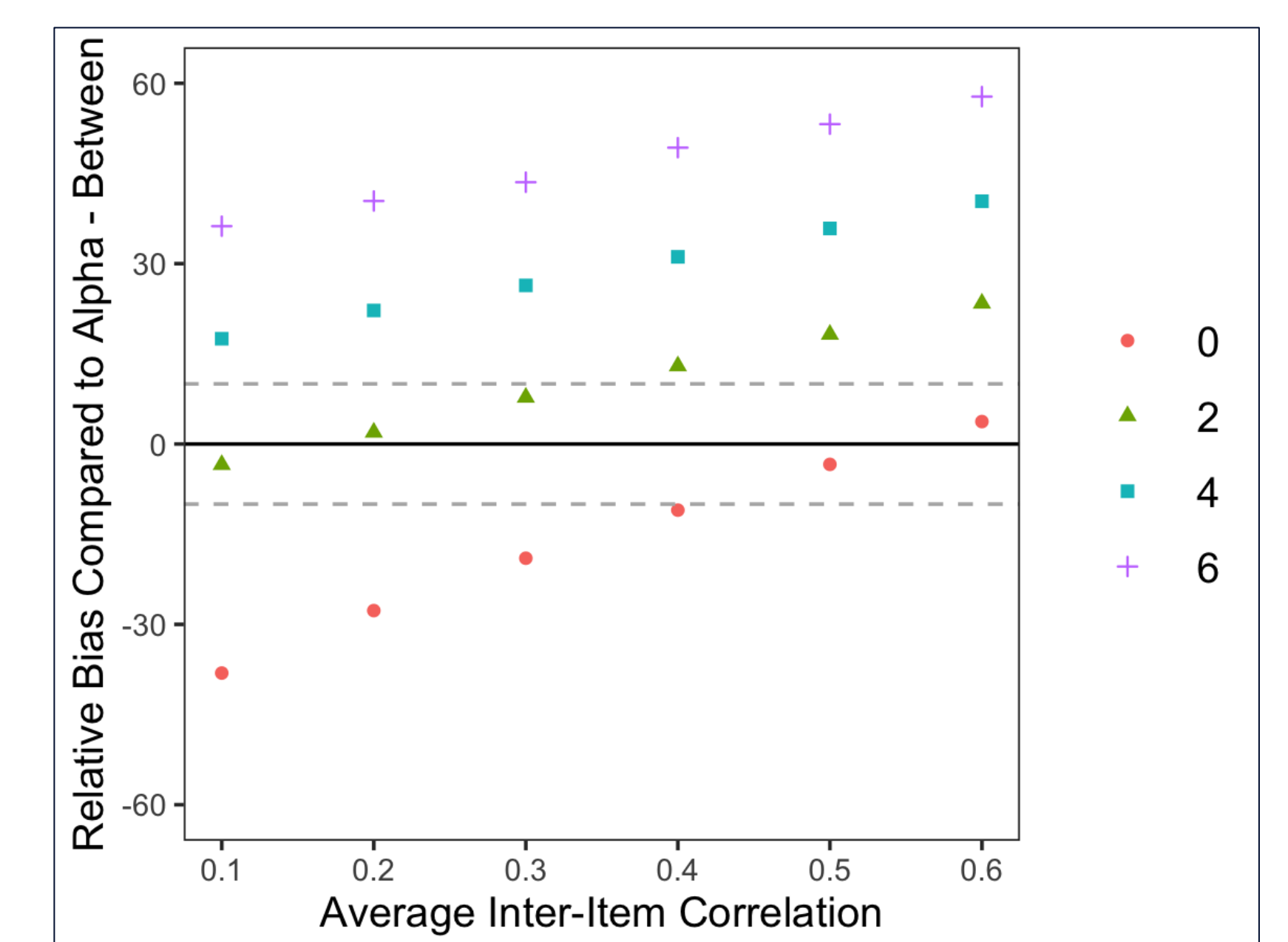
Multilevel Composite



Multilevel Within



Multilevel Between



References

Benson, N. F., Floyd, R. G., Kranzler, J. H., Eckert, T. L., Fefer, S. A., & Morgan, G. B. (2019). Test use and assessment practices of school psychologists in the United States: Findings from the 2017 national survey. *Journal of School Psychology, 72*, 29–48. <https://doi.org/10.1016/j.jsp.2018.12.004>

Center on PBIS (2024). Positive Behavioral Interventions & Supports [Website]. www.pbis.org.

Jak, S., & Jorgensen, T. D. (2017). Relating measurement invariance, cross-level invariance, and multilevel reliability. *Frontiers in Psychology, 8*. <https://doi.org/10.3389/fpsyg.2017.01640>

Lai, M. H. (2021). Composite reliability of multilevel data: It's about observed scores and construct meanings. *Psychological Methods, 26*(1), 90–102.

Lebrun-Harris, L. A., Ghandour, R. M., Kogan, M. D., & Warren, M. D. (2022). Five-year trends in US children's health and well-being, 2016-2020. *JAMA Pediatrics, 176*(7), e220056-e220056. <https://doi.org/10.1001/jamapediatrics.2022.0056>

Oakes, W. P., Lane, K. L., Cantwell E. D., & Royer, D. J. (2017). Systematic screening for behavior in K-12 settings as regular school practice: Practical considerations and recommendations. *Journal of Applied School Psychology, 33*, 369-393. <https://doi.org/10.1080/15377903.2017.1345813>

Pelton, K. S. L., Lane, K. L., Oakes, W. P., Buckman, M. M., Royer, D. J., Sherod, R. L. (2023). Mapping the research base for universal behavior screeners. *Manuscript submitted for review*.

Schatschneider, C., Lane, K. L., Oakes, W. P., & Kalberg, J. R. (2014). The Student Risk Screening Scale: Exploring Dimensionality and Differential Item Functioning. *Educational Assessment, 19*(3), 185–203. <https://doi.org/10.1080/10627197.2014.934608>