Effect Size Measures for Single-Case Research: Conceptual, Practical, and Statistical Considerations

James E. Pustejovsky
pustejovsky@wisc.edu
https://jepusto.com

April, 2023

Small-is-Beautiful (Once More)
3rd International N=1 Symposium

Acknowledgement

- The work reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grants R324U190002 and 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.
- This work has been shaped by many collaborators, although the views expressed are my own (as are any errors).

John Ferron
Wendy Machalicek
Daniel Swan



Ethan van Norman
David Klingbeil
Man Chen
Jennifer Ledford




Overview

- Premises
  - Effect sizes
  - Research Synthesis
- Three conceptual questions
  - Form of intervention effects (functional relation)
  - Level of analysis
  - Outcomes
- Practical and statistical considerations

PREMISES

Effect size

- Broadly:
  a quantitative [index] of relations among variables
  (Hedges, 2008, p. 167).

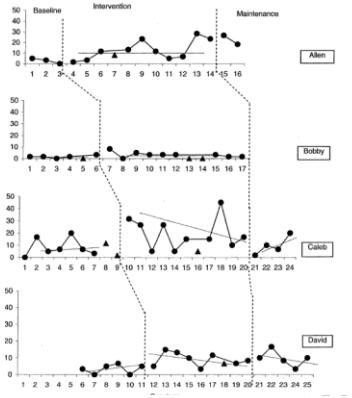
- In context of SCD / N-of-1:
  a quantitative index describing the direction and magnitude of a functional relationship (i.e., effect of intervention on an outcome) in a way that allows for comparison across cases and studies
  (Pustejovsky & Ferron, 2017)

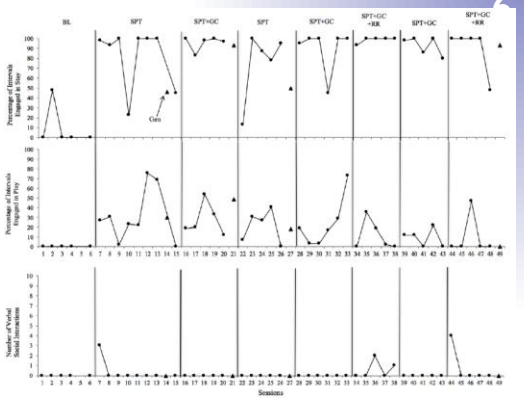
- ***"Reporting and interpreting effect sizes in the context of previously reported effects is essential to good research. It enables readers to evaluate the stability of results across samples, designs, and analyses. Reporting effect sizes also informs power analyses and meta-analyses needed in future research."***

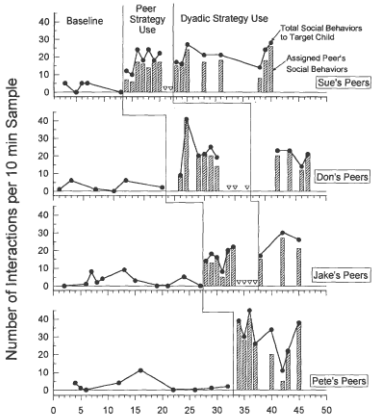
  (Wilkinson & APA Task Force on Statistical Inference, 1999)

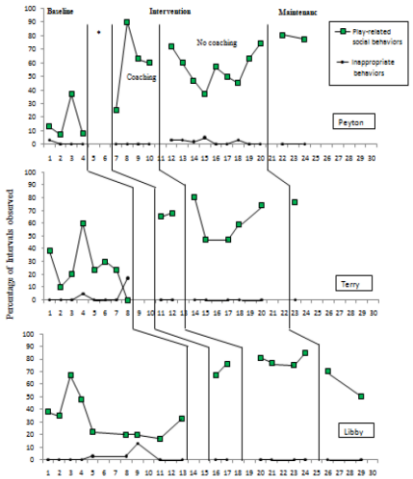
Research Synthesis and Meta-Analysis

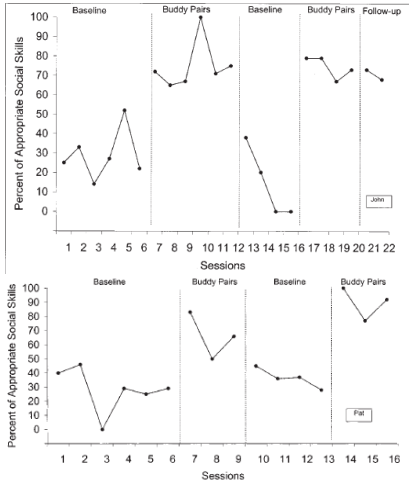
- ***Summarize*** magnitude of intervention effects.
- Characterize ***variation*** in effect magnitude.
- Identify ***systematic predictors*** of effectiveness.

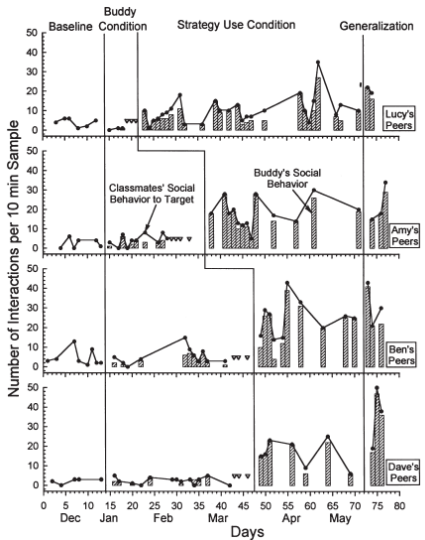


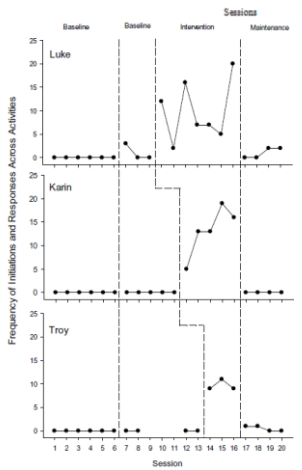






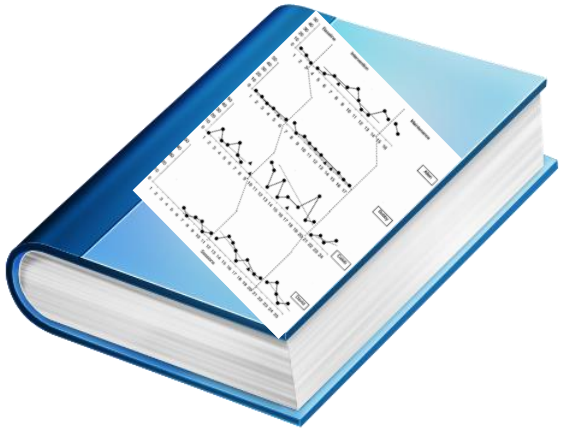


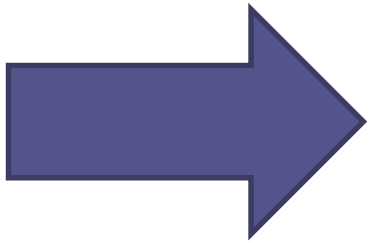


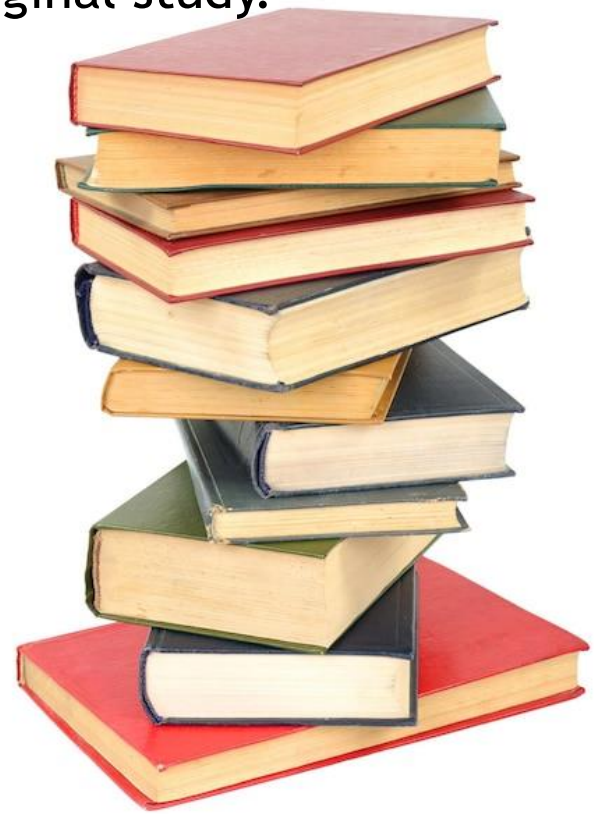


Selecting an effect size

- The goal is to relate findings from a given study to a broader literature.
  - Effect size should describe an intervention's effect in a way that makes sense beyond the context of the original study.







Effect sizes for single-case research

Non-overlap measures

- Non-overlap of all pairs (Parker & Vannest, 2009)
- Tau-AB (Parker, Vannest, Davis, & Sauber 2011)
- Percentage of non-overlapping data (PND; Scruggs et al., 1987)
- Percentage exceeding the median (PEM; Ma, 2006)
- Others: PAND, RIRD, Tau-U,…

Parametric within-case measures

- Within-case standardized mean differences (Busk & Serlin, 1992)
- Response ratio/log-response ratio (Pustejovsky, 2018)
- Ratio of medians (Bonett & Price, 2020)
- Odds ratio / log-odds ratio (Pustejovsky, 2015)
- Percentage of Goal Obtained (Ferron, Goldstein, Olszewski, & Rohrer, 2020)

Between-case standardized mean difference

- Pustejovsky, Hedges, & Shadish (2014)
- Maggin, Swaminathan, Rogers, O'Keeffe, Sugai, & Horner (2014)
- Chen, Pustejovsky, Klingbeil, & Van Norman (2023)

Raw Data Synthesis

- Van den Noortgate & Onghena (2008)
- Moeyaert, Ugille, Ferron, Beretvas, & Van den Nortgate (2013, 2014)

CONCEPTUAL QUESTIONS

- Level of analysis
- Dependent variable metric(s)
- Form of intervention effects (functional relation)

LEVEL OF ANALYSIS

Level of analysis

- Study-level average effects?
- Individual-level summary effect?
- Something more specific/detailed?

Level of analysis

	Goal/level of analysis	ES metrics	Assumptions
Study-level summary effect sizes	Study	BC-SMD	Hierarchical model of each study
Case-level summary effect sizes	Case	Non-overlap, parametric measures	Case-specific
Raw data synthesis	Time-point	Raw mean difference, within-case SMD	Hierarchical model across studies

- Higher level of analysis is more reductive, but also simpler to explain.
- Level of analysis should be determined by research aims/research questions.

DEPENDENT VARIABLE MEASURE(S)

Dependent variable measures

- ES metric needs to be meaningful and interpretable across dependent variables measured in a range of ways.

- Case-level effect size measures aim to put effects on a common scale (metric) ***even when DVs are measured in a variety of ways***.

Case-level effect size metrics

- Difference in raw scores
- Difference standardized by variability
- Proportionate change
- Distributional overlap

Difference standardized by variability

- Within-case standardized mean difference proposed by Gingerich (1984) and Busk & Serlin (1992)
- Parameter definition: $\delta = \frac{\mu_B - \mu_A}{\sigma_A}$

  $\mu_B - \mu_A$ ← Difference in means

  σ_A ← Baseline SD (within-person)

  - Difference in means, "standardized" by variability in baseline phase.
  - NOT equivalent to between-case SMD, because σ_A represents within-individual variation.

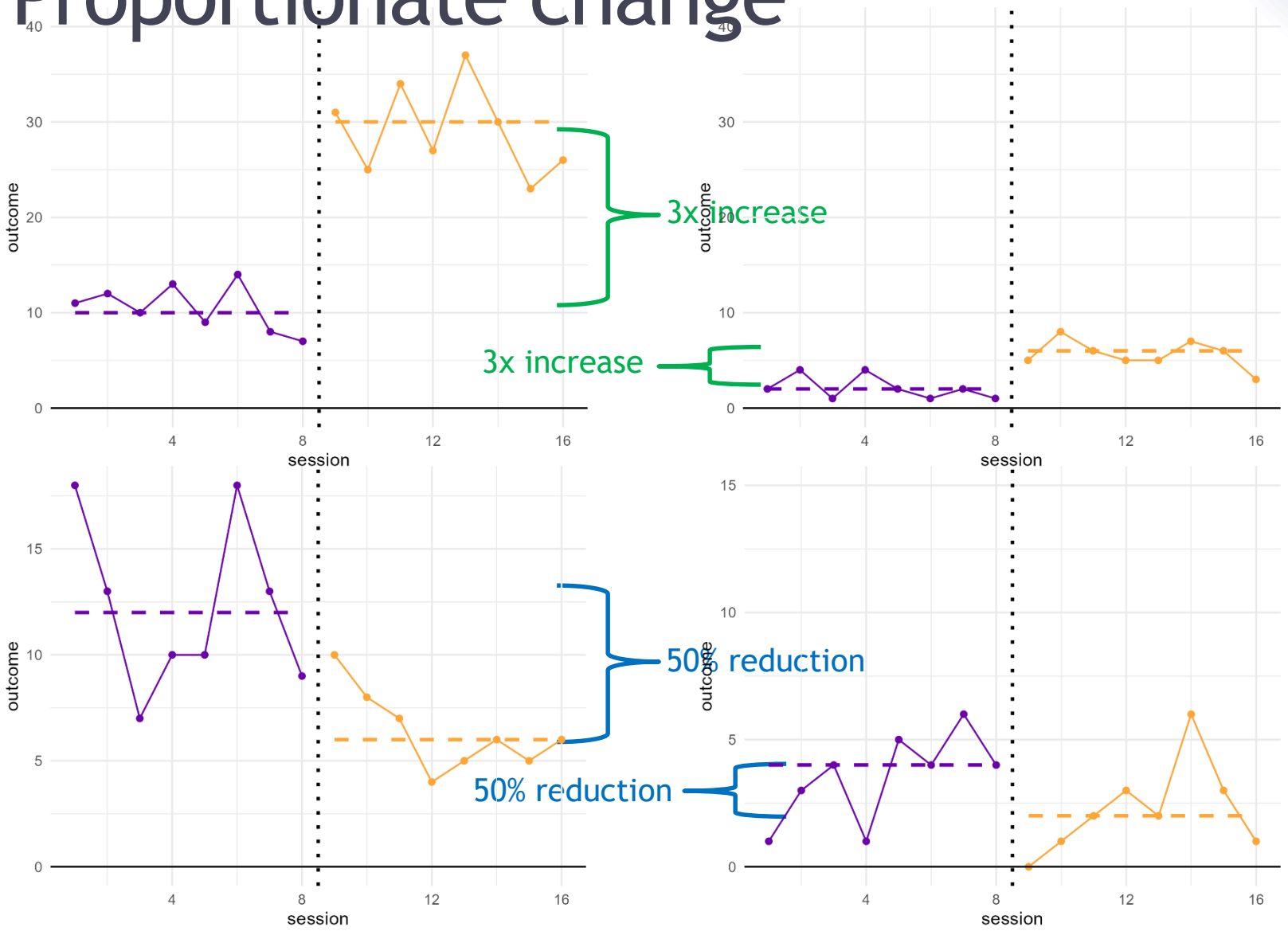
- Not a good metric if…
  - If DVs in different studies have very different reliability.
  - If DVs show little or no variation

Proportionate change

- Percentage/proportionate change from baseline to intervention is a common, easily interpretable “informal” effect size measure (Campbell & Herzinger, 2010).

- Effect size measures that use proportionate change:
  - Response ratio / log-response ratio (Pustejovsky, 2018)
  - Ratio of medians (Bonett & Price, 2020)
  - Odds ratio / log-odds ratio (Pustejovsky, 2015)

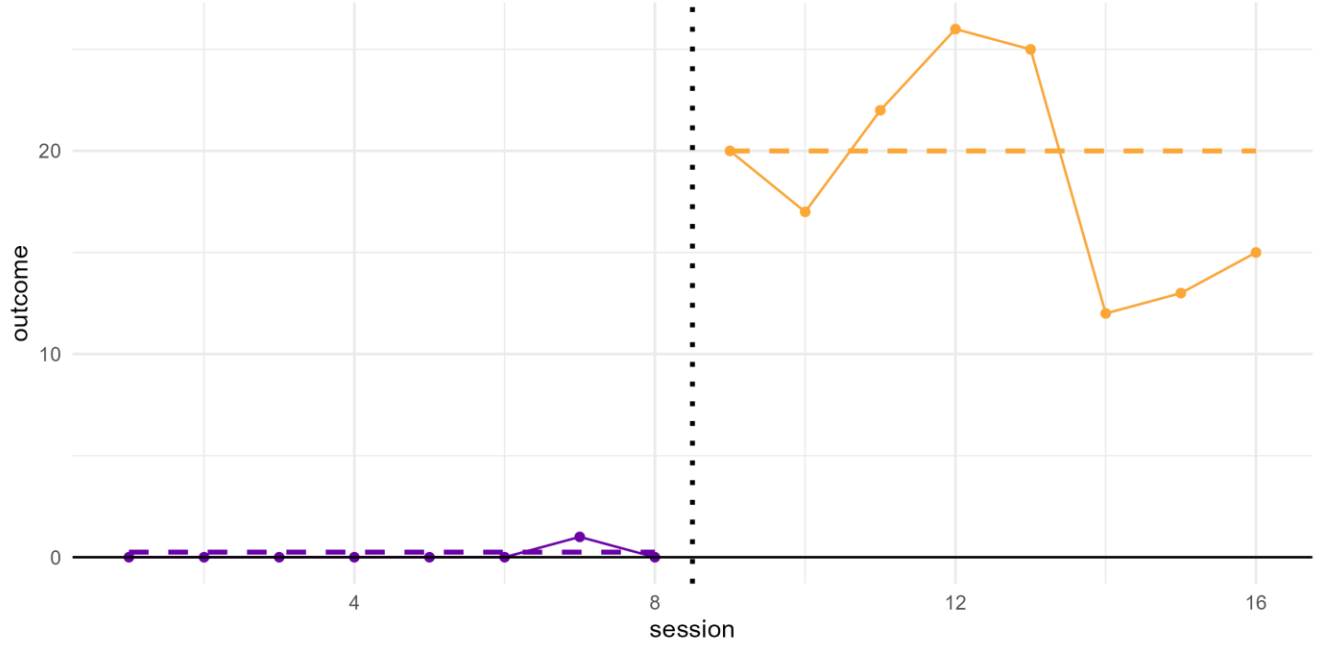
Proportionate change



Proportionate change in levels

- Proportionate change requires dependent variables that are on a ratio scale (i.e., meaningful zero).
  - Frequency count of behavioral incidence
  - Percent occurrence of a behavior
  - Percentage of correct responses

- Does not work well when baseline outcomes are at or near zero.



Distributional overlap / non-overlap

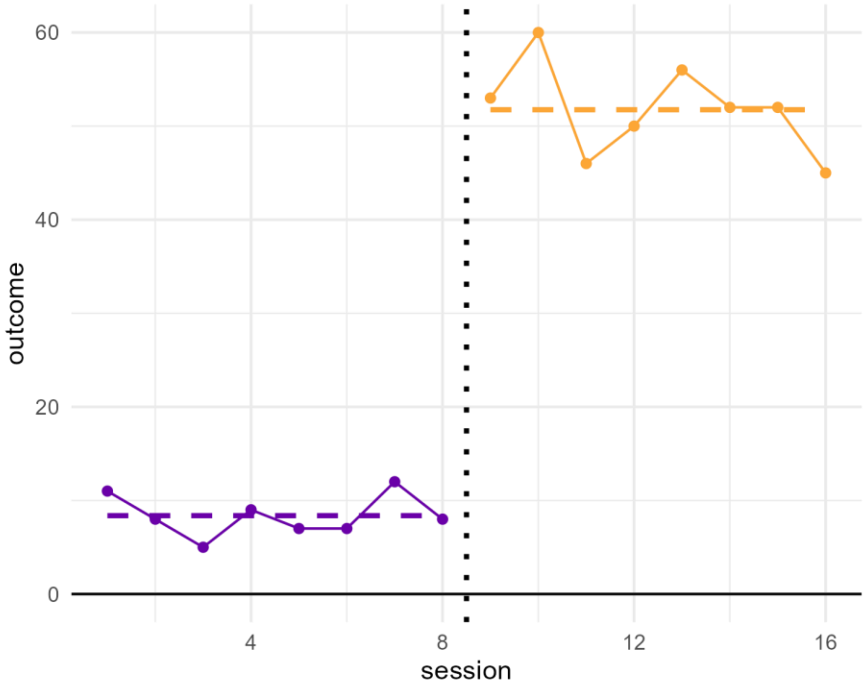
- Defined in terms of ***ordinal comparisons*** of outcomes
  - Meant to be "agnostic" to how dependent variables are measured
  - But still affected by reliability of measurements (Pustejovsky, 2019)

- NAP (Parker & Vannest, 2009) and Tau-AB (Parker, Vannest, Davis, & Sauber, 2011) defined in terms of all pairs of one baseline phase outcome and one intervention phase outcome.

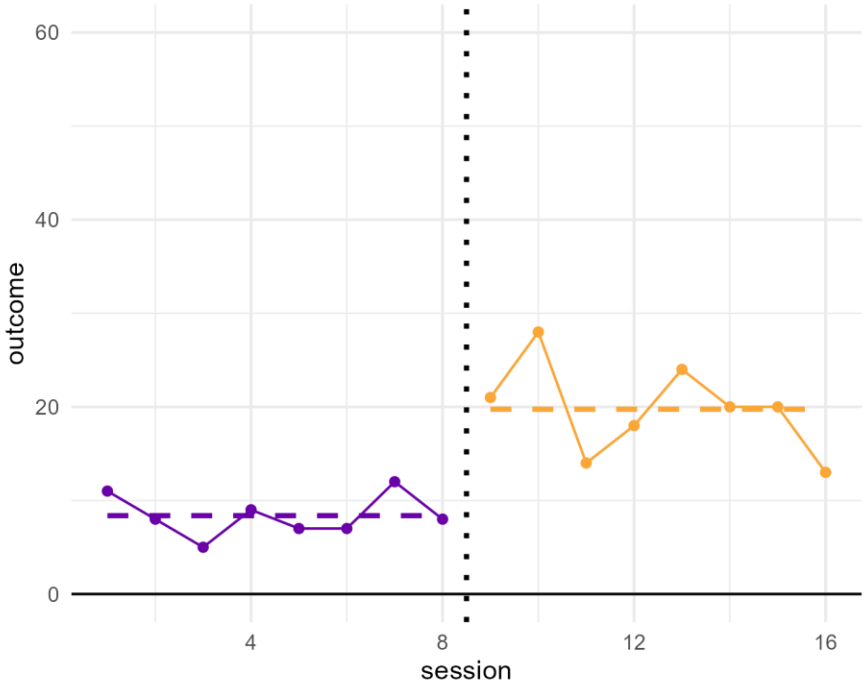
$$\text{Tau-AB} = (\text{Proportion of pairs where B} > \text{A}) - (\text{Proportion of pairs where B} < \text{A})$$

- Limited range of effects where non-overlap measures are sensitive to change.

Limited range of sensitivity



No overlap
Tau-AB = 1.00



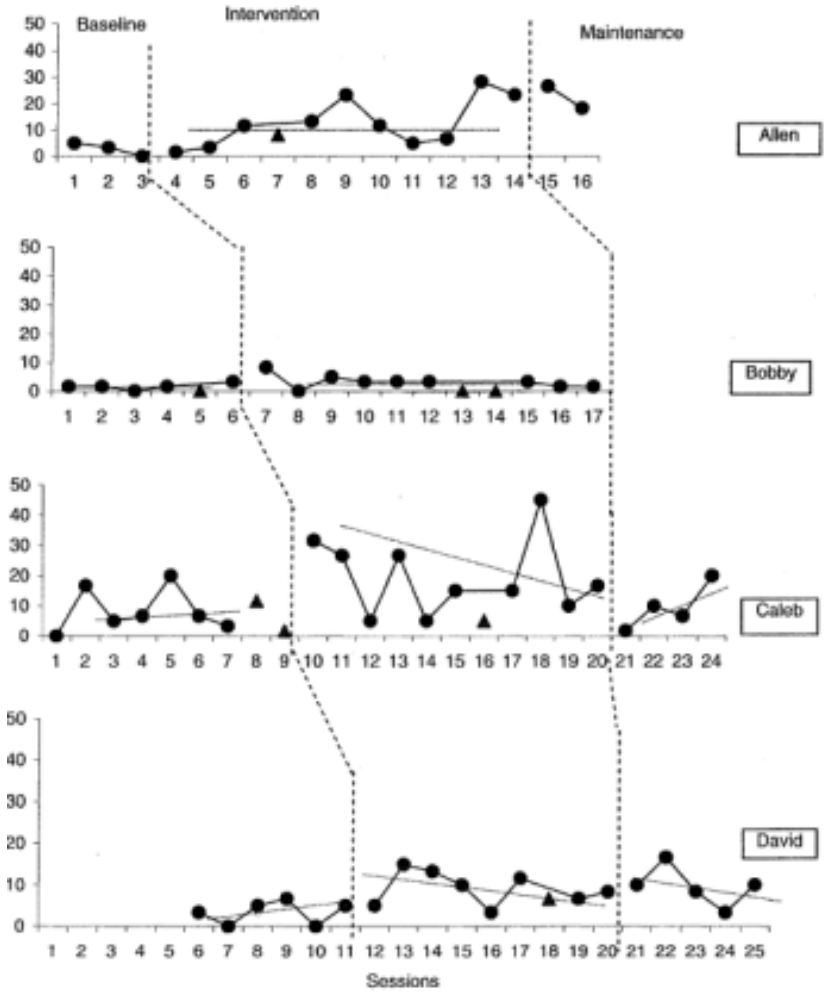
No overlap
Tau-AB = 1.00

Dependent variable measures

- Choose case-level effect size measures based on ***how dependent variables are measured*** in your research area.
- In research synthesis projects, this might mean using ***multiple*** effect size measures for different types of outcomes.

FORM OF INTERVENTION EFFECTS (FUNCTIONAL RELATIONS)

Form of intervention effects

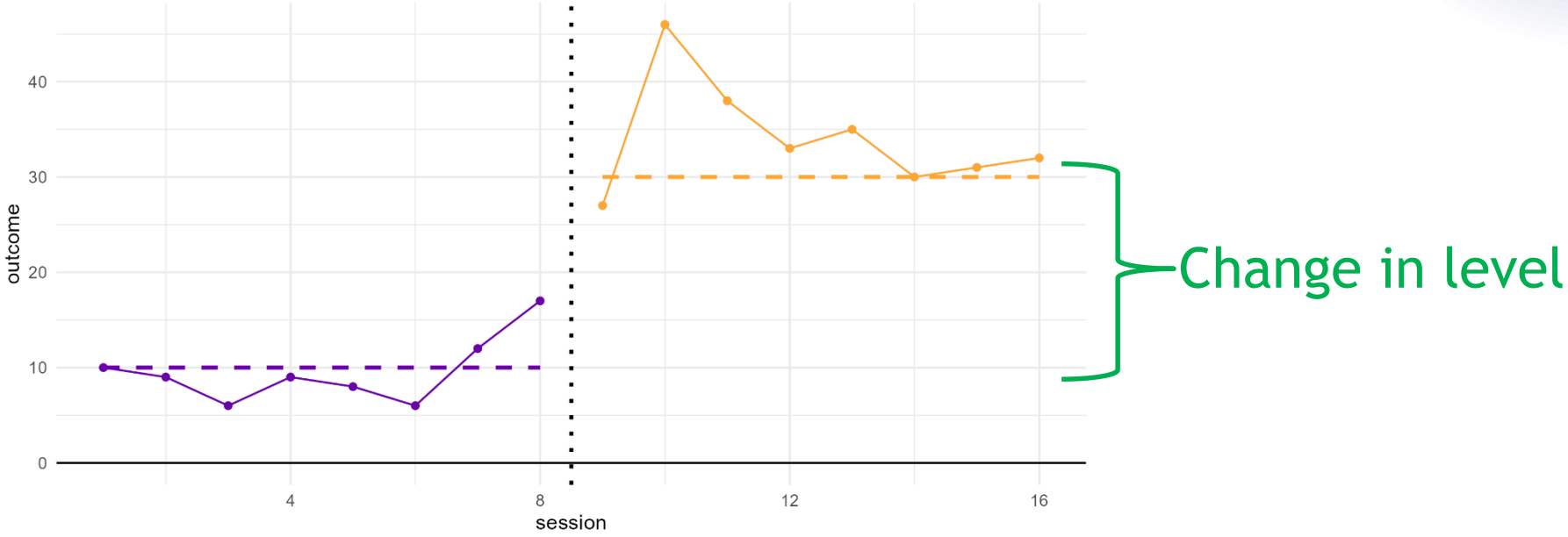


Visual analysis dimensions

- Change in level
- Change in trend
- Change in variability
- Immediacy of change
- Degree of overlap
- Consistency across phases

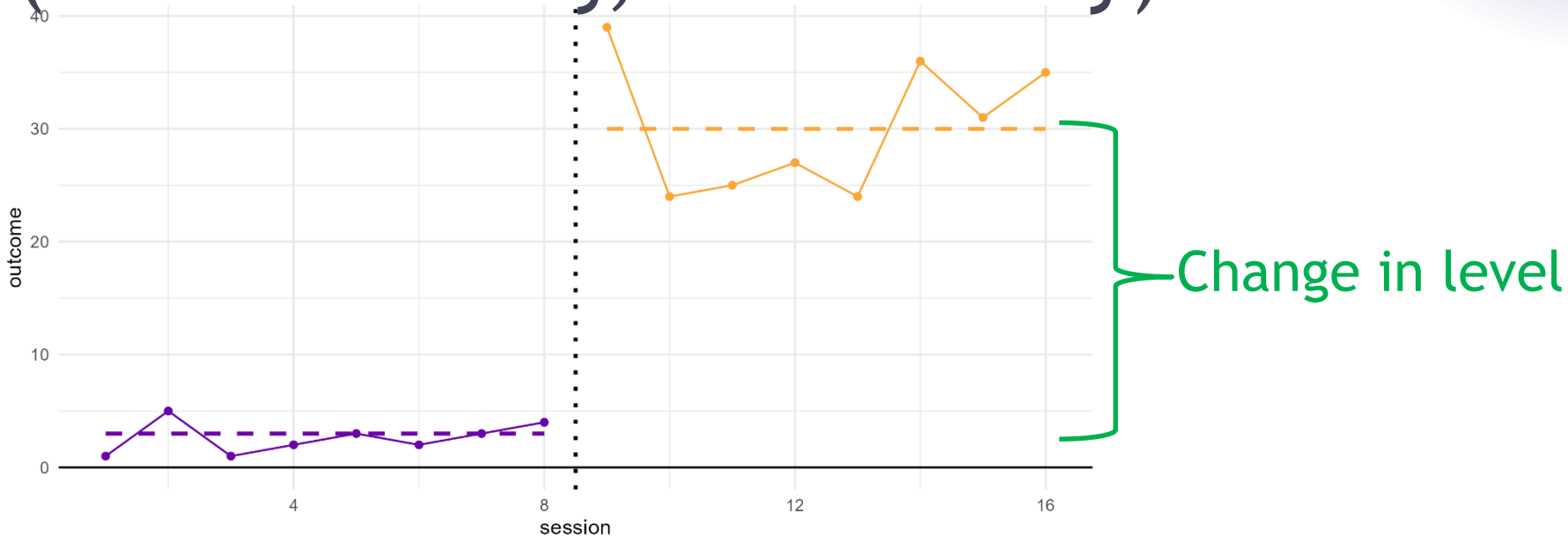
Integrative judgement about presence of functional relation.

Change in level



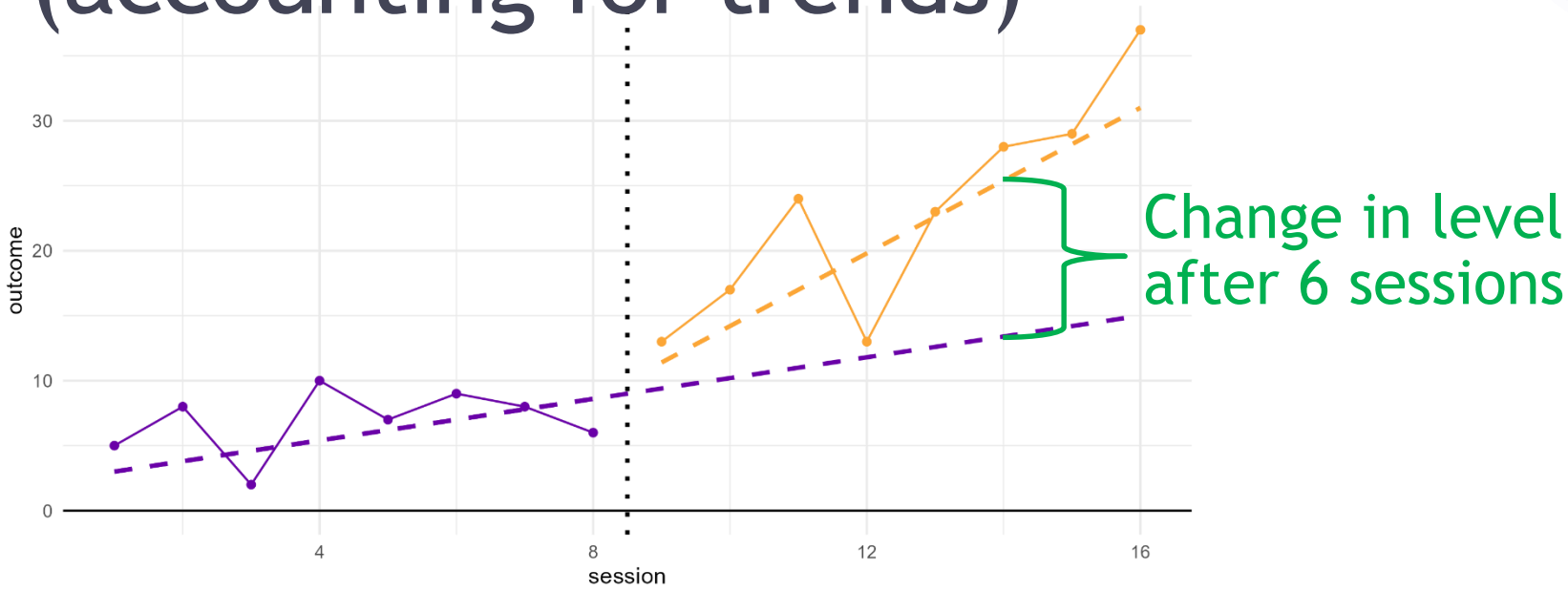
- Change in average level
  - Non-overlap measures (NAP, Tau, PND, PEM)
  - Parametric within-case measures (SMD, LRR, LOR, LRM, PoGO)

Change in level (and variability, incidentally)



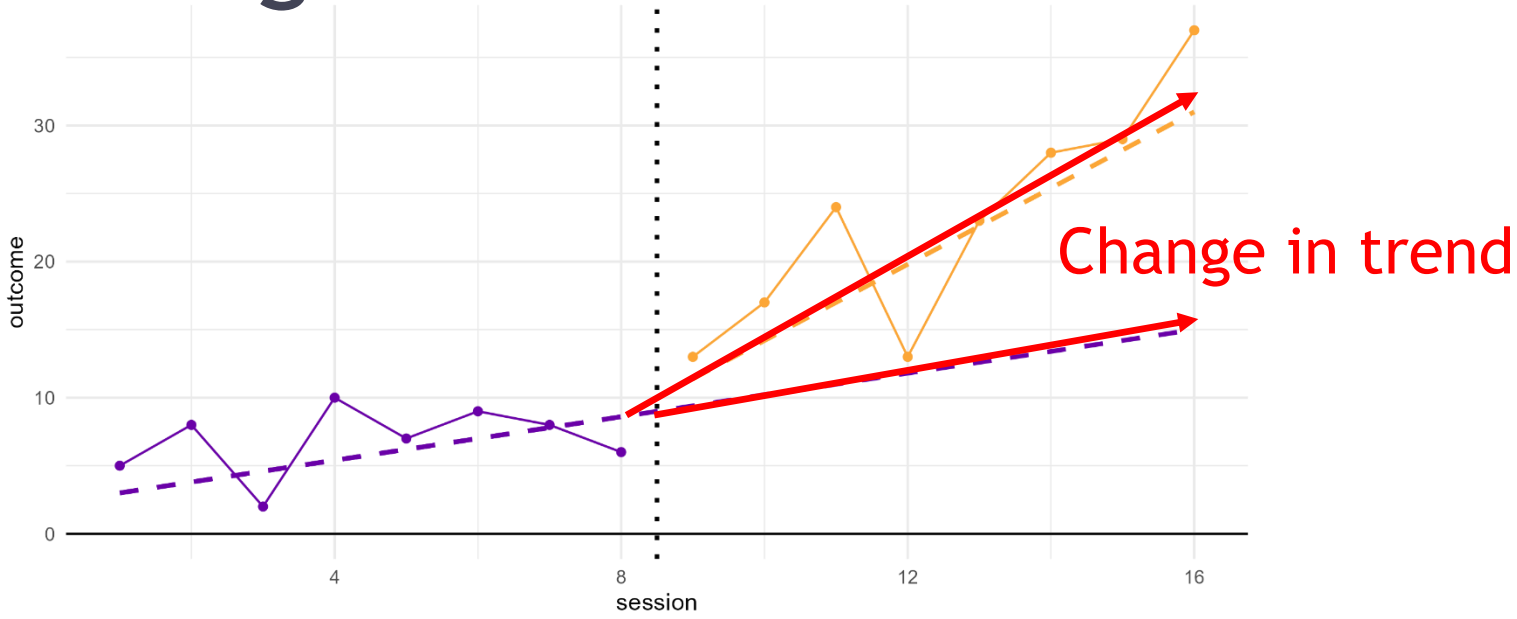
- For some types (distributions) of outcomes, we should expect change in variability to coincide with change in level.

Change in level (accounting for trends)



- Baseline trend adjustment
  - Baseline-corrected Tau (Tarlow, 2017)
- Change in level at focal follow-up time
  - Gradual effects model (Swan & Pustejovsky, 2018)
  - Between-case standardized mean difference

Change in trend



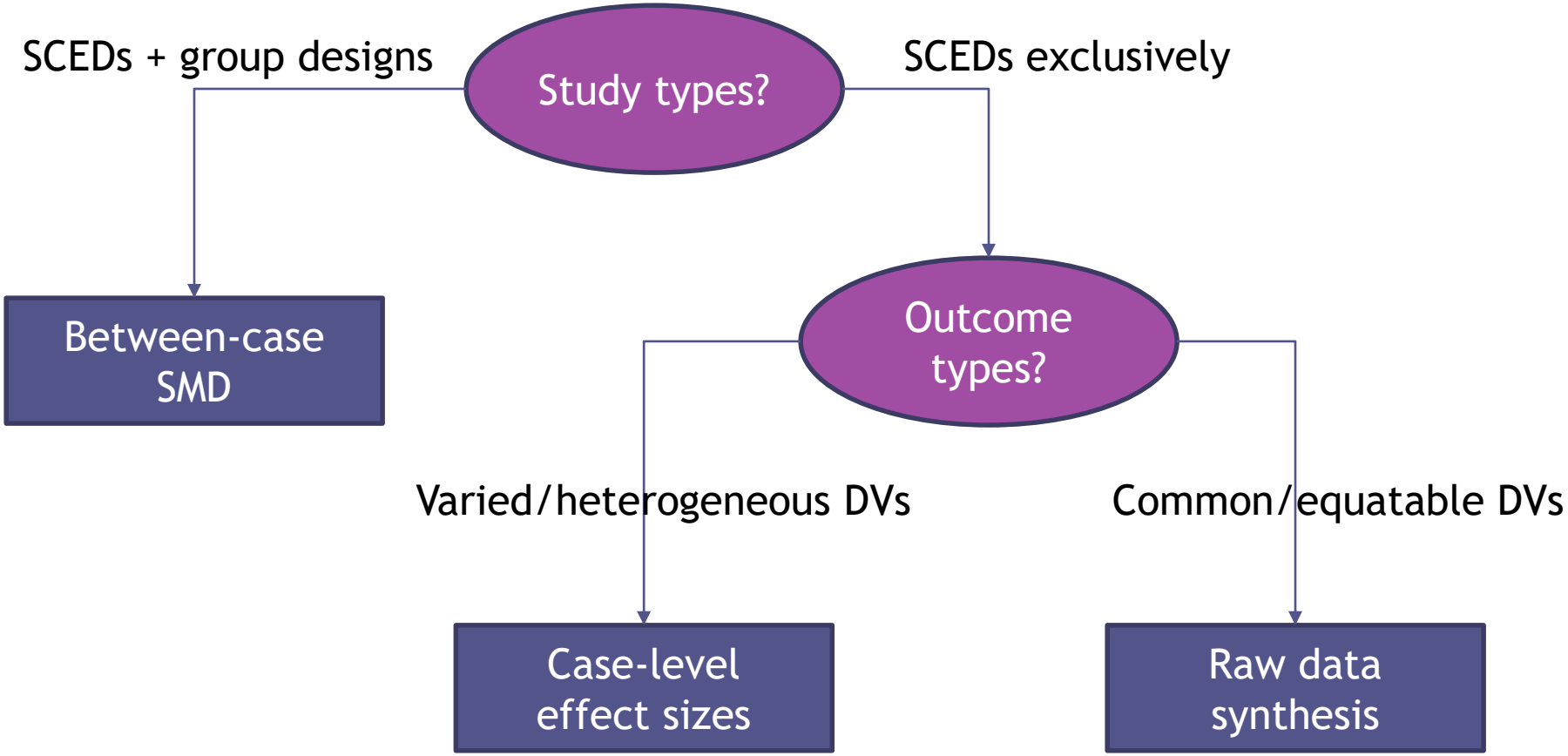
- Can be described by time-by-phase interaction in a regression model (Jamshidi et al., 2020).
- Mostly applied in context of raw data synthesis.

Form of intervention effects

What *forms of intervention effects* are we interested in quantifying numerically?

PRACTICAL AND STATISTICAL CONSIDERATIONS

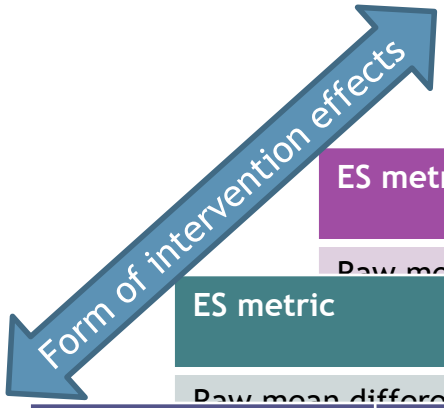
Effect size decision tree



Resources

- Within-case effect size calculator: https://jepusto.shinyapps.io/SCD-effect-sizes/
- Between-case standardized mean difference calculator: https://jepusto.shinyapps.io/scdhlm/
- MultiSCED raw data synthesis tool: http://34.251.13.245/MultiSCED/

A broad space of possibilities



ES metric	Study-level analysis	Case-level analysis	Time-point-level analysis
Raw mean difference		X	X
Standardized mean difference (within)		X	X
Standardized mean difference (between)	X		
Response ratio		X	
Odds ratio		X	
Non-overlap		X	
…			

Statistical Assumptions

- Currently, little recognition of the connection between study procedures and statistical modeling assumptions.
  - How do response-guided design practices affect assumptions (Joo et al., 2018; Swan et al., 2020)?

- Both substantive single-case researchers and methodologists need to work on ***clarifying and scrutinizing our assumptions***.

- Need better tools for ***investigating model fit***, building confidence in statistical summaries of data from single-case research.

Effect Size Measures for Single-Case Research: Conceptual and Statistical Considerations

James E. Pustejovsky
pustejovsky@wisc.edu
https://jepusto.com

April, 2023

Small-is-Beautiful (Once More)
3rd International N=1 Symposium