Empirical Benchmarks for Between-Case Standardized Mean Differences from Single-Case Multiple Baseline Designs Examining Academic Interventions

James E. Pustejovsky
Man Chen
David Klingbeil
Ethan van Norman

American Educational Research Association
2023 Annual Convention
Chicago, IL, April 16th, 2023

Acknowledgement

- The work reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant R305D190023. The opinions expressed are those of the author and do not represent the views of the Institute or the U.S. Department of Education.

Man Chen

Ethan van Norman

David Klingbeil

Outline

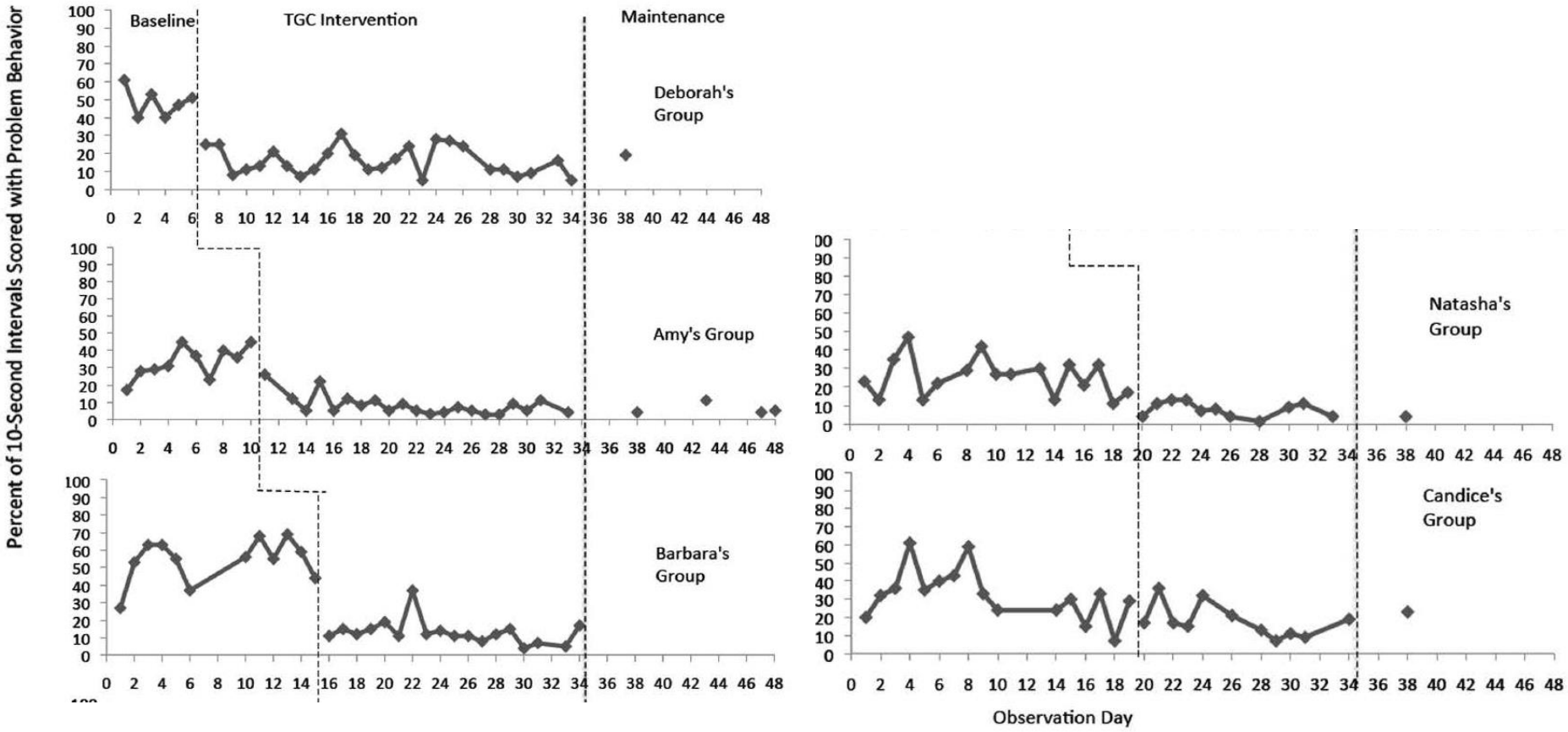
- **Single-case multiple baseline designs**
- Between-case standardized mean differences
- Systematic review of academic multiple baseline designs
- Analytic strategy
- Empirical benchmarks

Single-case multiple baseline designs

- Used for investigating effects of interventions / practices for individual participants across a variety of settings

- Essential features of multiple baseline designs
  - One or small number of participants
  - Repeated measurement of outcomes on each individual participant
  - Researcher controls introduction of intervention for each participant
  - Intervention initiation is staggered in time

Rodriguez & Anderson (2014). Integrating a social behavior intervention during small group academic instruction using a total group criterion intervention



Design schematic

A multiple baseline design across four participants

T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18
X	X	X	X	X	T	X	X	X	X	X	X	X	X	X	X	X	X
X	X	X	X	X	X	X	X	T	X	X	X	X	X	X	X	X	X
X	X	X	X	X	X	X	X	X	X	X	T	X	X	X	X	X	X
X	X	X	X	X	X	X	X	X	X	X	X	X	X	T	X	X	X	X

Outline

- Single-case multiple baseline designs
- **Between-case standardized mean differences**
- Systematic review of multiple baseline designs
- Analytic strategy
- Empirical benchmarks

Between-case standardized mean difference (BC-SMD)

- Shadish, Rindskopf, & Hedges (2008) asked:

  Can we estimate an effect size based on the data from a single-case design that is in the same metric as the standardized mean difference effect size from a between-groups design?

- Why do this? (Shadish, Hedges, Horner, & Odom, 2015)
  - **Translation** of single-case research for researchers who work primarily with between-groups designs.
  - **Comparison** of results from single-case studies and between-groups studies, for purposes of understanding the utility and limitations of each type of design.
  - **Synthesis** involving both single-case and between-groups designs.

SMD in between-group experiments

- What is the SMD from a between-groups experiment?

$$\delta_{BC} = \frac{\begin{pmatrix}\text{Average outcome if}\\ \text{everybody gets treatment}\end{pmatrix} - \begin{pmatrix}\text{Average outcome if}\\ \text{nobody gets treatment}\end{pmatrix}}{\begin{pmatrix}\text{Outcome standard dev.}\\ \text{nobody gets treatment}\end{pmatrix}}$$

$$\delta_{BC} = \frac{\begin{pmatrix}\text{Average outcome if}\\ \text{everybody gets treatment}\end{pmatrix} - \begin{pmatrix}\text{Average outcome if}\\ \text{nobody gets treatment}\end{pmatrix}}{\sqrt{\begin{pmatrix}\text{Within participant}\\ \text{variance}\end{pmatrix} + \begin{pmatrix}\text{Between participant}\\ \text{variance}\end{pmatrix}}}$$

- These quantities can be estimated from multiple baseline design data using a hierarchical linear model.
  - We'll need to have a sample of multiple participants (bare minimum of 3, more for more complex models).
  - We'll need to be specific about ***timing*** of intervention and follow-up.

Estimating BC-SMDs: The broad strategy

Pustejovsky, Hedges, and Shadish (2014):

1. Develop a hierarchical linear model that describes
   a) the form of time trends and intervention effects
   b) how the trends and intervention effects ***vary across cases***.

2. Imagine a ***hypothetical between-groups experiment*** with the same population of participants, same treatment, same outcomes.
   - When is treatment initiated?
   - When are outcomes assessed?

3. Use the hierarchal model to estimate the between-case SMD for the hypothetical experiment.

Design translation

A multiple baseline design

T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18
X	X	X	X	X	T	X	X	X	X	X	X	X	X	X	X	X	X
X	X	X	X	X	X	X	X	T	X	X	X	X	X	X	X	X	X
X	X	X	X	X	X	X	X	X	X	X	T	X	X	X	X	X	X
X	X	X	X	X	X	X	X	X	X	X	X	X	X	T	X	X	X

A hypothetical between-group design (with pre-test)

T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18
				X	T										X		
				X	T										X		
				X											X		
				X											X		

Motivation

- BC-SMD has been used in many primary single-case design studies, as well as many systematic reviews of single-case research.
- What Works Clearinghouse recently adopted BC-SMDs for summarizing findings from single-case designs.
- But no reference benchmarks available.
  - Theoretically comparable to between-group effect sizes.
  - But multiple baselines are used in different contexts that group designs, so existing group design benchmarks are probably not be appropriate.

Outline

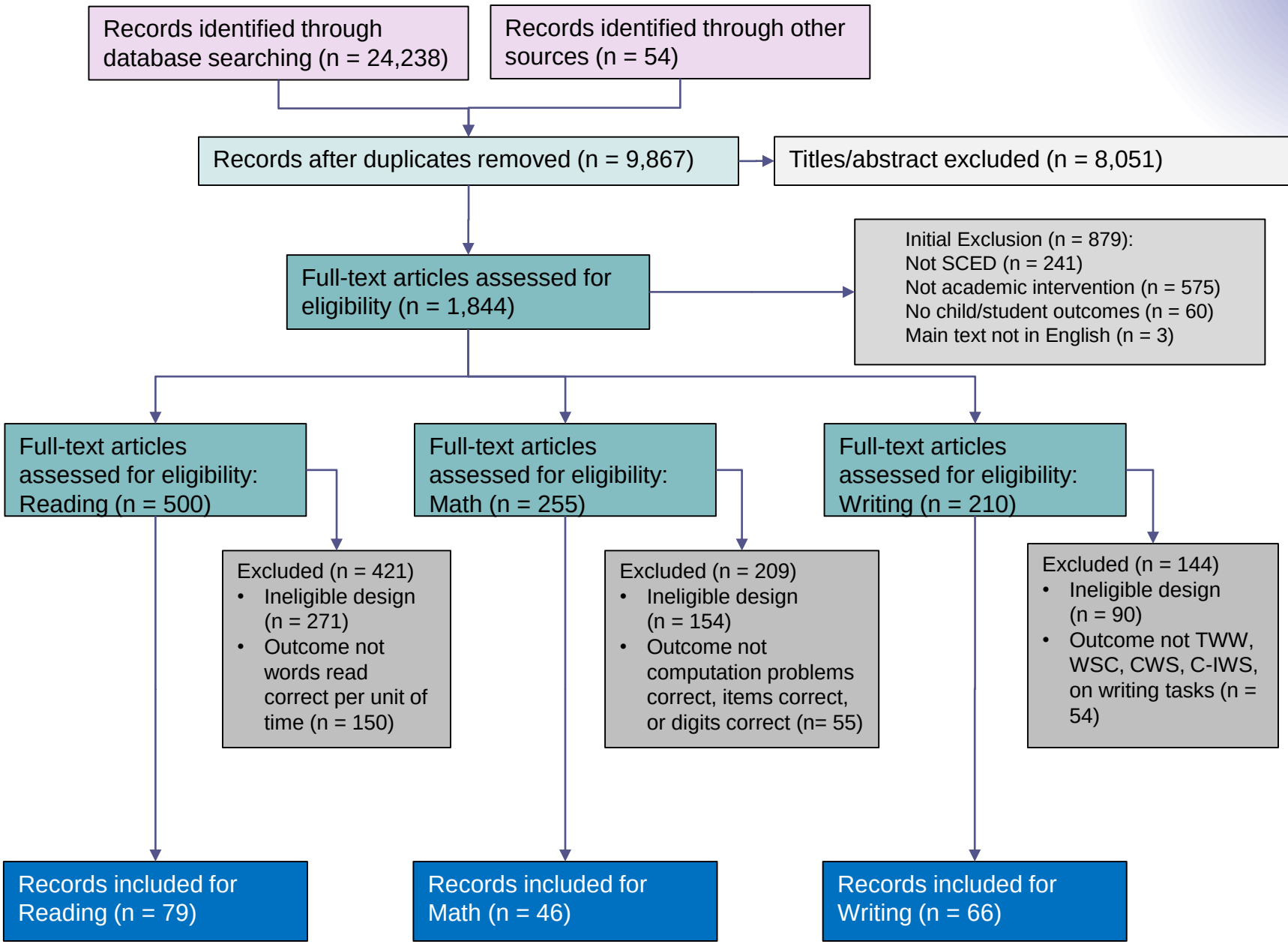
- Single-case multiple baseline designs
- Between-case standardized mean differences
- **Systematic review of multiple baseline designs**
- Analytic strategy
- Empirical benchmarks

Inclusion criteria and search strategy

- **Design:** Across-participant multiple baseline design with 3+ participants.
- **Participants:** Students in pre-kindergarten through 12th grade (or Special Education up to age 21)
- **Intervention:** Any intervention targeting an academic skill
- **Comparison:** Baseline prior to intervention
- **Outcomes:** Specific, curriculum-based measures of math, reading, or writing

- Databases: Academic Search Complete, ERIC, PsycInfo

- Search string: “single-case” AND (“read*” OR “math*” OR “writ*” OR “spell*” OR “academic*” OR “learn*”)


Records identified through database searching (n = 24,238)
Records identified through other sources (n = 54)
Records after duplicates removed (n = 9,867)
Titles/abstract excluded (n = 8,051)
Full-text articles assessed for eligibility (n = 1,844)
Initial Exclusion (n = 879):
Not SCED (n = 241)
Not academic intervention (n = 575)
No child/student outcomes (n = 60)
Main text not in English (n = 3)
Full-text articles assessed for eligibility: Reading (n = 500)
Full-text articles assessed for eligibility: Math (n = 255)
Full-text articles assessed for eligibility: Writing (n = 210)
Excluded (n = 421)
• Ineligible design (n = 271)
• Outcome not words read correct per unit of time (n = 150)
Excluded (n = 209)
• Ineligible design (n = 154)
• Outcome not computation problems correct, items correct, or digits correct (n= 55)
Excluded (n = 144)
• Ineligible design (n = 90)
• Outcome not TWW, WSC, CWS, C-IWS, on writing tasks (n = 54)
Records included for Reading (n = 79)
Records included for Math (n = 46)
Records included for Writing (n = 66)

Outline

- Single-case multiple baseline designs
- Between-case standardized mean differences
- Systematic review of multiple baseline designs
- **Analytic strategy**
- Empirical benchmarks

Estimating BC-SMD effect sizes

- Initial visual analysis of every included design to determine appropriate functional forms for modeling.

- Effect size estimates generated using scdhlm R package (Pustejovsky, Chen, & Hamilton, 2022) based on a model with
  - Linear time trends for baseline phases
  - Intervention-by-time interactions
  - Random intercepts (but no random slopes)
  - Auto-correlated errors (AR1)

- Hypothetical between-group design parameters (defaults)
  - Intervention time equal to actual intervention time for first participant
  - Follow-up time based on the length of the shortest intervention phase

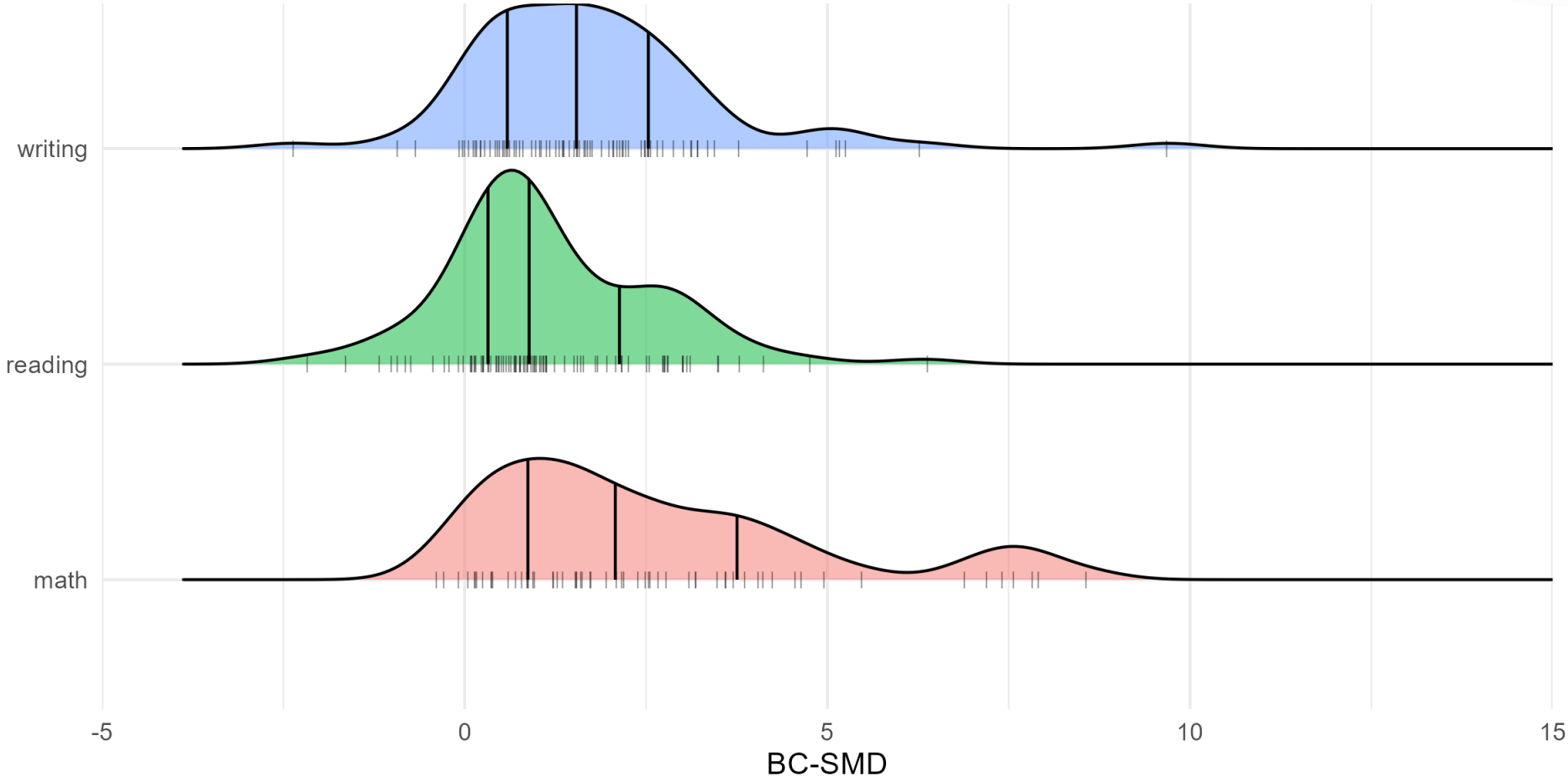
Summarizing distribution of effect sizes

- Empirical distribution of effect size estimates
- Meta-analytic model
  - Multi-level random effects
  - Prediction intervals for center of distribution
- Distribution of Empirical Bayes estimates
  - Non-parametric bootstrap intervals

Outline

- Single-case multiple baseline designs
- Between-case standardized mean differences
- Systematic review of multiple baseline designs
- Analytic strategy
- **Empirical benchmarks**

Empirical densities



Meta-analysis and empirical Bayes

Domain	Estimator	10th	20th	30th	40th	50th	60th	70th	80th	90th
Math	Percentile	0.21	0.77	1.10	1.54	2.08	2.59	3.49	4.20	7.13
	Meta-Analysis	-0.21	0.71	1.37	1.93	2.45	2.97	3.53	4.19	5.11
	Empirical Bayes	0.24	0.74	1.21	1.67	2.12	2.56	3.37	4.22	6.67
Reading	Percentile	-0.25	0.19	0.44	0.48	0.89	1.10	1.65	2.51	3.01
	Meta-Analysis	-0.05	0.25	0.46	0.64	0.81	0.98	1.16	1.37	1.67
	Empirical Bayes	0.11	0.39	0.57	0.73	0.84	0.94	1.09	1.29	1.60
Writing	Percentile	0.14	0.49	0.77	1.30	1.54	1.97	2.22	2.67	3.30
	Meta-Analysis	0.10	0.59	0.95	1.25	1.53	1.81	2.11	2.47	2.96
	Empirical Bayes	0.30	0.74	0.96	1.33	1.55	1.81	2.05	2.40	3.02

Reference benchmarks

- Using middle 40% of distribution (30th-70th percentile)
- empirical Bayes estimates

Domain		30th		70th	
Math	Small	1.2	Medium	3.4	Large
Reading	Small	0.6	Medium	1.1	Large
Writing	Small	1.0	Medium	2.1	Large

Observations and limitations

- Compared to group designs, distributions of BC-SMD effects from single-case multiple baseline designs cover ***substantially larger values*** and are ***more dispersed***.

- Differences between designs could be due to differences in
  - Populations
  - Interventions
  - Dependent variables
  - Settings
  - Temporal horizons

- BC-SMDs from multiple baseline designs are sensitive to follow-up time.

- Critical to interpret findings within the context of the topic area and based on the logic of single-case designs.

References

Chen, M., Pustejovsky, J. E., Klingbeil, D. A., & Van Norman, E. (2022, March 22). Between-Case Standardized Mean Differences: Flexible Methods for Single-Case Designs. https://doi.org/10.31234/osf.io/3pk5q

Hedges, L. V, Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods*, *3*, 224–239. https://doi.org/10.1002/jrsm.1052

Hedges, L. V, Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods*, *4*(4), 324–341. https://doi.org/10.1002/jrsm.1086

Pustejovsky, J. E., Chen, M., & Hamilton, B. (2022). scdhlm: A web-based calculator for between-case standardized mean differences (Version 0.5.2) [Web application]. https://jepusto.shinyapps.io/scdhlm

Pustejovsky, J. E., Hedges, L. V, & Shadish, W. R. (2014). Design-comparable effect sizes in multiple baseline designs: A general modeling framework. *Journal of Educational and Behavioral Statistics*, *39*(5), 368–393. https://doi.org/10.3102/1076998614547577

Shadish, W. R., Rindskopf, D. M., & Hedges, L. V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention*, *2*(3), 188–196. https://doi.org/10.1080/17489530802581603

Shadish, W. R., Hedges, L. V, & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. *Journal of School Psychology*, *52*(2), 123–147. https://doi.org/10.1016/j.jsp.2013.11.005

Shadish, W. R., Hedges, L. V, Horner, R. H., & Odom, S. L. (2015). *The role of between-case effect size in conducting, interpreting, and summarizing single-case research*. Washington, DC. Retrieved from http://ies.ed.gov/ncser/pubs/2015002/

Swaminathan, H., Rogers, H. J., & Horner, R. H. (2014). An effect size measure and Bayesian analysis of single-case designs. *Journal of School Psychology*, *52*(2), 213-230. https://doi.org/10.1016/j.jsp.2013.12.002

Rodriguez, B. J., & Anderson, C. M. (2014). Integrating a social behavior intervention during small group academic instruction using a total group criterion intervention. *Journal of Positive Behavior Interventions*, *16*(4), 234–245. https://doi.org/10.1177/1098300713492858