Bayesian multilevel factor analysis of the Emotion Regulation Checklist

EXAMINING CROSS-LEVEL AND LONGITUDINAL MEASUREMENT INVARIANCE

STEVEN J. PIERCE, KYUNGSOOK LEE, HOLLY BROPHY-HERB, & LAURIE A. VAN EGEREN

Introduction

We are evaluating the efficacy of an early childhood science education curriculum (Head Start on Science; Ritz, 2007) and teacher professional development program for improving the following in settings serving low-income, diverse children and families:

- Teachers' attitudes and practices related to science instruction.
- Children's scientific reasoning and knowledge, language, math, and socioemotional outcomes.
- Parents' attitudes about the value of science education and provision of opportunities for exposure to science for their children.

Objective

We wanted a reliable measure of preschool children's ability to regulate emotions that was comparable both over time and across classrooms.

MICHIGAN STATE UNIVERSITY

This research was funded by grant #DRL-1119327 from the National Science Foundation.

Research Design and Methods

Multi-site, cluster randomized longitudinal design with two 2-year cohorts.

- 8 sites (Head Start programs serving racially/ethnically diverse populations).
- 66 classrooms randomly assigned to intervention vs. control. Ten classrooms dropped out before child data collection, leaving 56 for this analysis.
- 522 children were rated by their teachers. Each teacher rated multiple children (Mean N_j = 9.3).
- 2 sets of ratings per child (Time 1 in Fall, Time 2 in Spring).

Both teachers and parents rated children on the 24-item Emotion Regulation Checklist (ERC; Shields & Cicchetti, 1997, 1998; Miller et al., 2004).

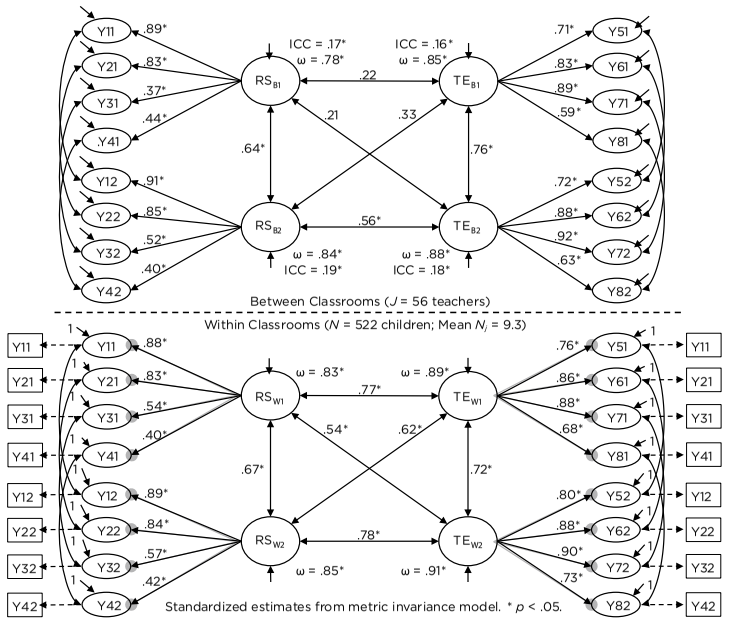
- 4-point Likert response format (treated as ordinal data).
- We could not replicate the published factor structure, so we used item content to select 8 core items for further analysis.

Multilevel EFA with half the Time 1 data suggested two 4-item factors: regulatory skills and temperament. We report the final multilevel CFA on the full dataset of teachers' ratings here.

- Used Bayesian estimation to better handle small classroom sample size (Hox, Van de Schoot, & Matthijsse, 2012; Muthén & Asparouhov, 2012) and ordinal indicators (Liang & Yang, 2014).
- Estimated ICC for latent factors and items (Jak, Oort, & Dolan, 2014).
- Estimated level-specific reliability (ω) of factors (Geldhof, Preacher, & Zyphur, 2014; Raykov & Marcoulides, 2011).

We tested measurement invariance longitudinally (Coertjens et al., 2012) and across levels (Jak et al., 2014) simultaneously.

- Longitudinal residual correlations among parallel items are method effects.
- Tested for exact configural invariance first (loadings free across time & level).
- Tested for exact metric invariance (equal loadings across time & level).
- Bayesian estimator doesn't support threshold constraints in multilevel models, preventing tests of exact longitudinal scalar invariance.



Parameter Type	Priors	Constraints
Factor loadings	N(0,1)*	Free for within time 1; within time 2; between = within.
Latent & res. variances	IW(1,3)**	Within res. variances = 1 (theta parameterization).
Latents covariances	IW(0,3)**	Free.
Latent means	N/A	Fixed = 0 for identification.
Res. covariances	IW(0,3)**	Free longitudinally, else fixed = 0.
Thresholds	N(0,5)**	Free (constraints not supported).

*Informative, lowers bias (Asparouhov & Muthen, 2010). **Non informative (Liang & Yang, 2014).

Findings

Model fit was adequate, but far from ideal. Adding approximate zero cross-loadings with informative priors may improve model fit (Muthén & Asparouhov, 2012).

Our model of regulatory skills and temperament measures demonstrates:

- Substantial non-independence due to classroom-level variability (ICCs ≥ .15 for all individual items and for the latent factors).
- Simultaneous longitudinal and cross-level metric invariance with teachers' ratings of children for both factors.
- High composite reliability across time and level for both factors.
- Bayes factors testing whether between-classroom residual variances < 0.01 all suggest thresholds are not invariant across classrooms

Model	Sampler	Chains	BITER	Thin	Convergence	# Param.	95% CI (χ^2_O-χ^2_R	PPP
Configural MI	Gibbs PX1	4	(10,000)	100	0.01	132	[-14.58, 142.10]	0.056
Metric MI	Gibbs PX1	4	(10,000)	100	0.01	108	[-12.16, 141.96]	0.051

Factor/Item	Time 1			Time 2			T1<->T2	
	ICC	R^2_W	R^2_B	ICC	R^2_W	R^2_B	r_W	r_B
Regulatory skills (RS)								
1. Transitions well from one activity to another	.17*	.78	.80	.18*	.80	.83	.33*	.21
2. Can recover quickly from upset or distress	.18*	.68	.69	.19*	.71	.73	.29*	.34
3. Is able to delay gratification	.31*	.30	.14	.22*	.33	.27	.22*	.76*
4. Can say when s/he is feeling sad, angry or mad, fearful or afraid	.15*	.16	.19	.20*	.18	.16	.33*	.32
Temperament (TE)								
5. [Does not] exhibit wide mood swings	.18*	.57	.51	.21*	.64	.52	.16	.32
6. Is [not] easily frustrated	.17*	.73	.68	.18*	.78	.77	.00	.24
7. Is [not] prone to angry outbursts/tantrums easily	.16*	.77	.78	.17*	.82	.85	.30	.02
8. Is [not] impulsive	.20*	.46	.34	.22*	.53	.39	.47*	.60*

Note: Likert format: 1. Rarely/never; 2. Sometimes; 3. Often; 4. Almost always. *p < .05.

Discussion

The ERC was developed with single-level samples and single-level analyses. Our work is the first to examine its psychometric properties in a multilevel sample with multilevel analysis. We constructed two factors (regulatory skills and temperament) that demonstrate good reliability and metric invariance.

Single level analyses of parents' ERC ratings of the children yielded a similar factor structure with longitudinal metric invariance for regulatory skills and temperament (results omitted for brevity). This increases the utility of the measures we created.

Software limits prevented directly testing longitudinal scalar invariance. Equivalence tests (Mascha & Sessler, 2011) on thresholds via inequality hypotheses (van de Schoot et al., 2012) may be a viable alternative; we will investigate this option soon.

Cross-level scalar invariance implies perfect between-classroom reliability (ω = 1) because residual variances must be zero (Geldhof et al., 2014; Jak et al., 2014). So, our high reliabilities (ω ≥ .78) suggest that the cluster bias caused by threshold non-invariance across classrooms is small enough to proceed with examining intervention effects.

CHILD CHARACTERISTICS (N = 522)

Variable	n (%)
Female	262 (50.2%)
Hispanic	125 (24%)
Race	
AA	74 (16%)
AI/NA	23 (5%)
White	247 (52%)
Other	127 (27%)
Age (years)	3.7 (SD = .48)

References

Coertjens, L., Donche, V., De Maeyer, S., Vanthournout, G., & Van Petegem, P. (2012). Longitudinal measurement invariance of Likert-type learning strategy scales: Are we using the same ruler at each wave? *Journal of Psychoeducational Assessment, 30*(6), 577-587. doi:10.1177/0734282912438844

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19*(1), 72-91. doi:10.1037/a0032138

Hox, J., van de Schoot, R., & Matthijsse, S. (2012). How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods, 6*(2), 87-93.

Jak, S., Oort, F. J., & Dolan, C. V. (2014). Using two-level factor analysis to test for cluster bias in ordinal data. *Multivariate Behavioral Research, 49*(6), 544-553. doi:10.1080/00273171.2014.947353

Liang, X., & Yang, Y. (2014). An evaluation of WLSMV and Bayesian methods for confirmatory factor analysis with categorical indicators. *International Journal of Quantitative Research in Education, 2*(1), 17-38. doi:10.1504/IJQRE.2014.060972

Mascha, E. J., & Sessler, D. I. (2011). Equivalence and noninferiority testing in regression models and repeated measures designs. *Anesthesia & Analgesia, 112*(3), 678-687. doi:10.1213/ANE.0b013e318206f872

Miller, A. L., Gouley, K. K., Seifer, R., Dickstein, S., & Shields, A. (2004). Emotions and behaviors in the Head Start classroom: Associations among observed dysregulation, social competence, and preschool adjustment. *Early Education & Development, 15*(2), 147-166. doi:10.1207/s15566935eed1502_2

Muthén, B. O., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods, 17*(3), 313-335. doi:10.1037/a0026802

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory.* New York, NY: Routledge.

Ritz, W. C. (Ed.). (2007). *A head start on science: Encouraging a sense of wonder.* Arlington, VA: National Science Teachers Association Press. www.nsta.org.

Shields, A., & Cicchetti, D. (1997). Emotion regulation among school-age children: The development and validation of a new criterion Q-sort scale. *Developmental Psychology, 33*(6), 906-916.

Shields, A., & Cicchetti, D. (1998). Reactive aggression among maltreated children: The contributions of attention and emotion dysregulation. *Journal of Clinical Child Psychology, 27*(4), 381-395. doi:10.1207/s15374424jccp2704_2

van de Schoot, R., Hoijtink, H., Hallquist, M. N., & Boelen, P. A. (2012). Bayesian evaluation of inequality-constrained hypotheses in SEM models using Mplus. *Structural Equation Modeling, 19*(4), 593-609. doi:10.1080/10705511.2012.713267