Symposium:
Issues in and Application of Meta-Analyses and Syntheses of Single-Case Experimental Research in Autism and Developmental Disabilities

Chair: Jeni Ganz
Texas A&M University

Discussant: Kimberly Vannest
Texas A&M University

To cite the symposium: Ganz, J. B. (symposium chair), & Vannest, K. (discussant). (2015, May). *Issues in and application of meta-analyses and syntheses of single-case experimental research in autism and developmental disabilities.* Presenters: Wendt, O., *Current issues in research synthesis and meta-analysis of single-case experiments on autism treatment*; Ledford, J., *The potential effects of interval-based measurement on the estimation of effect sizes: Considerations for single case researchers*; Hong, E. R., Ganz, J. B., Neely, L., Boles, M., Gerow, S., & Ninci, J., *A meta-analytic review of single-case studies on primary caregiver-implemented communication interventions with individuals with ASD;* Machalicek, W. A., Hansen, S., & Raulston, T., *Parent-implemented interventions: Evaluation of utility of Tau-U, Hedges' g, R-IRD, and visual analysis.* San Antonio, TX: Presented at the Association for Behavior Analysis International Convention. [Peer-reviewed].

Current Issues in Research Synthesis of Single-Case Experiments on Autism Treatment

Oliver Wendt
Purdue University

The Potential Effects of Interval-Based Measurement on the Estimation of Effect Sizes

Jennifer Ledford
Vanderbilt University

A Meta-Analytic Review of Single-Case Studies on Primary Caregiver-Implemented Communication Interventions with Individuals with ASD

Ee Rea Hong
University of Tsukuba
Jeni Ganz, Leslie Nelly, Margot Boles, Stephanie Gerow, & Jennifer Ninci
Texas A&M University

Parent-Implemented Interventions: Evaluation of Utility of Tau-U, Hedges' g, R-IRD, and Visual Analysis

Wendy Machalicek, Sarah Hansen, & Tracy Raulston
University of Oregon

ABAI Convention 2015
San Antonio, TX

Current Issues in Research Synthesis and Meta-Analysis of Single-case Experiments on Autism Treatment

Oliver Wendt, PhD

Department of Speech, Language, and Hearing Sciences, and Department of Educational Studies
Purdue University

PURDUE
HEALTH & HUMAN SCIENCES

To cite this paper:

- Wendt, O. (2015, May). *Current issues in research synthesis and meta-analysis of single-case experiments on autism treatment*. In J. B. Ganz (symposium chair) & K. Vannest (discussant), *Issues in and application of meta-analyses and syntheses of single-case experimental research in autism and developmental disabilities.* San Antonio, TX: Presented at the Association for Behavior Analysis International Convention. [Peer-reviewed].

Three Major Issues

1. Critical appraisal:
   - Evaluating quality of SCDs crucial for research synthesis and documenting evidence-based practice
2. Selection of effect size metrics:
   - Controversy regression- vs. non-regression-based measures
3. Mixed methods synthesis:
   - New trend of combining quantitative and qualitative evidence in treatment meta-analysis

ABAI 2015: Current Issues in Research Synthesis

QUALITY APPRAISAL TOOLS

Table 1
Current Quality Appraisal Tools for Single-Subject Experimental Designs (SSEDs)

Characteristics/ Properties*	Tool (maximum quality score)						
	Certainty Framework (no max. score)	Evaluative Method (max. = 12)	EVIDAAC Scales (max. = 10 or 19)	Logan et al. Scale (max. = 14)	SCED Scale (max. = 10)	Smith et al. Scale (max. = 15)	WWC Standards (no max. score)
Composition of tool	Ranks certainty of evidence as "conclusive" (highest), "preponderant", "suggestive", or "inconclusive" (lowest), based on research design, interobserver agreement of dependent variable, and treatment integrity	12-item rating scale divided into primary and secondary indicators; strength of research ranked "strong", "adequate", or "weak" based on number and level of indicators achieved	One treatment scale: 10 items; two or more treatments scale: 19 items; higher score = higher quality	14 questions containing 16 items; studies are rated "strong" (11-14 points), "moderate" (7-10 points), or "weak" (less than 7 points)	11-item rating scale; item 1 assesses clinical history information; items 2-11 allow calculation of quality score; higher score = higher quality	15-item rating scale; higher score = higher quality	*Design Standards* rank internal validity as "Meets Standards", "Meets Standards with Reservations", and "Does not Meet Standards"; *Evidence of Effect Standards* rate effects strength as (1) "Strong Evidence", (2) "Moderate Evidence," or (3) "No Evidence "
Content validity established	No	Yes	No	No	Yes	No	No
Inter-rater reliability provided	No	Yes, including expert and novice raters	No	Yes, including the four authors of the scale	Yes, including expert and novice raters	No	No

Note. EVIDAAC = Evidence in Augmentative and Alternative Communication; SCED = Single-Case Experimental Design; WWC = What Works Clearinghouse. *An extended version of this table containing further details on the various tools is available from the first author upon request.

Application of Appraisal Tools

Small field test to compare seven appraisal tools:

- Four SSED articles on autism treatment
- All major design types:
  - Withdrawal design (Crozier & Tincani, 2005),
  - Changing criterion design (Ganz & Sigafoos, 2005)
  - Multiple baseline design (Ozdemir, 2008)
  - Alternating treatment design (Tincani, 2004)
- First and second author independently applied each appraisal tool to each article
- Calculated inter-rater agreement using percentage agreement yielded an agreement rate of 85%

Table 3

Comparison of Quality Appraisal Tools When Applied to Four Different Types of Single-Subject Experimental Designs

Article (Authors; Year)	Single-Subject Research Design	Quality Appraisal Scores and Rankings based on SSED Appraisal Tools (maximum quality score)*						
		Certainty Framework	Evaluative Method (max.=12)	EVIDAAC Scales (max.=10)	Logan et al. Scale (max.=14)	SCED Scale (max.=10)	Smith et al. Scale (max.=15)	WWC Standards
Crozier & Tincani, 2005	Withdrawal (A-B-A-C)	"Inconclusive"	58% "weak"	80%	71% "moderate"	70%	87%	"Does not meet evidence standards"
Ganz & Sigafoos, 2005	Changing Criterion	"Suggestive"	50% "weak"	90%	54% "moderate"	60%	53%	"Meets standards with reservations": "Moderate evidence"
Ozdemir, 2008	Multiple Baseline Across Participants	"Preponderant"	67% "adequate"	80%	39% "weak"	60%	60%	"Meets standards with reservations": "Strong evidence"
Tincani, 2004	Alternating Treatment	"Preponderant"	67% "adequate"	79% CSSEDARS	61% "moderate"	70%	80%	"Meets standards": "Moderate evidence"

Note. CSSEDARS = Comparative Single-Subject Experimental Design Rating Scale; EVIDAAC = Evidence in Augmentative and Alternative Communication; IOA = Interobserver agreement; IV = Independent variable; SCED = Single-Case Experimental Design; SSED = Single-Subject Experimental Design; WWC = What Works Clearinghouse. *An extended version of this table containing further appraisal details is available from the first author upon request.

Results and Conclusions

- Different tools yield variable quality appraisals when applied to the same research reports
- Lack of agreement on a “gold standard”
- Keep context, focus, and limitations of the tool in mind
- Four tools appeared more rigorous, yielded more consistent results; need to distinguish different purposes:
  - **The Evaluative Method**: comprehensive systematic reviews to inform both clinical/educational practice and policy.
  - **The Certainty Framework**: For time-efficient literature reviews such as rapid evidence reviews (United Kingdom Civil Service, 2011) or critically appraised topics (Wendt, 2006).
  - **The WWC Standards**: thorough assessment of internal validity.
  - **The EVIDAAC Scales**: useful when considerable proportion is comparative treatment designs. The user-friendliness of the scale—that is, an easily accessible format and clear instructions how to use the instrument—also make it an option for the less experienced reviewer.

References

Wendt, O., & Miller, B. (2012). Quality appraisal of single subject experimental designs: An overview and comparison of different appraisal tools. *Education and Treatment of Children*, *35*, 235-268.

Heyvaert, M., Wendt, O., Van den Noortgate, W., & Onghena, P. (In-press). Randomization and data-analysis items in tools for reporting and evaluating single-case experimental studies. *The Journal of Special Education.*

ABAI 2015: Current Issues in Research Synthesis

SELECTION OF EFFECT SIZE METRICS

Current Debate

- What "effect size metrics" are most appropriate to measure effect size and synthesize SSED results?
- Regression-based approaches
  - 4-parameter model (Beretvas & Chung, 2008)
  - Generalized least squares regression approach (Maggin et al., 2011)
  - Multilevel Models (Van den Noortgarte & Onghena, 2003a, 2003b, 2008)
  - *d*-statistic (e.g., Hedges, Pustejovsky, & Shadish, 2012)
- Non-regression-based approaches
  - Family of "non-overlap" metrics, e.g.,
  - Improvement Rate Difference (IRD; Parker et al., 2009)
  - Non-overlap of All Pairs (NAP; Parker & Vannest, 2009)
  - Percentage of Non-overlapping Data (PND; Scruggs, et al., 1987)

How Do PND, PNCD, PEM, PAND, PDO, NAP, and IRD Compare?

- All seven effect size metrics were applied to "real data", previous studies used fabricated or convenience data
- Data set taken from systematic review of school-based instructional interventions for students with autism spectrum disorders (Machalicek et al., 2008)
  - *N*=11 studies, 30 participants, various designs, 133 A-B phases extracted
- Outcomes: communication skills (e.g., gestures, natural speech, use of comm. device) ⇨ behavior increase data
- Followed mostly methodology outlined in Parker, Vannest, and Brown (2009), Parker and Vannest (2007)
  - Focused on overlap calculation only, no transformation to group design effect sizes

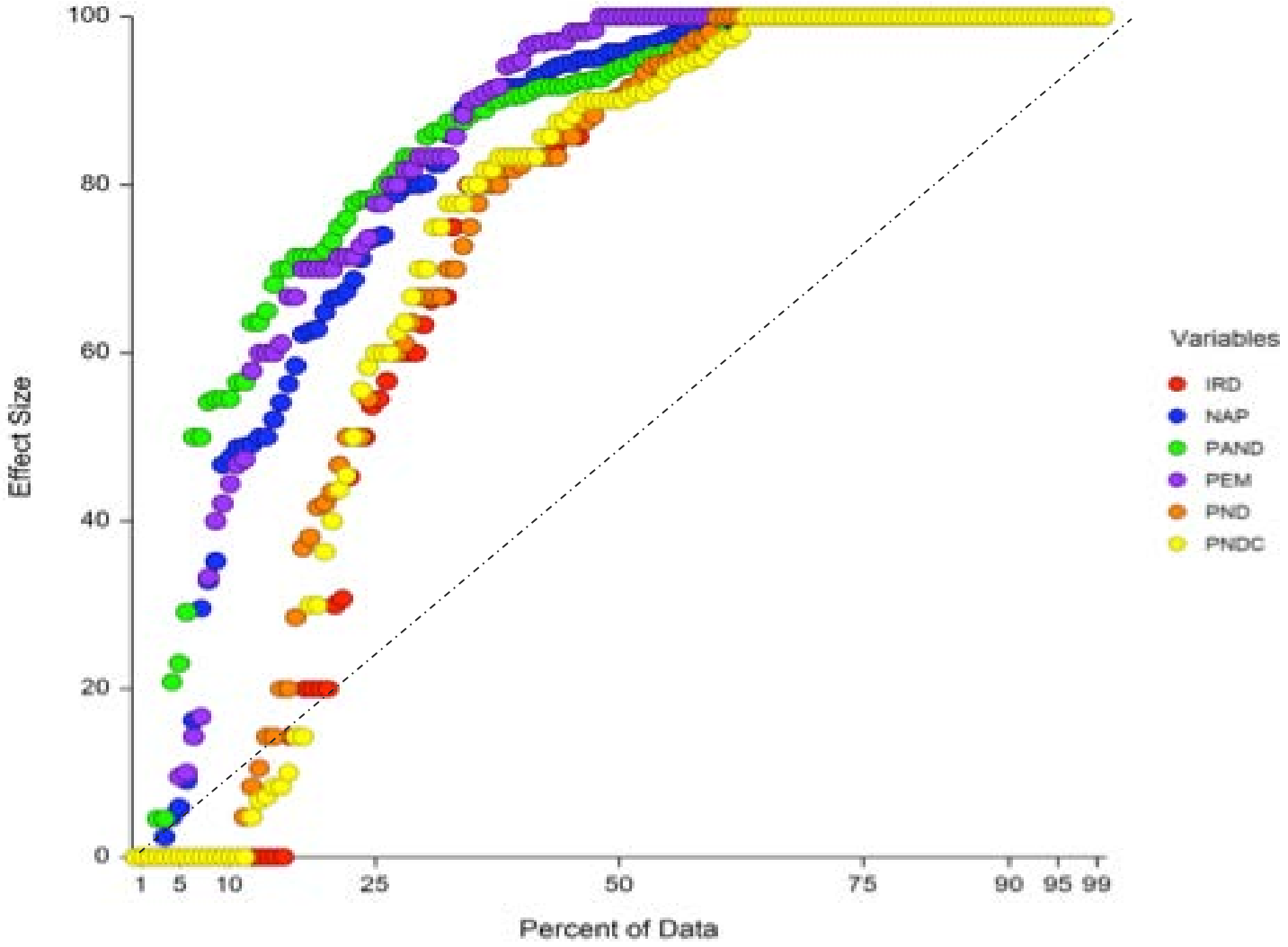
Results

Correlations between non-parametric effect size indices and visual analysis (VA)

	IRD	NAP	PAND	PEM	PDO^2	PND	PND-C	VA
IRD	1	.632*	.831*	.824*	.913*	.914*	.644*	.828*
NAP	.632*	1	.608*	.606*	.610*	.620*	.711*	.549*
PAND	.831*	.608*	1	.858*	.872*	.849*	.607*	.781*
PEM	.824*	.606*	.858*	1	.934*	.845*	.637*	.829*
PDO^2	.913*	.610*	.872*	.943*	1	.941*	.679*	.879*
PND	.914*	.620*	.849*	.845*	.941*	1	.631*	.892*
PND-C	.644*	.711*	.607*	.673*	.679*	.631*	1	.631*
VA	.828*	.549*	.781*	.829*	.879*	.892*	.631*	1

* Correlation is significant at the .01 level (2-tailed)

Discriminability: Uniform Probability Plot



Conclusions/Recommendations

- PND maybe not as bad as originally thought
  - Strong correlation with visual analysis
  - Discriminability better than other metrics but not quite as good as IRD
- IRD looks promising -- superior in discriminability but needs stronger conventions
  - Allows confidence intervals
- NAP is appealing for pairwise comparisons instead of reliance on single data point
  - Discriminability may be an issue -- are the refined procedures of Tau-U the solution? ⇨ needs empirical evaluation

Conclusions/Recommendations (cont.)

- PEM shows by far the weakest performance – confirming previous results (Parker & Vannest, 2007, 2009)
  - PEM leads to inflated ES and does not correlate well with other metrics, use not recommended
- Recommending multiple tier approach
  - Visual analysis – stats test – effect size report

Further Discussion

- Pustejovsky, J.E. (2015). “Effects of Measurement Operations on the Magnitude of Nonoverlap Effect Sizes for Single-Case Experimental Designs”
  - How does measurement system impact effect size magnitude of the non-overlap metrics? Simulation study
  - Factors that varied: recording procedure, interval length, length of observation session, number of observations in baseline and treatment phases ➔ should not impact effect size
  - But: non-overlap metric were sensitive to the length of the observation session and the choice of recording procedure
    - PND or PAND were influenced by length of baseline
  - Results question use of non-overlap metrics when measurement system involves direct observation of behavior

Further Discussion (cont.)

- American Educational Research Association (AERA): Annual poster session on "Single-Case Experimental Designs: Developments in Statistical Analysis, Effect Size Metrics, and Meta-Analysis Methods"
- More applied researchers need to get involved!
  - Collaborate on framework and guidelines for selection and application of suitable data analysis and synthesis methods for SSEDs
  - What are the needs of the applied research community?
  - Showcase meta-analyses and systematic reviews of SSEDs
  - One repository of resources for everyone
- Next meeting April 2016 in Washington, DC

ABAI 2015: Current Issues in Research Synthesis

MIXED METHODS REVIEWS

Complex / Compound Questions

- E.g.: questions about interventions that require mixed methods to answer
  - To what extent and in what ways does the person who delivers the intervention affect the outcomes attained?
  - Who does this intervention work for, and why?
  - What works to achieve outcome x – for whom, in what circumstances, etc.?
- Focus is on the causes of variations in outcome

(Thomas, 2014)

Mixed Methods Reviews

- More complete, concrete, and nuanced answers can be given to complex research questions
- Combination of qualitative and quantitative synthesis approaches holds the possibility to help confirm or refute a theory to a greater degree than either one method can do on its own (Risjord et al. 2002)

Autism Mixed Method Review Example

- Frantzen, Lauritsen, Joergensen, Tanggaard, Aikens, Fetters, & Bjerrum (2014). *Parental Self-perceptions in Autism Spectrum Disorder: Systematic Mixed Method Review and Synthesis of the Literature*
  - Aim was to identify and review suitable self-report measures
  - Parents play a key role integrating and generalizing the ASD treatment strategies into family life
  - Identified most preferable self-report scales, and those less meaningful

Conclusions

- As more and more policy questions will have to be answered, mixed methods reviews will receive more attention
- Review-producing organizations promoting mixed methods reviews
  - Campbell Collaboration (www.campbellcollaboration.org)
  - Cochrane Collaboration (www.cochrane.org)
- Calls for collaboration and team work (single-case researchers pair up with qualitative researchers)

Further Information / Online Workshop

Center on Knowledge Translation for Disability and Rehabilitation Research (KTDRR)

Home | About Us | Training | Technical Assistance | Resources | KTDRR Products | Working Groups

Home » Training » Workshops » Qualitative Res. Synth.

- Home
- Training Overview
- Annual KT Conference
- Workshops
- Webcasts

Qualitative Research Synthesis: KTDRR's Web-based Workshop Series

Research evidence in the field of Disability and Rehabilitation (D&R) research often includes studies that follow a variety of qualitative research paradigms. Such evidence is difficult to summarize using traditional systematic research review procedures. The goal of this series of online webinars is to introduce D&R researchers to the methodology of qualitative evidence reviews. Participants will be provided a state-of-the-art overview on current approaches and will learn to apply those to the literature base.

CRC-CEUs: The Center on KTDRR's 4-part online workshop on Qualitative Research Synthesis has been pre-approved through 2-28-16 by the Commission on Rehabilitation Counselor Certification (CRCC) for 4 CRC-CEUs for anyone who participates in **all four sessions** of the workshop and **completes the required evaluation**. If you would like to request a Verification of Participation form to submit to CRCC, please complete this evaluation: http://www.surveygizmo.com/s3/2083043/Evaluation-Qualitative-Research-Synthesis

Session 1

Introduction to reviewing and synthesizing qualitative evidence
Presenter: Karin Hannes
Date: Feb. 18, 2015

http://ktdrr.org/training/workshops/qual/

Contact Information

Oliver Wendt, Ph.D.
Speech, Language, and Hearing Sciences
LYLE 3160, Purdue University
West Lafayette, IN 47907-2122, USA
Phone: (+1) 765-494-2462
Fax: (+1) 765-496-0125
E-mail: olli@purdue.edu

Questions ???

References

- Beretvas, S. N., & Chung, H., (2008, May). *Computation of regression-based effect size measures.* Paper presented at the annual meeting of the International Campbell Collaboration Colloquium, Vancouver, Canada.
- Braithwaite, K. L., & Richdale, A. L. (2000). Functional communication training to replace challenging behaviors across two behavioral outcomes. *Behavioral Interventions, 15*, 21-36.
- Buffington, D. M., Krantz, P. J., McClannahan, L. E., & Poulson, C. L. (1998). Procedures for teaching appropriate gestural communication skills to children with autism. *Journal of Autism and Developmental Disorders, 28*(6), 535-545.
- Campbell, J. M. (2003). Efficacy of behavioral interventions for reducing problem behavior in persons with autism: A quantitative synthesis of single-subject research. *Research in Developmental Disabilities, 24,* 120-138.
- Campbell, J. M. (2004). Statistical comparison of four effect sizes for single-subject designs. *Behavior Modification, 28*(2), 234-246.
- Campbell, J. M. & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Chicago: RandMcNally.

References (cont.)

- Center, B. A., Skiba, R. J., & Casey, A. (1986). A methodology for the quantitative synthesis of intra-subject design research. *Journal of Special Education, 19*(4), 387-400.
- Chacon, M., & Wendt, O., & Lloyd, L.L. (2006, November). *Using Functional Communication Training to reduce aggressive behavior in autism.* Paper presented at the Annual Convention of the American Speech-Language-Hearing Association (ASHA), Miami, FL.
- Chambers, J., Cleveland, W., Kleiner, B., & Tukey P. (1983). *Graphical methods for data analysis.* Emeryville, CA: Wadsworth.
- Day, H. M., Horner, R. H., & O'Neill, R. E. (1994). Multiple functions of problem behaviors: Assessment and intervention. *Journal of Applied Behavior Analysis, 27*, 279-289.
- Hetzroni, O. E., & Shalem, U. (2005). From logos to orthographic symbols: A multilevel fading computer program for teaching nonverbal children with autism. *Focus on Autism and Other Developmental Disabilities, 20*(4), 201-212.

References (cont.)

- Hetzroni, O., & Tannous, J. (2004). Effects of a computer-based intervention program on the communicative functions of children with autism. *Journal of Autism and Developmental Disorders, 34*(2), 95-113.
- Horner et al. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165-179.
- Horner, R. H., & Day, H. M. (1991). The effects of response efficiency on functionally equivalent competing behaviors. *Journal of Applied Behavior Analysis, 24*, 719-732.
- Huitma, B. E., & McKean, J. W. (2000). Design specification issues in time-series intervention models. *Educational and Psychological Measurement, 60*(1), 38-58.

References (cont.)

- Keen, D., Sigafoos, J., & Woodyatt, G. (2001). Replacing prelinguistic behaviors with functional communication. *Journal of Autism and Developmental Disorders, 31*(4), 385-398.
- Ma, H. (2006). An alternative method for quantitative synthesis of single-subject researchers: Percentage of data points exceeding the median. *Behavior Modification, 30*(5), 598-617.
- Machalicek et al. (2008). A review of school-based instructional interventions for students with autism spectrum disorders. *Research in Autism Spectrum Disorders, 2*, 395-416.
- O'Brien, S., & Repp, A. C. (1990). Reinforcement-based reductive procedures: A review of 20 years of their use with persons with severe or profound retardation. *Journal of the Association for Persons with Severe Handicaps, 15,* 148-159.

References (cont.)

- Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping data (PAND): An alternative to PND. *The Journal of Special Education, 40*(4), 194-204.
- Parker, R.I. & Vannest, K.J. (2009). An improved effect size for single case research: Non-Overlap of All Pairs (NAP), *Behavior Therapy, 40*, 357-367.
- Parker, R.I. & Vannest, K.J. (2007). *Pairwise data overlap for single case research.* Unpublished manuscript.
- Parker, R.I., Vannest, K.J., & Brown (2009). The Improvement Rate Difference for single-case research, *Exceptional Children, 75*, 135-150.

References (cont.)

- Schindler, H. R., & Horner, R. H. (2005). Generalized reduction of problem behavior of young children with autism: Building trans-situational interventions. *American Journal on Mental Retardation, 110*, 36-47.
- Schlosser, R. W. (2003). *The efficacy of augmentative and alternative communication: Toward evidence-based practice.* San Diego, CA: Academic Press.
- Schlosser, R. W., Lee, D.L., & Wendt, O. (2008). Application of the percentage of non-overlapping data (PND) in systematic reviews and meta-analyses: A systematic review of reporting characteristics. *Evidence-based Communication Assessment and Intervention (Special Issue on Meta-analysis of Single-Subject Research), 2*(3), 163-187.

References (cont.)

- Scotti, J. R., Evans, I. M., & Meyer, L. H., & Walker, P. (1991). A meta-analysis of intervention research with problem behavior: Treatment validity and standards of practice. *American Journal on Mental Retardation, 96*(3), 233-256.
- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single subject research methodology: Methodology and validation. *Remedial and Special Education, 8,* 24-33.
- Scruggs, T. E., Mastropieri, M. A., Cook, S. B., & Escobar, C. (1986). Early intervention for children with conduct disorders: A quantitative synthesis of single-subject research. *Behavioral Disorders, 11,* 260-271.
- Sigafoos, J., & Meikle, B. (1996). Functional communication training for the treatment of multiply determined challenging behavior in two boys with autism. *Behavior Modification, 20*(1), 60-84.

References (cont.)

- Van den Noortgarte, W., & Onghena, P. (2003a). Combining single-case experimental data using hierarchical linear models. *School Psychology Quarterly, 18*, 325-346.
- Van den Noortgarte, W., & Onghena, P. (2003b). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments, and Computers, 35*(1), 1-10.
- Van den Noortgate, W., & Onghena, P. (2008). A multi-level meta-analysis of single-subject experimental design studies. *Evidence-based Communication Assessment and Intervention, 2*(3), 142-151.

References (cont.)

- Wacker, D. P., Steege, M. W., Northup, J., Sasso, G., Berg, W., Reimers, T., Cooper, L., Cigrand, K., & Donn, L. (1990). A component analysis of functional communication training across three topographies of severe behavior problems. *Journal of Applied Behavior Analysis, 23,* 417-429.
- Wendt, O., & Lloyd, L.L. (2005). *Evidence-based practice in AAC.* Unpublished PowerPoint Presentation, Purdue University, West Lafayette.
- White, O.R., & Haring, N.G. (1980). *Exceptional teaching.* Columbus, OH: Charles Merrill.
- Wolery, M., Busick, M., Reichow, B., & Barton, E.E. (2010). Comparison of overlap methods for quantitatively synthesizing single subject data. *The Journal of Special Education, 44*(1), 18-28.

The Potential Effects of Interval-Based Measurement on the Estimation of Effect Sizes

Considerations for Single Case Researchers

To cite this paper:

Ledford, J. (2015, May). *The potential effects of interval-based Measurement on the estimation of effect sizes: Considerations for single case researchers.* In J. B. Ganz (symposium chair) & K. Vannest (discussant), *Issues in and application of meta-analyses and syntheses of single-case experimental research in autism and developmental disabilities.* San Antonio, TX: Presented at the Association for Behavior Analysis International Convention. [Peer-reviewed].

VANDERBILT
PEABODY COLLEGE

Interval Based Measurement Systems

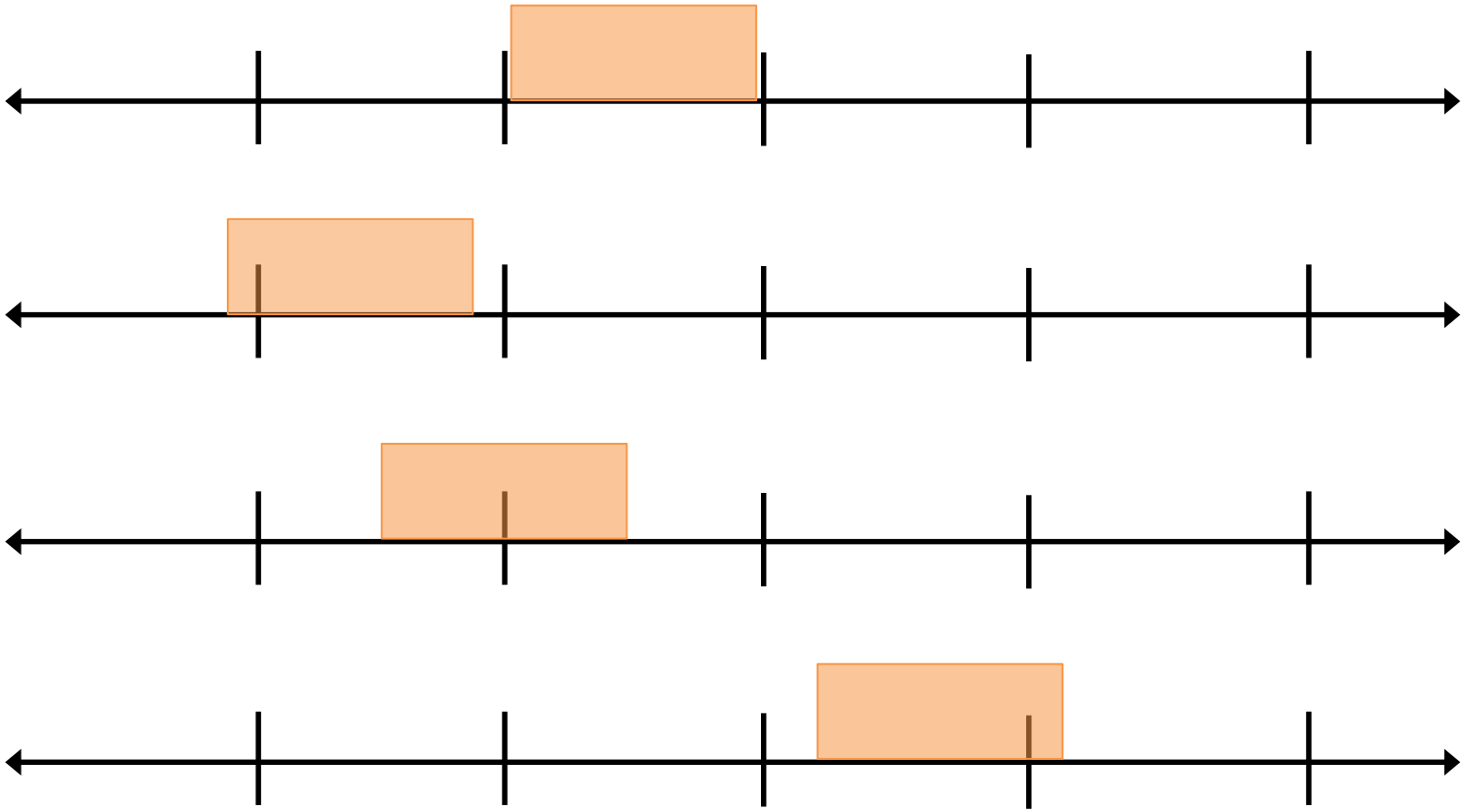
1. Mark an occurrence when the behavior occurs **at any point** during a given time period
2. Mark an occurrence when the behavior occurs **for the entire** time period
3. Mark an occurrence if the behavior **is occurring at the end** of a given time period

Why do people use IBS?

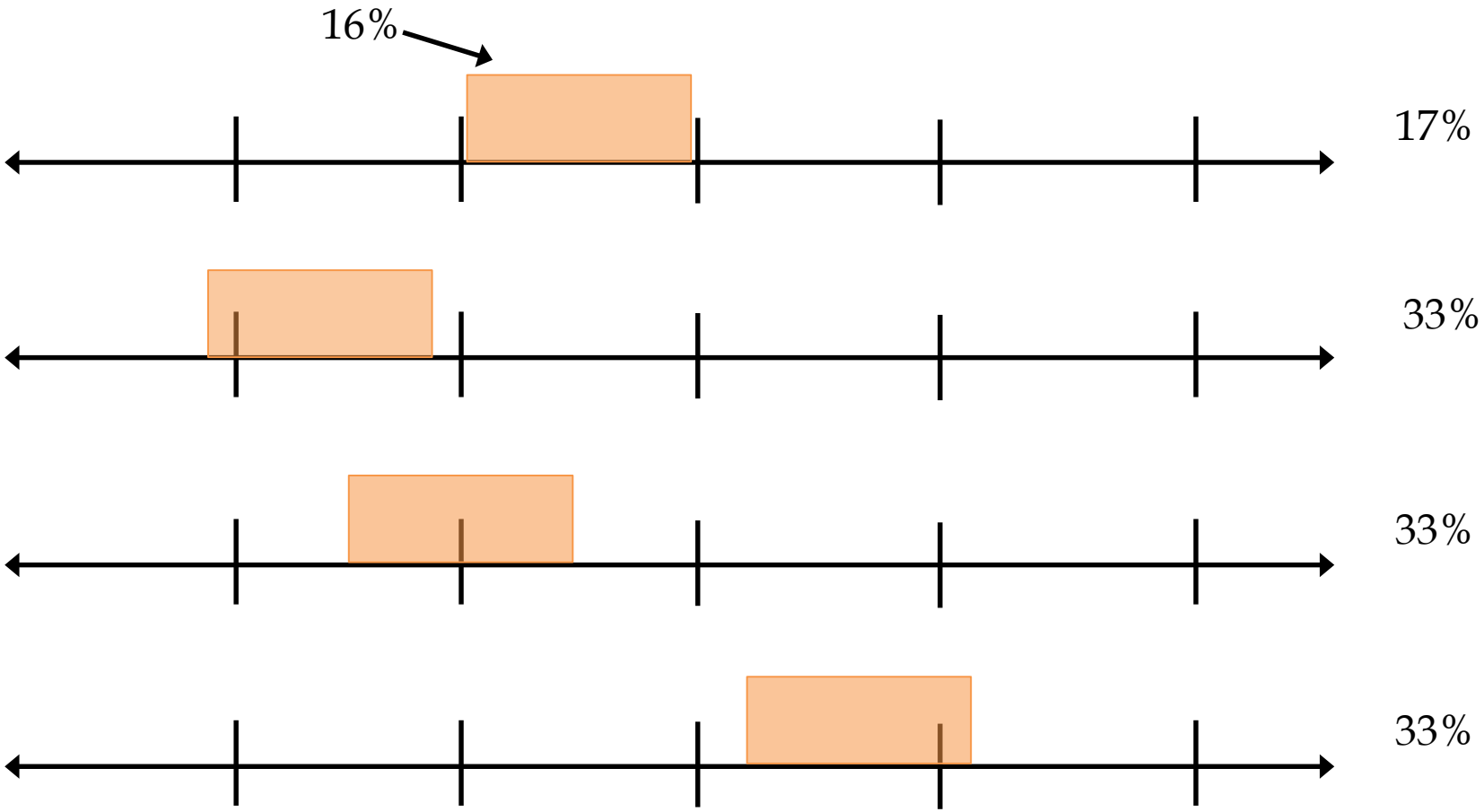
Use of Interval-Based Systems

- Use small intervals
  - Smaller than average DPO
- Often-cited behavior of IBS:
  - PIR overestimates occurrence
  - WIR underestimates occurrence
  - MTS both overestimates and underestimates occurrence

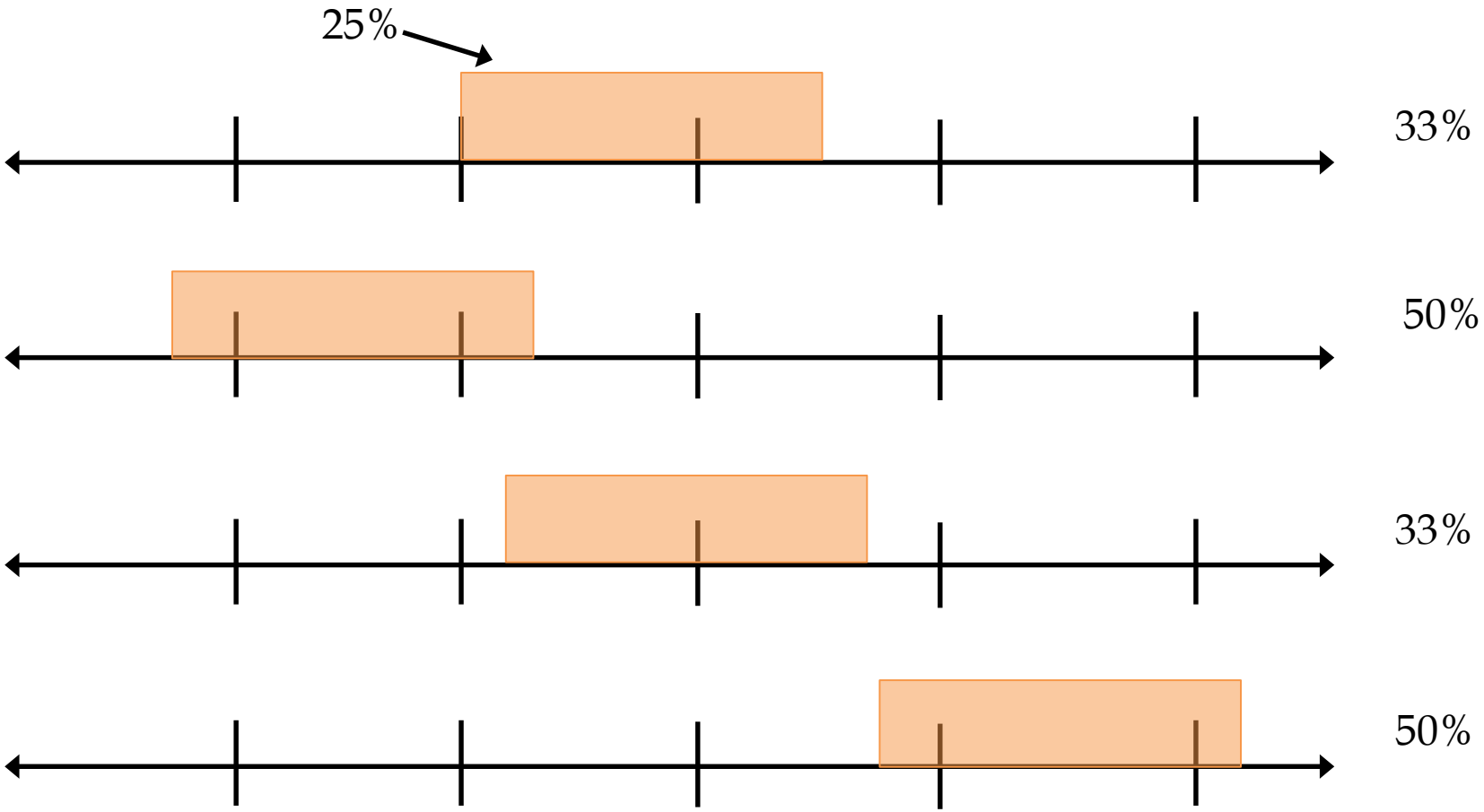
PIR



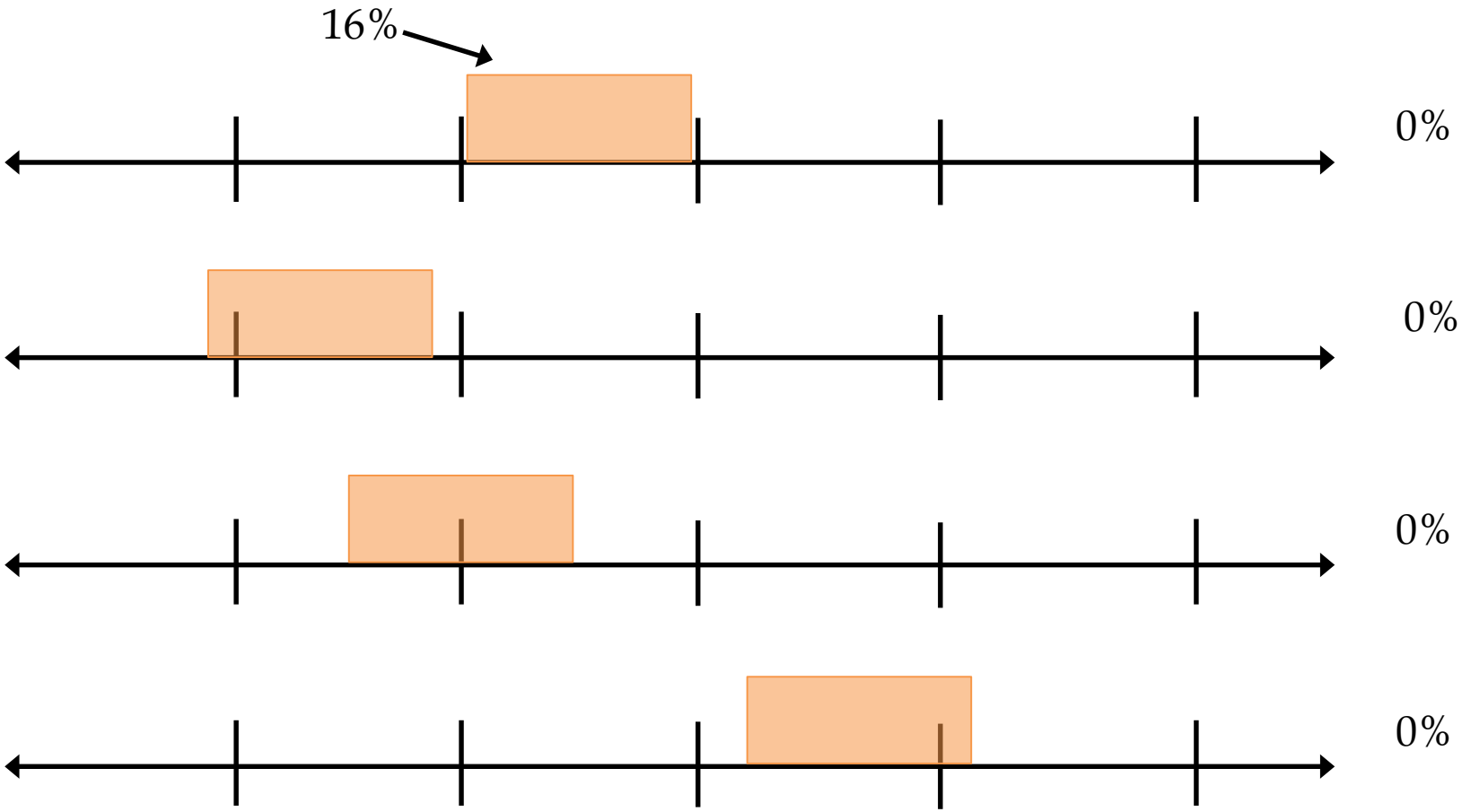
PIR



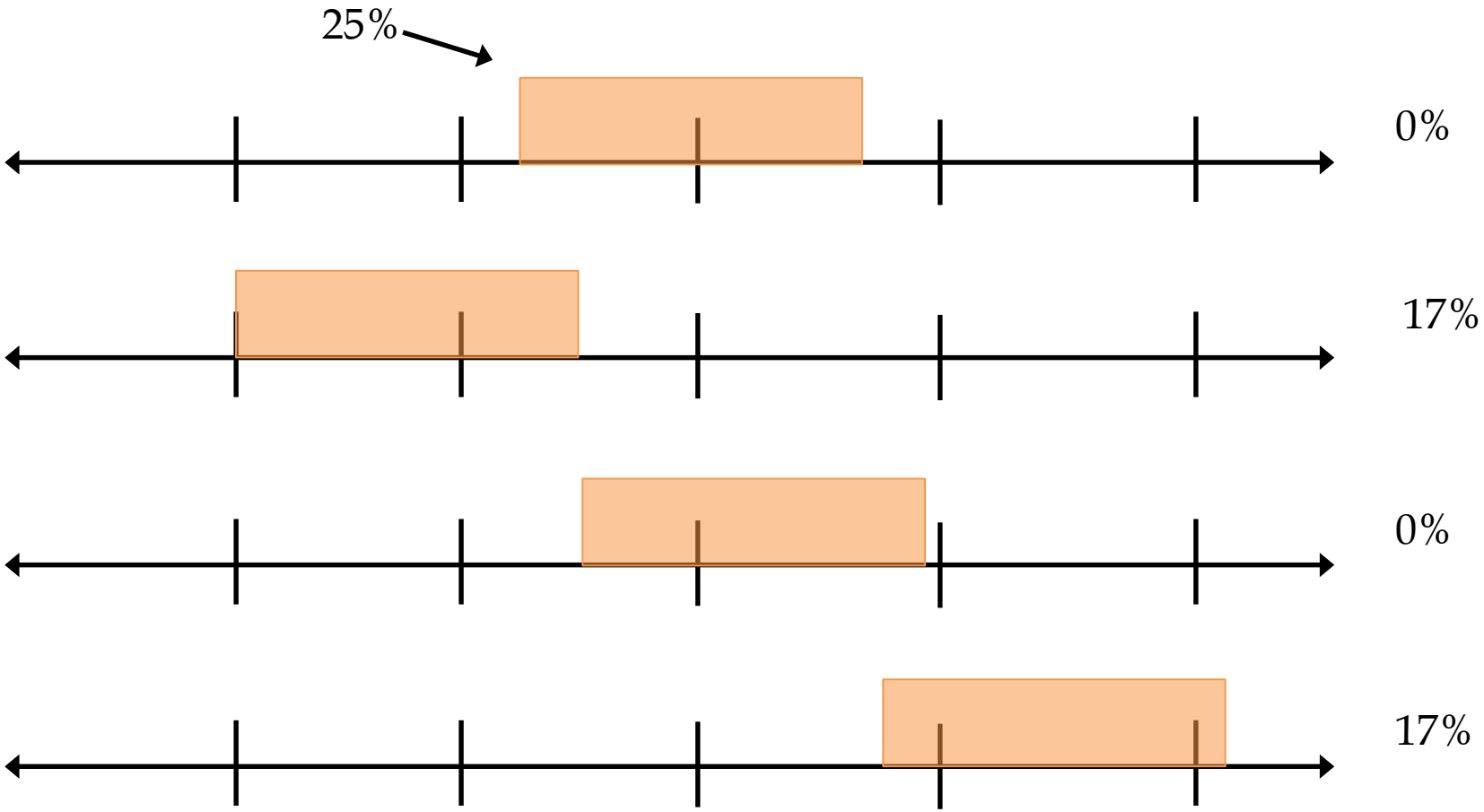
PIR



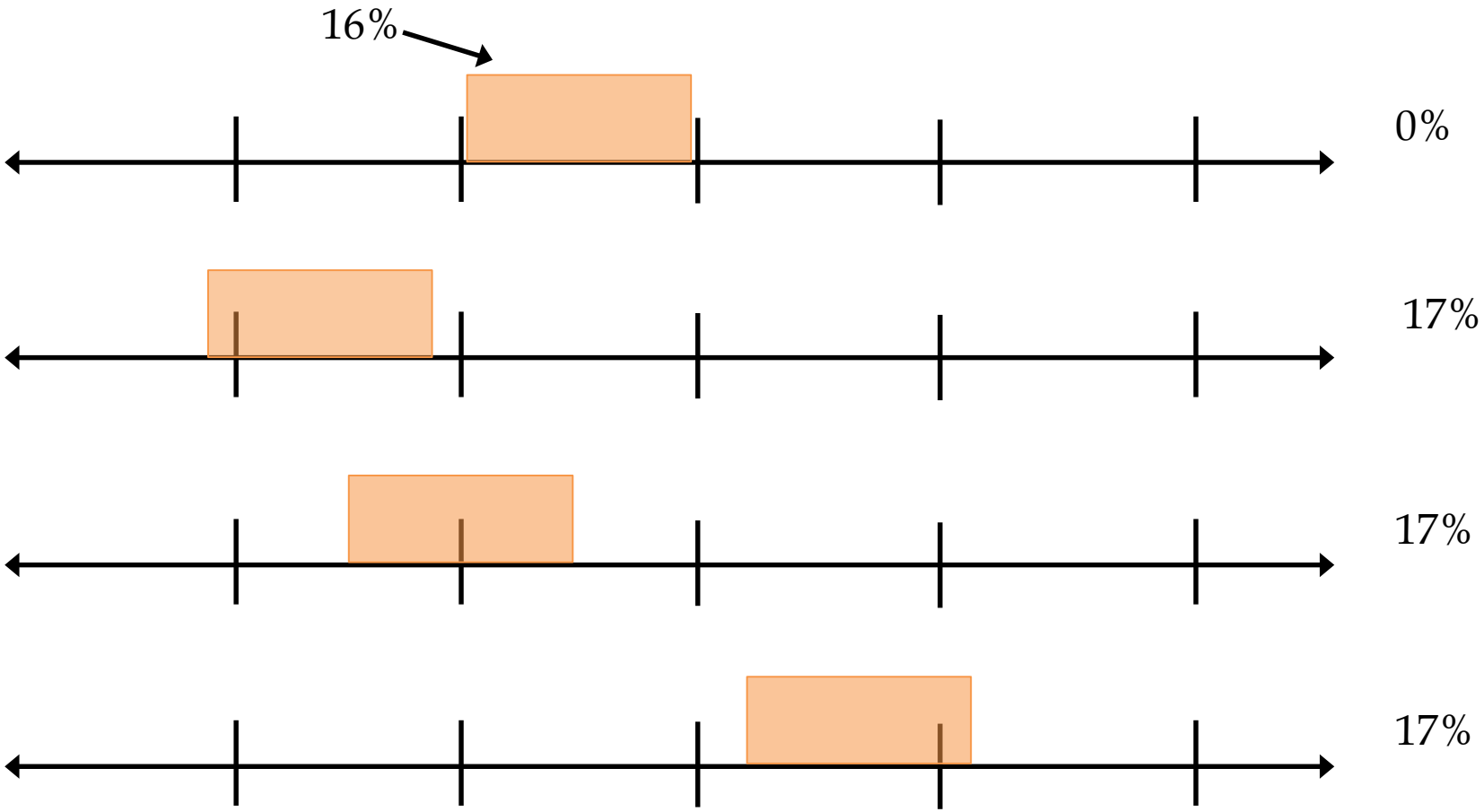
WIR



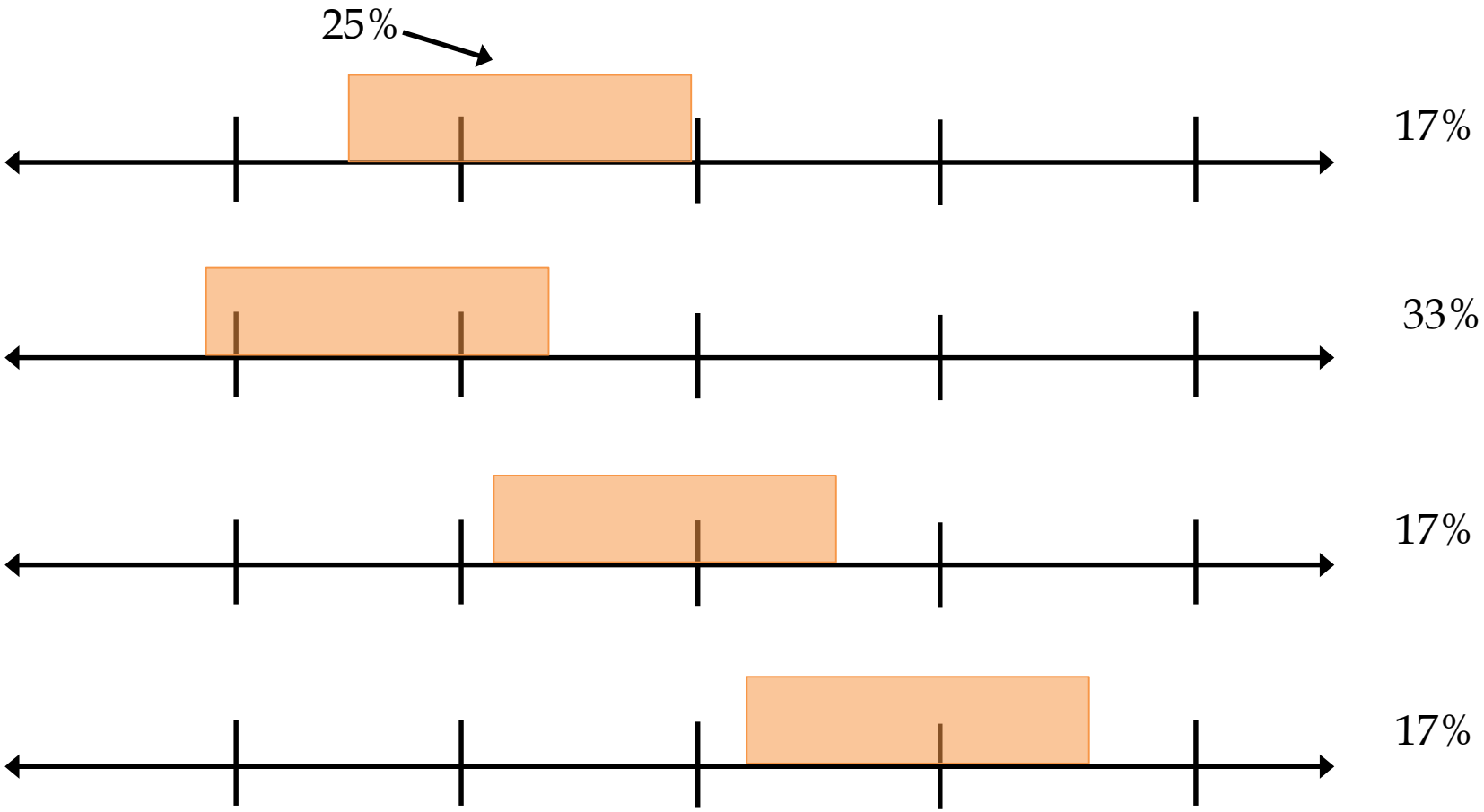
WIR



MTS

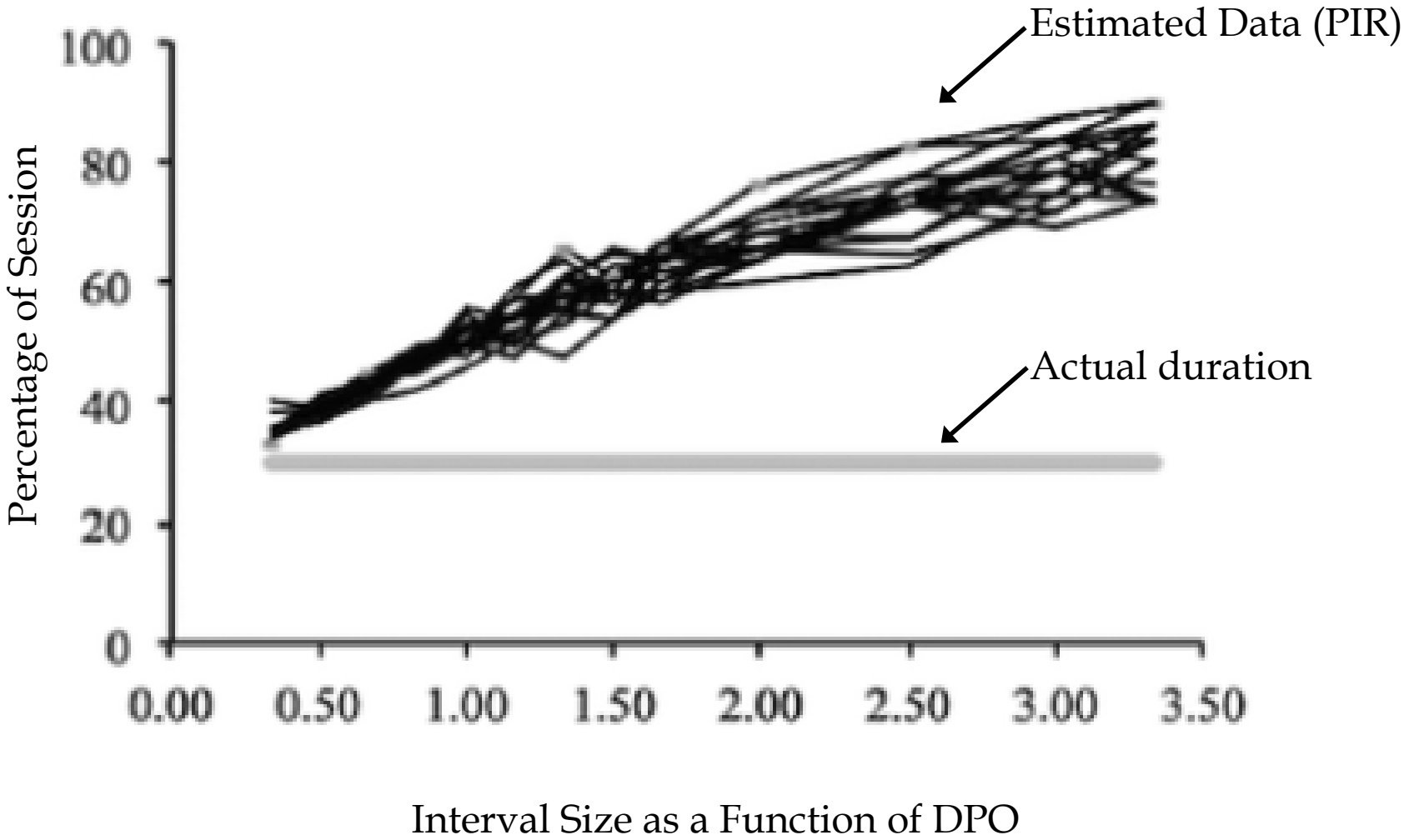


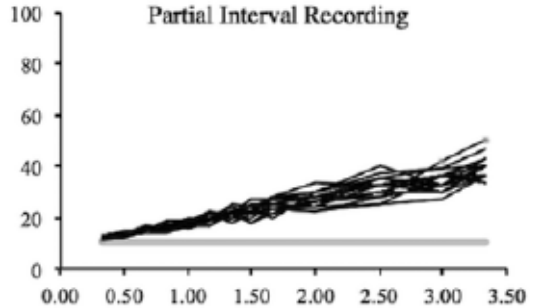
MTS

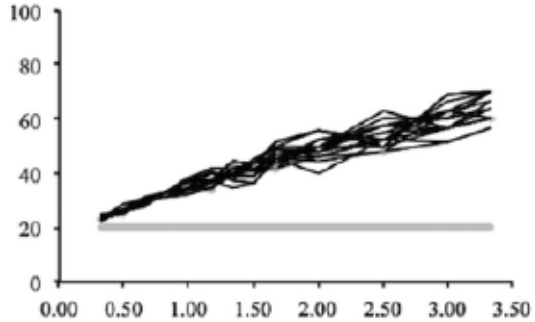


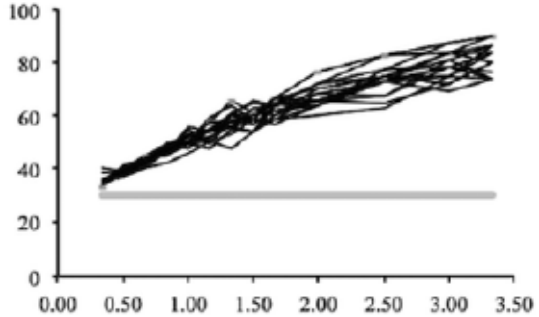
IBS Distortions

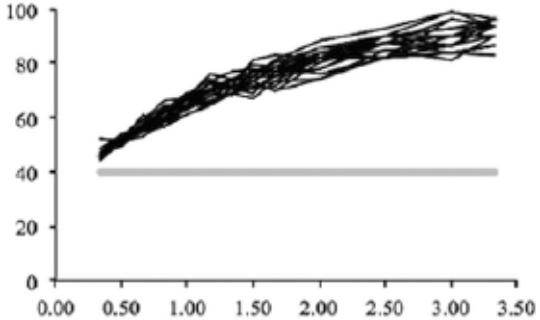
Ledford, Ayres, Lane, & Lam (2015). JoSE

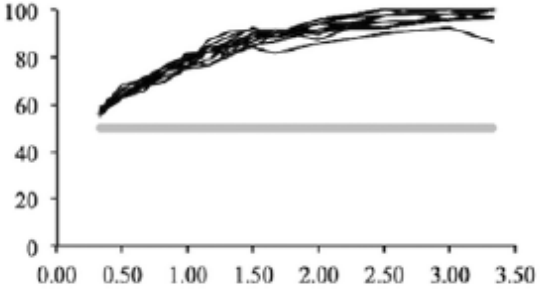


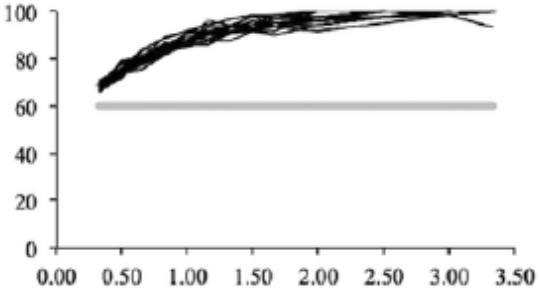


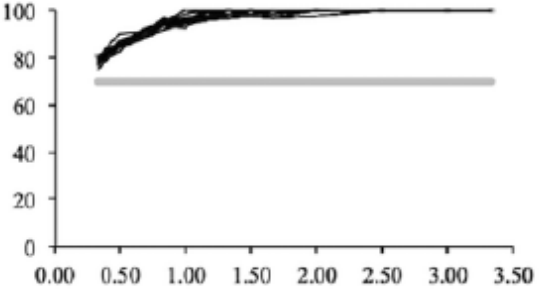








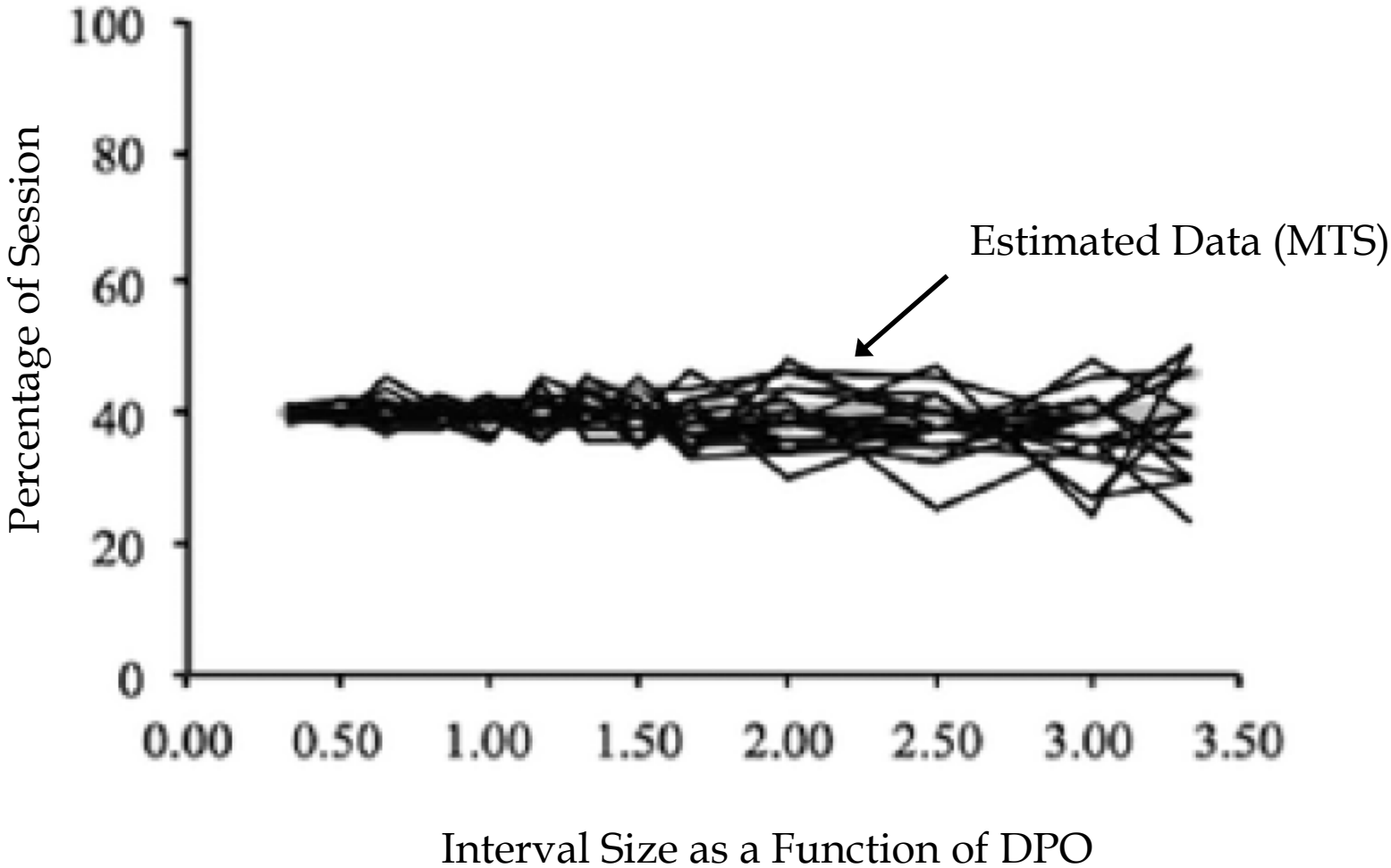




Partial Interval Recording
100
80
60
40
20
0
0.00
0.50
1.00
1.50
2.00
2.50
3.00
3.50

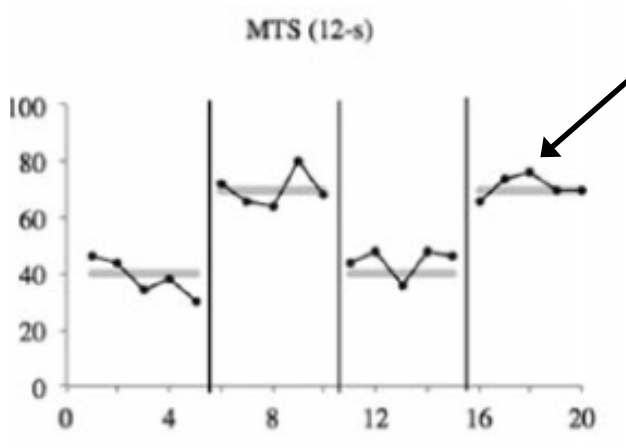
IBS Distortions

Ledford, Ayres, Lane, & Lam (2015). JoSE

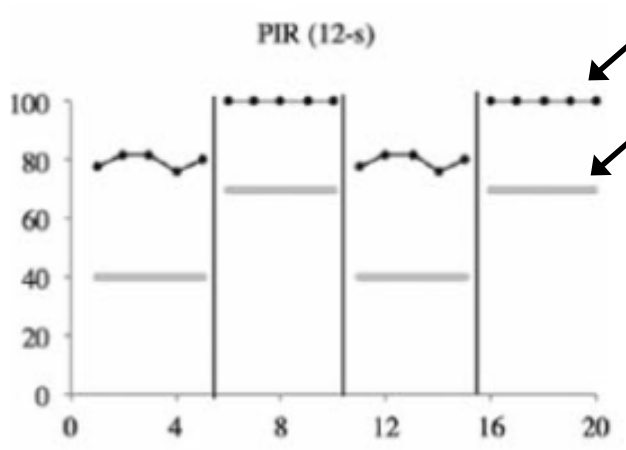


IBS Distortions

Ledford, Ayres, Lane, & Lam (2015). JoSE



Estimated Data (MTS)



Estimated Data (PIR)

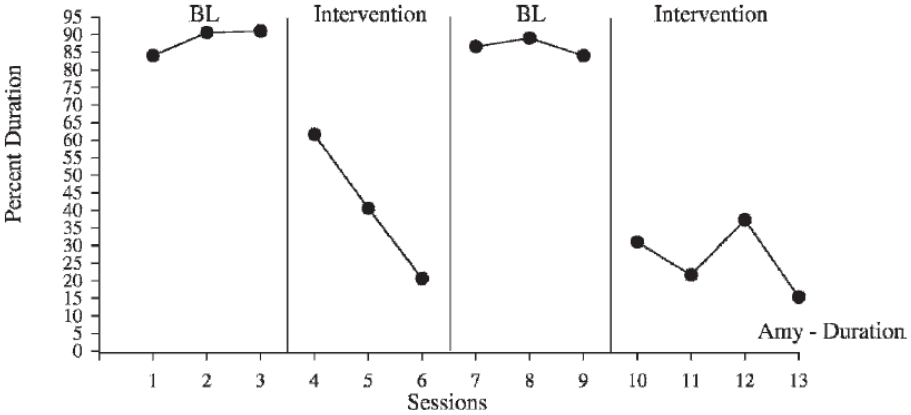
Actual duration

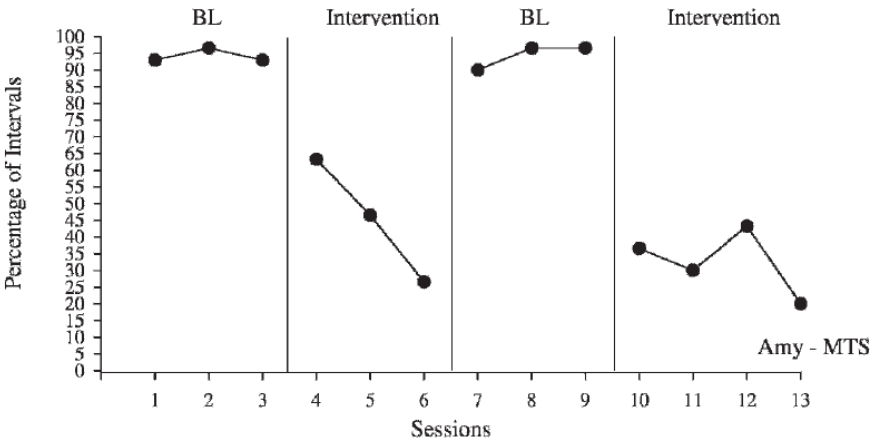
Use of Interval-Based Systems

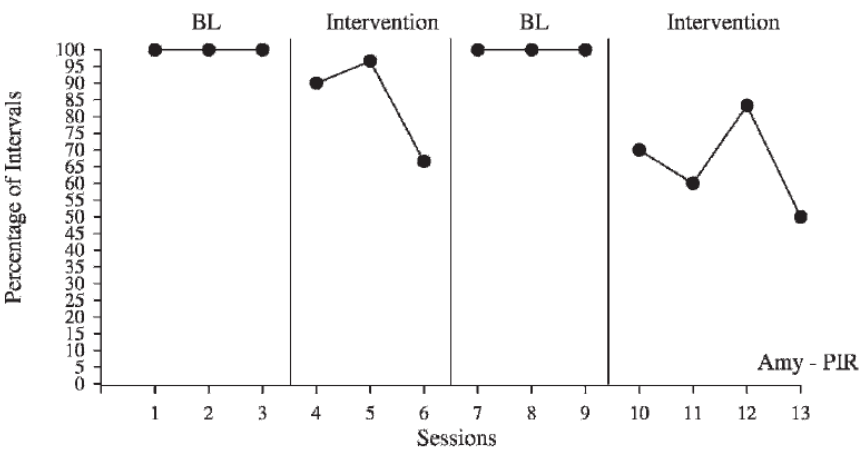
- Previous research has suggested use of IBS has resulted in inaccurate data *that do not* impact conclusions regarding functional relations

What Does this Mean for Effect Sizes?

- From three articles designed to determine differences related to the use of MTS and PIR (with intervals < 1 min) to estimate duration
- Estimated data values (www.plotdigitizer.sourceforge.net)
- Calculated Tau-U values (www.singlecaseresearch.org)
  - Duration
  - PIR
  - MTS







Meany-Daboul, M. G., Roscoe, E. M., Bourret, J. C., & Ahearn, W. H. (2007). A comparison of momentary time sampling and partial-interval recording for evaluating functional relations. *Journal of Applied Behavior Analysis, 40,* 501-514.

A-B-A-B Designs

PIR

↓ 10%

↑ 0%

✓ 90%

	Duration	PIR		MTS	
1	1.00	--	--	1.00	✓
2	1.00	--	--	1.00	✓
3	0.98	1.00	✓	--	--
4	1.00	1.00	✓	1.00	✓
5	0.55	0.45	↓	0.48	↓
6	1.00	1.00	✓	1.00	✓
7	1.00	0.95	✓	1.00	✓
8	1.00	0.99	✓	1.00	✓
9	0.44	0.44	✓	0.49	✓
10	0.98	0.96	✓	0.97	✓
11	1.00	1.00	✓	1.00	✓
12	1.00	1.00	✓	1.00	✓

MTS

↓ 9%

↑ 0%

✓ 91%

A-B-A-B Designs

PIR

↓ 25%

↑ 0%

✓ 75%

	Duration	PIR		MTS	
~~1~~	~~1.00~~	~~--~~	~~✓~~	~~1.00~~	~~✓~~
~~2~~	~~1.00~~	~~--~~	~~✓~~	~~1.00~~	~~✓~~
3	0.98	1.00	✓	--	--
~~4~~	~~1.00~~	~~1.00~~	~~✓~~	~~1.00~~	~~✓~~
5	0.55	0.45	↓	0.48	↓
~~6~~	~~1.00~~	~~1.00~~	~~✓~~	~~1.00~~	~~✓~~
~~7~~	~~1.00~~	~~0.95~~	~~✓~~	~~1.00~~	~~✓~~
~~8~~	~~1.00~~	~~0.99~~	~~✓~~	~~1.00~~	~~✓~~
9	0.44	0.44	✓	0.49	✓
10	0.98	0.96	✓	0.97	✓
~~11~~	~~1.00~~	~~1.00~~	~~✓~~	~~1.00~~	~~✓~~
~~12~~	~~1.00~~	~~1.00~~	~~✓~~	~~1.00~~	~~✓~~

MTS

↓ 33%

↑ 0%

✓ 67%

ATD Designs

PIR

↓ 14%

↑ 27%

✓ 59%

MTS

↓ 27%

↑ 14%

✓ 59%

	Duration	PIR		MTS	
1	0.00	0.50	↑	0.38	↑
2	1.00	1.00	✓	1.00	✓
3	1.00	1.00	✓	1.00	✓
4	1.00	1.00	✓	1.00	✓
5	1.00	1.00	✓	1.00	✓
6	1.00	1.00	✓	1.00	✓
7	1.00	1.00	✓	1.00	✓
8	0.18	0.29	↑	0.21	↓
9	0.46	-0.11	↓	0.32	↓
10	0.63	0.25	↓	0.56	↓
11	0.38	0.21	↓	0.08	↓

	Duration	PIR		MTS	
12	0.36	0.44	↑	0.28	↓
13	1.00	1.00	✓	0.93	↓
14	0.25	0.25	✓	0.25	✓
15	1.00	1.00	✓	1.00	✓
16	1.00	1.00	✓	1.00	✓
17	0.30	0.50	↑	0.50	↑
18	1.00	1.00	✓	1.00	✓
19	0.88	1.00	↑	0.88	✓
20	1.00	1.00	✓	1.00	✓
21	1.00	1.00	✓	1.00	✓
22	0.44	0.50	↑	0.50	↑

ATD Designs

	Duration	PIR		MTS	
1	0.00	0.50	↑	0.38	↑
~~2~~	~~1.00~~	~~1.00~~	~~✓~~	~~1.00~~	~~✓~~
~~3~~	~~1.00~~	~~1.00~~	~~✓~~	~~1.00~~	~~✓~~
~~4~~	~~1.00~~	~~1.00~~	~~✓~~	~~1.00~~	~~✓~~
~~5~~	~~1.00~~	~~1.00~~	~~✓~~	~~1.00~~	~~✓~~
~~6~~	~~1.00~~	~~1.00~~	~~✓~~	~~1.00~~	~~✓~~
~~7~~	~~1.00~~	~~1.00~~	~~✓~~	~~1.00~~	~~✓~~
8	0.18	0.29	↑	0.21	↓
9	0.46	-0.11	↓	0.32	↓
10	0.63	0.25	↓	0.56	↓
11	0.38	0.21	↓	0.08	↓

	Duration	PIR		MTS	
12	0.36	0.44	↑	0.28	↓
~~13~~	~~1.00~~	~~1.00~~	~~✓~~	~~0.93~~	~~↓~~
14	0.25	0.25	✓	0.25	✓
~~15~~	~~1.00~~	~~1.00~~	~~✓~~	~~1.00~~	~~✓~~
~~16~~	~~1.00~~	~~1.00~~	~~✓~~	~~1.00~~	~~✓~~
17	0.30	0.50	↑	0.50	↑
~~18~~	~~1.00~~	~~1.00~~	~~✓~~	~~1.00~~	~~✓~~
19	0.88	1.00	↑	0.88	✓
~~20~~	~~1.00~~	~~1.00~~	~~✓~~	~~1.00~~	~~✓~~
~~21~~	~~1.00~~	~~1.00~~	~~✓~~	~~1.00~~	~~✓~~
22	0.44	0.50	↑	0.50	↑

ATD Designs

PIR

↓ 30%

↑ 60%

✓ 10%

	Duration	PIR		MTS	
1	0.00	0.50	↑	0.38	↑
2	0.18	0.29	↑	0.21	↓
3	0.46	-0.11	↓	0.32	↓
4	0.63	0.25	↓	0.56	↓
5	0.38	0.21	↓	0.08	↓
6	0.36	0.44	↑	0.28	↓
7	0.25	0.25	✓	0.25	✓
8	0.30	0.50	↑	0.50	↑
9	0.88	1.00	↑	0.88	✓
10	0.44	0.50	↑	0.50	↑

MTS

↓ 50%

↑ 30%

✓ 20%

Conclusions

- Use of IBS can substantially change "size of effect"
- More research is needed
- Until then, acknowledgement is needed that synthesizing results when measurement systems vary may be inappropriate

(Current &) Future Directions

- Interesting large scale simulation study also shows overlap-based metrics are influenced by measurement system
  - Session length, number of data points in baseline and/or intervention sessions
- Can we mathematically control for differences?
  - I don't think so—some overestimation and some underestimation occurs
  - Part of problem may be related to sensitivity of overlap metrics to very small changes in a single data point—probably a problem with our notion of "effect size" rather than with the measurement

References

jennifer.ledford@vanderbilt.edu

1. **Devine, S. L., Rapp, J. T., Testa, J. R., Henrickson, M. L., & Schnerch, G. (2011). Detecting changes in simulated events using partial-interval recording and momentary time sampling III: Evaluating sensitivity as a function of session length. *Behavioral Interventions, 26,* 103 – 124.**
2. Ledford, J. R., Ayres, K. A., Lane, J. D., & Lam, M. F. (2015). Accuracy of interval-based measurement systems in single case research. *Journal of Special Education.* Advance online publication. doi: 10.1177/0022466915568975
3. **Meany-Daboul, M. G., Roscoe, E. M., Bourret, J. C., & Ahearn, W. H. (2007). A comparison of momentary time sampling and partial-interval recording for evaluating functional relations. *Journal of Applied Behavior Analysis, 40,* 501-514.**
4. Pustejovsky, J. E. (2015). *Operational sensitivities of non-overlap effect sizes for single case designs.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
5. **Rapp, J. T., Carroll, R. A., Stangeland, L., Swanson, G., & Higgins, W. J. (2011). A comparison of reliability measures for continuous and discontinuous methods: Inflated agreement scores with partial interval recording and momentary time sampling for duration events. *Behavior Modification, 35,* 389 – 402.**
6. **Rapp, J. T., Colby, A. M., Vollmer, T. R., Roane, H. S., Lomas, J., & Britton, L. N. (2007). Interval recording for duration events: A re-evaluation. *Behavioral Interventions, 22,* 319 – 345.**

A Meta-Analytic Review of Single-Case Studies on Family-Implemented Social-Communication Interventions with Individuals with ASD and other DD

Ee Rea Hong, Jennifer B. Ganz, Ph.D., Leslie Neely,
Margot Boles, Stephanie Gerow, Jennifer Ninci

To cite this paper:

Hong, E. R., Ganz, J. B., Neely, L., Boles, M., Gerow, S., & Ninci, J. (2015, May). *A meta-analytic review of single-case studies on primary caregiver-implemented communication interventions with individuals with ASD.* In J. B. Ganz (symposium chair) & K. Vannest (discussant), *Issues in and application of meta-analyses and syntheses of single-case experimental research in autism and developmental disabilities.* San Antonio, TX: Presented at the Association for Behavior Analysis International Convention. [Peer-reviewed].

TEXAS A&M UNIVERSITY

Agenda

I. Overall Introduction and Rationale
II. Research Questions and Purpose of the Study
III. Methodology
    i. Literature search
    ii. Evidence standards
    iii. Moderator coding/Extract raw data/Calculate effect sizes
IV. Results
V. Discussion
VI. References
VII. Self-Disclosure

Introduction & Rationale

Autism Spectrum Disorder (ASD) and Developmental Disabilities (DD):

- Providing early and intensive communication interventions has been emphasized by researchers and educators (Reichow, 2012).
- Most of intervention approaches are teacher- or researcher-mediated interventions and conducted in school or structured settings: A lack of skill generalization (Smith, 2001)
- Delivering early intensive communication interventions by all key communication partners promotes improvement of social-communication skills (Strauss et al., 2012)
- It is critical to involve primary caregivers in the development and implementation of communication interventions for individuals with ASD and other DD (Meadan, Ostrosky, Zaghlawan, & Yu. 2009).

Continued..

Evidence-Based Practice (EBP):

- IDEA 2004 and NCLB 2001
- Use scientifically and empirically validated practices, EBP (Horner et al., 2005).

Single-Case Research (SCR) in Special Education:

- Most commonly implemented type of research design (Horner et al., 2005).

Meta-analytic Review:

- Meta-analytic techniques allow synthesizing and analyzing the data from different single-case design studies and help determine EBP through the use of a single metric applied to all studies (Banda & Therrien, 2008).

Continued..

- None of the prior reviews (e.g., Meadan, Ostrosky, Zaghlawan, & Yu, 2009) investigated how family-implemented interventions differentially affected social-communication skills of individuals with ASD and other DD differentially by the characteristics of those individuals with ASD, types of intervention, and outcome variables.

Purposes and Research Questions

- This study identified overall and specific effect sizes of family-implemented social-communication interventions according to each moderator variable.
- The moderator variables included participant characteristics, type of communication interventions, and social-communication outcome variables.

Research Questions:

1. What are the overall effects of the family-implemented intervention on improving the social-communication skills of individuals with ASD and other DD?
2. Are the effects of a family-implemented social-communication intervention moderated by characteristics of the individual with ASD or DD (i.e., age, communication and language ability)?
3. Which type of interventional approach (i.e., individual with disabilities-led instruction, adult-led didactic instruction) produces the largest improvement?
4. What are the effects of the family-implemented interventions, differentiated by categories of the social-communication outcomes (i.e., social play behaviors, joint attention, verbal or recognizable words, use of AAC system)?

Methodology
Article Identification

Search Procedures:

- Included peer-reviewed and non-peer reviewed papers.
- Unrestricted publication year.
- ERIC, PsychINFO, Academic Search Complete, Professional Development Collection, and Social Science Full Text.
- Keywords: *autis*, ASD, pervasive developmental disorder*, PDD*, Asperger*, development* disab*, low-incidence dis*, intellectual* disab*, mental* retard*,* and *multiple disab** were each combined with the terms, *parent* training, parent education, primary caregiver* training, caregiver* education, sibling training, famil* training, langu*, play*, communic*, langu*, social*,* and *social communic*.*
- Ancestral search

Continued..

Inclusion and Exclusion Criteria:

a) The article must have participants who have been diagnosed as having an ASD or other DD;

b) At least one of family members of those participants' primary caregivers (e.g., parent, other relative, paid in-home caregiver) must have played a role as an intervention implementer;

c) As an outcome measure, social-communication skills must have been targeted;

d) The article must have assessed the efficacy of any type of educational intervention;

e) The article must have conducted an experimental research design including a group design or single-case design, such as AB, alternating treatment, reversal, changing criterion, or multiple-baseline design;

f) In case of a group design, the paper must have reported time-series data for individual participants;

g) The article must have presented data in graphical displays that presented individual data points; and

h) The article was excluded if family members' data and outcome measures were not differentiated from other participants (such as paraprofessionals, teachers, researchers, etc.) or other outcome measures (such as behaviors, academic skills, etc.)

Application of Basic Design Standards

- After the initial screening, articles were reviewed based on basic design standards developed by WWC, outlined by Kratochwill et al. (2013) and adapted by Maggin, Briesch, and Chafouleas (2013).
- In order to meet evidence standards, six design standard indicators should be met.
- An overall score of 0, 1, or 2 was assigned for each design standard based on whether the article did not meet evidence standards, met standards with reservations, or met evidence standards.

Continued..

Design Standards (Kratochill et al., 2013)**:**

- Systematic manipulation of independent variable or intervention.
- Inter-observer agreement (IOA).
- *IOA was collected in each condition and on at least 20% of the data points in each condition.
- IOA averages .80 or higher measured by percentage agreement or at least .60 was measured by *Cohen's kappa* coefficient.
- * At least three attempts of demonstration of an intervention effect at three different points in time or with three different condition changes.
- * At least three data points in each condition.

added an intermediary rating for the current review

Overall Rating:

- Met basic standards: The article met all design six standards (score of 2)
- Met with reservation: Any of the standards were given a score of 1, but none were scored 0 (score of 1).
- Did not meet basic standards: Any of the standards were given a score of 0 (score of 0).

Isolation of Descriptive Information and Potential Moderators Coding

Participant Characteristics:

Ages:

PRES (<5), **ELEM** (5 to < 10), **SEC** (10 to < 15), **ADUL** (15 and older), and **OTHERS** (does not fit any of the categories).

Language and communication level:

NOSP (no speech, but may have had vocalizations); **SPNOTSPON** (some speech, but not spontaneous or functional, echolalia or prompted speech); **SPSOMESPON** (minimal spontaneous speech, large vocabulary, but usually prompted speech); and **OTHERS** (does not fit any of the categories); and **NP** (not provided).

Continued..

Intervention Variables:

IWD (individual with disabilities-led instruction); **ADI** (adult-led didactic instruction); **COMB** (combination of individual with disabilities-led and adult-directed instruction); and **OTHERS** (does not fit any of the categories).

Targeted Communication Skills:

VOC (vocalization, verbalize target words); **NOVOC** (nonverbal communication or gestures using a part of body); **AAC** (use of augmentative and alternative communication systems); **SOC** (social behaviors, joint attention, social play skills, social interpersonal skills), **COMB** (combination of two or more skills), and **OTHERS** (does not fit any of the categories).

Analysis: Tau-U

(Parker, Vannest, Davis, & Sauber, 2011)

- Currently considered among the most appropriate to use in SCR.
- Robust to autocorrelation of data.
- Combine non-overlap between phases with trend from within intervention phases and permits controlling an undesirable baseline trend.
- A “bottom-up” approach:
  - Can be calculated even though there are few data points and phases in the design.
  - Can be customized regarding the design and data.
  - Is in line with visual analysis.
  - Effect sizes can be calculated.
- A result of Tau-U can be summarized either as percent of non-overlap data between phases or percent of non-overlap with either or both phase A and phase B trend controlled.

Continued..

Data Extraction:

- Graphs of each study were saved using the snipping tool provided by Microsoft Windows and saved into an Excel file.
- A rank order for data points in a graph.

Phase Contrasts:

- Only two phases adjacent to each other were contrasted at a time (e.g., A_1 vs B_1 and A_2 vs B_2).
- In a case of a reversal and multiple baseline design, effect sizes of each phase contrast were aggregated.
- If there was more than one intervention phase used in one design (e.g., ABC), each adjacent phase was contrasted separately (e.g., A vs B and A vs C)
- Computed an omnibus effect size (see Parker & Brossart, 2006).

Continued..

Calculation of Effect Size:

- Tau-U software developed through the Maple platform was used to calculate effect sizes (Davis & Davis, 2014).
- The Tau-U effect size was calculated considering the "percent of nonoverlapping data" (as cited in Parker, Vannest, Davis, & Sauber, 2011, p. 6) between baseline and intervention phases.
- Scores ranges from -1.0 to 1.0.
- Tau-U scores were calculated for each participant and across all of the moderators coded.

Inter-rater Reliability: Inclusion/Exclusion Criteria, Design Standards, Moderator Variables, and Raw Data: *Chi-Squared* statistic (Cohen, 1976)

Inclusion/Exclusion Criteria	Kappa
1st criterion	.873
2nd criterion	1.000
3rd criterion	.978
4th criterion	1.000
5th criterion	1.000
6th criterion	.985
7th criterion	.986
8th criterion	.944

Design Standards	Kappa
Overall standard	.850
DS#1: Independent variable	1.000
DS#2A: [a]IOA Collected	.803
DS#2B: IOA 20%	.900
DS#2C: Minimum quality thresholds of IOA	.722
DS#3: Replication effects	.827
DS#4: Number of data points	.712

[a]IOA-interobserver agreement

Moderators	Kappa
Participant age	1.000
Participant communication/language level	.744
Independent variable	.843
Dependent variable	.739

Raw Data	Kappa
Baseline	.869
Intervention	.895

Results: Design Standards

- 368 separate AB contrasts across 40 studies with 156 participants were extracted to calculate effect sizes
  - 5 articles met the design standards.
  - 35 articles met the design standards with reservations.
  - 29 articles did not meet design standards.

Rating	Overall Standards	DS 1: Independent Variable	DS 2A: [a] IOA Collected	DS 2B: IOA 20%	DS 2C: Minimum Quality Thresholds of IOA	DS 3: Replication Effects	DS 4: Number of Data Points
2	**5**	NA	NA	26	NA	47	12
1	**35**	67	66	31	64	5	38
0	**29**	2	3	12	5	17	19

[a] IOA-Interobserver agreement

Results: Number of studies, participants, analyses and Tau results - Age

Age	Number of Studies	Number of Study Participants	Number of Analyses	Group Tau [CI_{95}]
[a]PRESCH	33	103	244	0.659904
[b]ELEM	21	45	113	0.604178
[c]SEC	0	0	0	NA
[d]ADULT	2	2	7	0.593097

Results: Number of studies, participants, analyses and Tau results -Communication/Language Level

	Number of Studies	Number of Study Participants	Number of Analyses	Group Tau [CI_{95}]
[a]SPSOMESPON	11	30	52	0.65077
[b]SPNOTSPON	14	42	100	0.594762
[c]NOSP	9	21	62	0.734505

[a]PRESCH-5<, [b]ELEM-5 to <10, [c]SEC-10 to <15, [d]ADULT-<15

[a]SPSOMESPON- minimal spontaneous speech, large vocabulary, but usually prompted speech, [b]SPNOTSPON- some speech, but not spontaneous or functional, echolalia or prompted speech, [c]NOSP-no speech, but may have had vocalizations

Results: Number of studies, participants, analyses and Tau results – Independent Variables

Independent Variable	Number of Studies	Number of Study Participants	Number of Analyses	Group Tau [CI_{95}]
[a]ADI	23	85	178	0.646182
[b]IWD	14	63	167	0.676521

Results: Number of studies, participants, analyses and Tau results – Dependent Variables

Dependent Variable	Number of Studies	Number of Study Participants	Number of Analyses	Group Tau [CI_{95}]
[a]AAC	6	11	35	0.765988
[b]VOC	18	78	130	0.546551
[c]SOC	11	32	70	0.663016
[d]NOVOC	6	24	36	0.789988

[a]ADI-adult-led didactic instructions, [b]IWD-individual with disabilities-led instructions

[a]AAC-use of augmentative and alternative communication systems, [b]VOC-vocalization, verbalize target words, [c]SOC-social behaviors, joint attention, social play skills, social interpersonal skills, [d]NOVOC-nonverbal communication or gestures using a part of body

Findings

- The first meta-analytic review on single-case research studies that evaluated the overall impacts of family-implemented social-communication interventions and differential impacts across the moderator variables analyzed in this review.
- The first review on this topic that only included single-case research studies that met the basic design standards developed by WWC (Kratochwill et al., 2013).
- Family-implemented social-communication interventions have a moderate effect on improving the social-communication skills among individuals with ASD and other DD.
- No statistically significant differences between the moderator levels.

Implications for Practice and Future Research

For Practice:

- No statistically significant differences between preschool- and elementary-aged individuals with ASD and other DD in terms of the treatment effects
  - The practice of family-implemented interventions can be broadly applied for those aged individuals with ASD and other DD.
- No statistically significant differences between the levels of communication /language skills of individuals with ASD and other DD
  - Family-implemented interventions can be utilized for individuals with ASD and other DD regardless of their level of communication/language skills.

Continued..

For Future Research:

- Slightly modified the basic design standards
  - More studies that have high quality designs should be conducted across the moderator levels.
- No generalization and maintenance conditions were analyzed
  - Evaluate data in generalization and maintenance conditions
  - Plan for collecting generalization and maintenance data more frequently throughout phases.
- No studies were conducted with secondary-aged individuals with ASD and other DD
  - More research on family-implemented social-communication interventions should be conducted with older-aged individuals with ASD and other DD.
- Several moderator levels included only a few studies
  - Conduct more studies with each category within the moderator variables
- Provide specific information regarding the minimum number of or the length of training sessions

References

Buckley, S. D., & Newchok, D. K. (2005). Differential impact of response effort within a response chain on use of mands in a student with autism. *Research in Developmental Disabilities*, *26*, 77-85. doi: 10.1016/j.ridd.2004.07.004

Centers for Disease Control and Prevention (2014). Prevalence of autism spectrum disorders among children aged 8 years–Autism and developmental disabilities monitoring network, 11 sites, United States, 2010.

Charlop-Christy, M. H., & Carpenter, M. H. (2000). Modified incidental teaching sessions: A procedure for parents to increase spontaneous speech in their children with autism. *Journal of Positive Behavior Interventions*, *2*, 98-112. doi: 10.1177/109830070000200203

Crockett, J. L., Fleming, R. K., Doepke, K. J., & Stevens, J. S. (2007). Parent training: Acquisition and generalization of discrete trials teaching skills with parents of children with autism. *Research in Developmental Disabilities*, *28*, 23-36. doi: 10.1016/j.ridd.2005.10.003

Drager, K. D. R., Postal, V. J., Carrolus, L., Castellano, M., Gagliano, C., & Glynn, J. (2006). The effect of aided language modeling on symbol comprehension and production in two preschoolers with autism. *American Journal of Speech-Language Pathology, 15*, 112–125.

Elder, J., Valcante, G., Yarandi, H., White, D., & Elder, T. (2005). Evaluating in-home training for fathers of children with autism using single-subject experimentation and group analysis methods. *Nursing Research, 54,* 22–32. doi: 10.1097/00006199-200501000-00004

Ganz, J. B., Earles-Vollrath, T. L., Mason, R. A., Rispoli, M. J., Heath, A. K., & Parker, R. I. (2011). An aggregate study of single-case research involving aided AAC: Participant characteristics of individuals with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *5*, 1500-1509. doi: 10.1016/j.rasd.2011.02.011

Ganz, J. B., Earles-Vollrath, T. L., Heath, A. K., Parker, R. I., Rispoli, M. J., & Duran, J. B. (2012). A meta-analysis of single case research studies on aided augmentative and alternative communication systems with individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *42*, 60-74. doi: 10.1007/s10803-011-1212-2

Hart, B., & Risley, T. R. (1978). Promoting productive language through incidental teaching. *Education and Urban Society*, *10* (4), 407-29. doi: 10.1177/001312457801000402

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*, 165-179.

Howlin, P., & Charman, T. (2011). Autism spectrum disorders: interventions and outcome. *The SAGE Handbook of Developmental Disorders*, 307. doi: 10.4135/9781446201107.n16

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings*. New York, NY: Oxford University Press.

Koegel, R. L., Symon, J. B., & Koegel, L. K. (2002). Parent education for families of children with autism living in geographically distant areas. *Journal of Positive Behavior Interventions*, *4*, 88-103. doi: 10.1177/109830070200400204

Continued..

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education*, *34*(1), 26-38. doi: 10.1177/0741932512452794

Lang, R., Machalicek, W., Rispoli, M., & Regester, A. (2009). Training parents to implement communication interventions for children with autism spectrum disorders (ASD): A systematic review. *Evidence-based Communication Assessment and Intervention*, *3*(3), 174-190. doi: 10.1080/17489530903338861

Lord, C., Risi, S., & Pickles, A. (2004). *Trajectory of Language Development in Autistic Spectrum Disorders.* Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Maggin, D. M., Briesch, A. M., & Chafouleas, S. M. (2013). An application of the what works clearinghouse standards for evaluating single-subject research synthesis of the self-management literature base. *Remedial and Special Education*, *34*(1), 44-58. doi: 10.1177/0741932511435176

Mayer, G. R., Sulzer-Azaroff, B., & Wallace, M. (2012). *Behavior analysis for lasting change*. Hudson, NY: Sloan Pub.

Meadan, H., Ostrosky, M. M., Zaghlawan, H. Y., & Yu, S. (2009). Promoting the social and communicative behavior of young children with autism spectrum disorders: A review of parent-implemented intervention studies. *Topics in Early Childhood Special Education*, *29*, 90-104. doi: 10.1177/0271121409337950

Minjarez, M. B., Williams, S. E., Mercier, E. M., & Hardan, A. Y. (2011). Pivotal response group treatment program for parents of children with autism. *Journal of Autism and Developmental Disorders*, *41*(1), 92-101. doi: 10.1007/s10803-010-1027-6

National Autism Center. (2008). National Standards Project. Randolf, MA: Author. Retrieved December 4, 2008, from www.nationalautismcenter.org/about/national.php

No child left behind act of 2001, 20 U.S.C. § 6319 (2001).

Parker, R. I., & Brossart, D. F. (2006). Phase contrasts for multiphase single case intervention designs. *School Psychology Quarterly*, *21*, 46. doi: 10.1521/scpq.2006.21.1.46

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy*, *42*(2), 284-299. doi: 10.1016/j.beth.2010.08.006

Reichle, J., Drager, K., & Davis, C. (2002). Using requests for assistance to obtain desired items and to gain release from nonpreferred activities: implications for assessment and intervention. *Education and Treatment of Children*, *25*, 47-66.

Reichle, J., McComas, J., Dahl, N., Solberg, G., Pierce, S., & Smith, D. (2005). Teaching an individual with severe intellectual delay to request assistance conditionally. *Educational Psychology, 25*, 275–286. doi: 10.1080/0144341042000301201

Reichow, B. (2012). Overview of meta-analyses on early intensive behavioral intervention for young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *42*, 512-520. doi: 10.1007/s10803-011-1218-9

Questions?

Thank you!

Contact: Ee Rea Hong

irehong@human.tsukuba.ac.jp

TEXAS A&M UNIVERSITY

Parent implemented interventions: Evaluation of utility of Tau-U, Hedges' g, R-IRD, and visual analysis

- Wendy Machalicek, Sarah Hansen, & Tracy Raulston
- Department of Special Education
- University of Oregon

To cite this paper:

Machalicek, W. A., Hansen, S., & Raulston, T. (2015, May). *Parent-implemented interventions: Evaluation of utility of Tau-U, Hedges' g, R-IRD, and visual analysis.* In J. B. Ganz (symposium chair) & K. Vannest (discussant), *Issues in and application of meta-analyses and syntheses of single-case experimental research in autism and developmental disabilities.* San Antonio, TX: Presented at the Association for Behavior Analysis International Convention. [Peer-reviewed].

Stiebel, D. (1999). Promoting augmentative communication during daily routines: A parent problem-solving intervention. *JPBI*, *1(*3), 159-169.

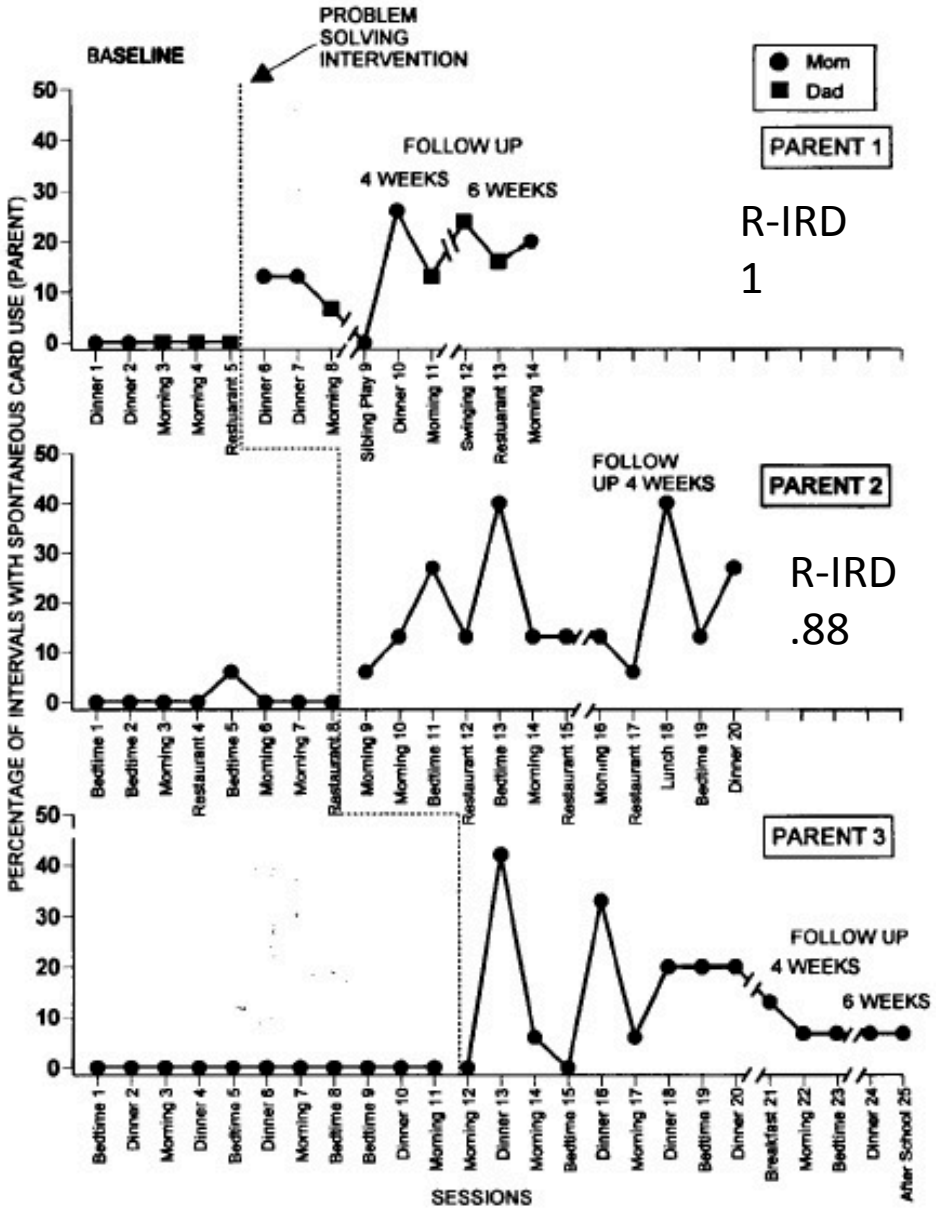


Figure 5. Percentage of opportunities the parents provided their children to use the cards during baseline, intervention, and follow-up.

Visual analysis

R-IRD 1

Tau-U 0.8964

R-IRD .88

Hedges g (across study)

2.19

R-IRD .89

Purpose of current analysis

- Early intensive ABA programs > less intensive EIBI>eclectic treatment>treatment as usual
- Inconclusive if parent implemented programs are as effective as center-directed programs
  - Center-directed > home-based (Smith, Groen et al., 2000)
  - Parent-implemented EIBI = center-directed (Sallows & Graupner, 2005)
  - Center-directed moderate to high effect on intellectual and language improvement; moderate effect on adaptive behavior
  - Amount of parent inclusion shifts benefit of intensive programs from intellectual to adaptive improvements

Strauss, Mancini, the SPC Group, & Fava (2013)

Purpose of current analysis

- Regarding change in parent behavior following parent education, coaching and performance feedback in parent implemented intervention research for children with autism
  - What are typical visual analysis, R-IRD, Tau-U, and Hedges-G values?
  - What is the ability of Tau-U and Hedges-G to discriminate SCR results?
  - How do Tau-U and Hedges-G correlate with one another?
  - What are the relationships between R-IRD, Tau-U, and Hedges-G to traditional visual analysis?

METHODS

Sample selection

- Previously conducted review of 11 high quality systematic reviews of parent implemented interventions for children with intellectual and developmental disabilities (IDD) published between 1997 and 2013 (Machalicek, Raulston, Knowles, Gerow, Hanson, Ruppert, Lang, in preparation)
  - ✓ Focus of the review was interventions with an aim to improve the functioning (i.e., addressed adaptive behavior domains) or quality of life (e.g., development of friendships) of children (ages birth-twelve years)

Additional inclusion criteria

- SCR study
  - Reversal (ABAB), Multiple baseline design, or Multiple-probe design with sufficient data points
- At least 5 data points in each experimental phase
- At least 3 demonstrations of a functional relation
- No inclusion of follow up/maintenance or fading phases

Resultant sample

- From 93 experimental group and SCR design studies, 13 SCR studies fit inclusion criteria
- 95 A-B comparisons published in 13 articles
  - 18 Hedges' g comparisons
- Median length of a full data series was 25 data points, with an interquartile range (IQR; middle 50% of scores) of 20 to 31
  - Phase A had Median=7, IQR= 6 to 12
  - Phase B length had Median=16, IQR= 8 to 24

Meets criteria set by previous studies (e.g., Matyas & Greenwood, 1996)

Journal	# studies included
Journal of Applied Behavior Analysis	3
Research in Developmental Disabilities	3
Journal of Early Intervention	2
Early Education and Development	1
Journal of Emotional and Behavioral Disorders	1
Journal of Speech, Language, Hearing Research	1
Journal of Positive Behavior Interventions	1
Scholarly Inquiry for Nursing Practice: An International Journal	1

Targeted parent behavior	# studies
Naturalistic language intervention (i.e. EMT, imitation training, natural language paradigm)	6
Instructional intervention (i.e. DTT, generalized teaching strategy)	2
Applied behavior analysis (instructional prompts, general teaching strategies)	2
Arrangement of opportunities for AAC use	1
Joint attention intervention	1
Prompting pro-social sibling interactions	1

Visual analysis

- Advanced doctoral students with 4-5 graduate level courses in Single-case Research Methodology who had previously reached reliability on www.singlecase.org (1 is BCBA) conducted visual analysis of each graph and assigned single (average) score
  - using a 6-item rubric with a 7 point scale for each item
  - Developed for this analysis. Based on www.singlecase.org (Swoboda, Kratochwill, Horner, Levin, & Albin, 2012)

Traditional visual analysis of single-case research (SCR) (Horner et al., 2005)

- **Level**
  - The mean of the data within a phase
  - Also can be used to assess the level of the last 3-5 data points within a phase.
- **Trend**
  - The slope of the best-fit straight line describing data within a phase
- **Variability**
  - The level deviation of data around the slope of the best fit straight line (range, standard deviation)
- **Immediacy of Effect**
  - The magnitude of change (in level, trend or variability) between the last 3-5 data points in one phase and the first 3-5 data points in the next phase.
- **Overlap**
  - The percentage of data from one phase (typically the intervention phase) that overlaps with the range of data from the previous phase (typically the baseline phase)
- **Consistency of Data Pattern in Similar Phases**
  - The extent to which phases with similar conditions are associated with data similar data patterns.

Data preparation

- Graphs were manually digitized (point by point) using Un-Scan-It Graph Digitizing Software for Mac OS X
  - http://www.silkscientific.com/graph-digitizer.htm
- Similar software (UnGraph) previously proven to be highly reliable with high confidence that digitized data is nearly identical to original data (Shadish et al., 2009)
  - Digitized data was compared to published graph to identify errors

2 overlap estimates & 1 effect size

- R-IRD
- Tau-U
- Hedges' g

R-IRD (Robust Improvement Rate Difference; Parker, Vannest, Brown, 2009)

- Comes from “Risk Difference” in medical research
- Looks at difference between 2 improvement rates between conditions (generally A-B phases)
- Baseline treated as “control condition”, intervention phase as “treatment condition”

- Allows for calculation of confidence intervals
- Compromised by within phase trends and variability in baseline, number of data points in intervention
- Is not affected by data point at floor or ceiling

Tau-U
Parker, Vannest, Davis, & Sauber, 2011)

- Kendall's Tau + Mann-Whitney U (share same S sampling distribution)
- Integrates non-overlap and trend

- Not affected by ceiling effect and autocorrelation is not an issue

SINGLE CASE RESEARCH™

Home | Calculators | Coming Soon | Idea Center | Papers | SCR Meta-Analysis | Videos | Contact

Tau-U Calculator

?

contrast | chart | clear all | correct baseline

Results

combine | to weighted | remove | clear all | download all

	id	Label	S	PAIRS	TAU	TAUb	VARs	SD	SDtau	Z	P Value	CI 85%	CI 90%
trend:													
	-	-	-	-	-	-	-	-	-	-	-	-	-
phase:													
	-	-	-	-	-	-	-	-	-	-	-	-	-
corrected baseline:													
	-	-	-	-	-	-	-	-	-	-	-	-	-
combined:													
	-	-	-	-	-	-	-	-	-	-	-	-	-

Weighted Average

Label	Tau	Var-Tau	Z	P-Value	CI 85%	CI 90%	CI 95%
-	-	-	-	-	-	-	-

www.singlecaseresearch.org

Hedges' g *d* estimator (Hedges et al., 2012)

- Effect size estimator for $(AB)^k$ designs, also MBD (Hedges et al., 2012)
- Corresponds to standardized mean difference between groups at post-test (Cohen's *d*)
- Takes following into account:
  - Autocorrelation
  - Number cases in each study, data points each phase
  - Ratio of between/total (between + within) variance
  - Corrects for small sample bias
- Still need 3 cases on same outcome, continuous outcomes, absence of trends, fixed treatment effect across cases within studies

Shadish (2014). IES Single-case methods and advanced analysis Institute.

DHPS SPSS Macro for AB^k and MBDs (Marso & Shadish)

- This macro produces Hedges' g (Hedges, 1981), which is comparable to Cohen's d, but also allows for small sample sizes as is typical in SCR.
- The Hedges' g effect size is calculated like a standard Cohen's d effect size, where control means are subtracted from treatment means and divided by standard error, with additions that make Hedges' g more appropriate for SCD.
  - Ability to compare sets of non-missing data across phases and tiers.
  - Effect size equation shows the difference between the unweighted means for all baseline and treatment data, over a denominator of the pooled standard deviation for both sets of data.
  - Hedges' g also features a small sample size correction, which mitigates positive bias of sample sizes using a small number of cases to some extent.

http://faculty.ucmerced.edu/wshadish/software/software-meta-analysis-single-case-design

Analysis

- Visual checks for autocorrelation-nothing noted
- Analyzed data by both graph and dependent variable (depending on study design)
- Pearson's *R* values obtained between *M* visual analysis scores and each of the following Tau-U and Hedges' g
  - Not calculated for R-IRD
- Percentile ranks calculated for Visual analysis, Tau-U and Hedges' g

RESULTS AND DISCUSSION

Visual analysis scores

- 26 cases
- M = 4.34/7
- Median score = 4/7

Visual Analysis Matrix

	1	2	3	4	5	6	7
Level and/or Trend	Data are stable in few phases or conditions and few tiers		Data are moderately stable in most phases or conditions and most tiers	X	Data are moderately stable in all phases or conditions and all tiers.		Data are highly stable in all phases or conditions and all tiers
Variability	An increase or reduction in variability as predicted is *not* apparent in any phases/conditions or tiers		An increase or reduction in variability as predicted is apparent in few phases and few tiers X		An increase or reduction in variability as predicted is apparent in most phases and most tiers		An increase or reduction in variability as predicted is apparent in all phases and all tiers
Overlap	There is between 0-15% nonoverlapping data for all phase/condition comparisons and all tiers		There is approximately 40% nonoverlapping data for all phase/condition comparisons and all tiers	X	There is no more than 70% nonoverlapping data for all phase/condition comparisons and all tiers		There is 100% nonoverlapping data for all phase/condition comparisons and all tiers
Immediacy of effect	An effect is *not* observed in the predicted direction within the first three to five data points for any phase changes or tiers		A small to moderate effect is observed in the predicted direction within the first three to five data points for at least one phases change and one tier		A moderate to large effect is observed in the predicted direction within the first first three to five data points for [illegible] phases changes X most tiers		A moderate to large effect is observed in the predicted direction within the first three to five data points for all phase changes and all tiers
Consistency of similar phases or conditions	No similar phases contain consistency with regard to level, trend, and variability, including within case and across cases		Few or some similar phases contain consistency with regard to level, trend, and variability, including within case and across cases.		Most similar phases contain consistency with regard to level, trend, and variability, including within case and across cases	X	All similar phases contain consistency with regard to level, trend, and variability, including within case and across cases
Overall impression of strength of functional relation					X		
Averaged score							

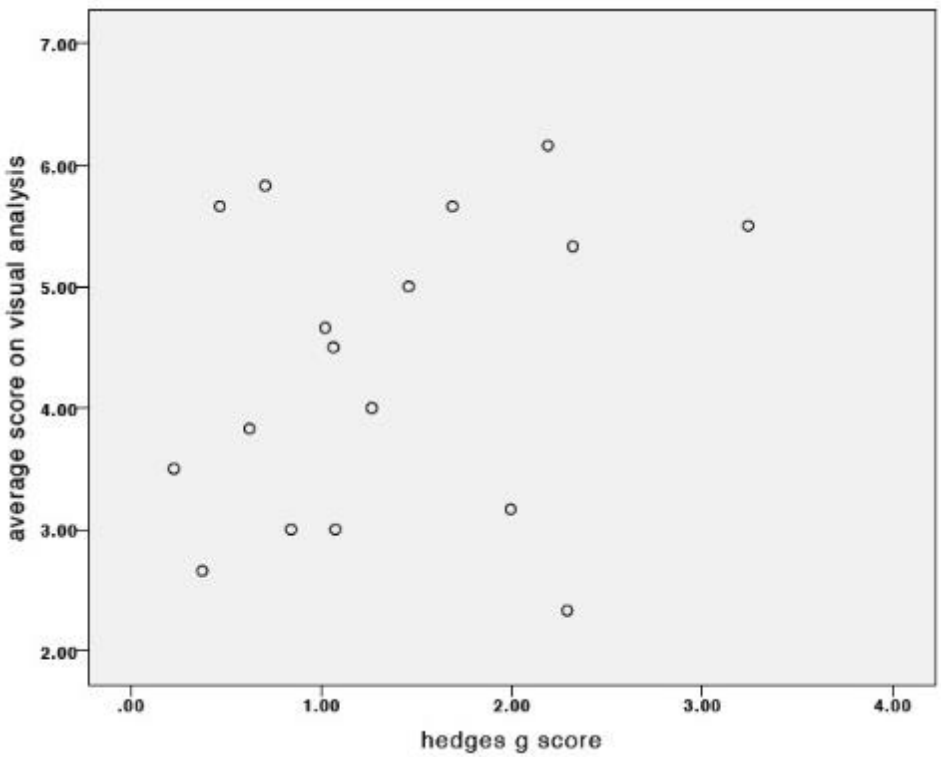
Correlation to traditional visual analysis

Pearson's R values among visual analysis, Tau-U, and Hedges' g

	M (*SD*)	Tau-U	Hedges' g
Visual analysis	4.34 (1.24)	.63**	.26
Tau-U	0.74 (0.17)		.50*
Hedges' g	1.34 (0.83)		

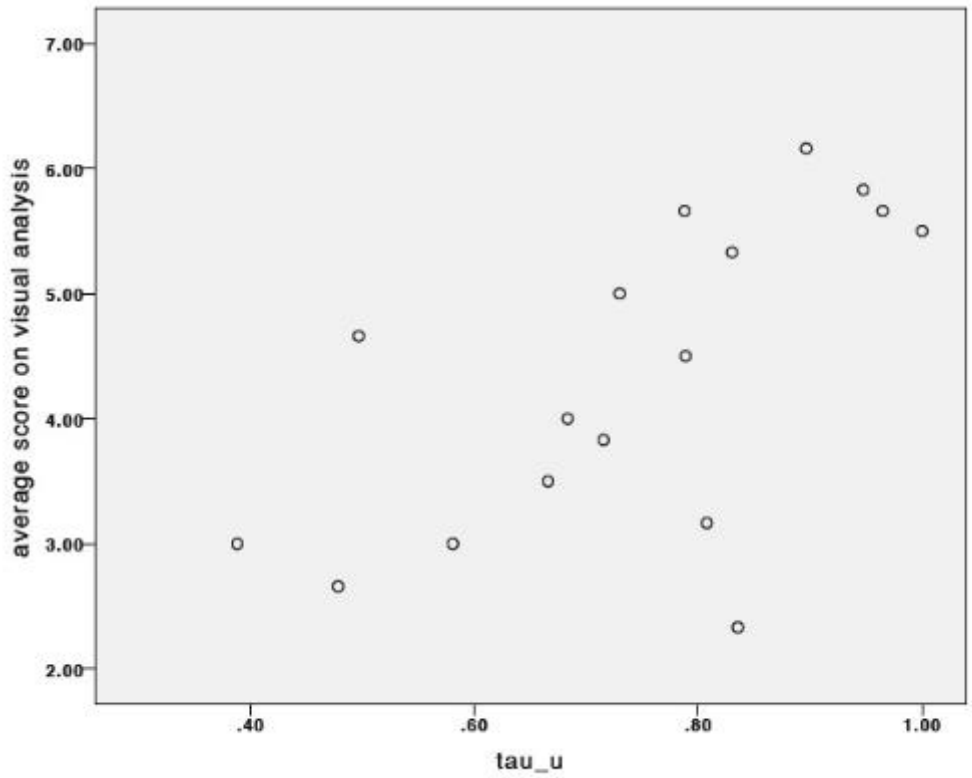
***significant at 0.05 level** ****significant at 0.01 level**

Hedges' G and Visual analysis



Hedges g-an effect size of 0.8 is a large effect (Cohen, 1988)

Tau-U and visual analysis



Tau-U
-effect size of 0.8 is a large effect
-effect size of .5-.79 is moderate

R-IRD

- 59 cases
- M score 0.72 (Moderate effects)
- Range 0-1
- Median score 0.78 (Large and very large effects)
- IQR = .63 to .95

Discriminability among SCR results

Percentile Ranks for Visual Analysis, Tau-U, and Hedges' g

	10th	25th	50th	75th	90th
Visual analysis	2.59	3.08	4.5	5.58	5.89
Tau-U	0.46	0.62	0.78	0.86	0.97
Hedges' g	0.35	0.66	1.07	2.09	2.50

Summary of findings

- Typical scores across measures are within moderate to very large effects range
- R-IRD corresponded to traditional visual analysis
- Tau-U corresponded to traditional visual analysis
  - Not surprising given recent meta analysis (e.g., Ninci et al., 2015)
- Hedges g did not correlate to visual analysis, but does to Tau-U
  - Large Hedges' g scores do not correlate to visual analysis

Limitations & Future research

- Data met assumptions for Hedges' g but a sample size over 20 is considered strong and we had 18 (Kline, 2004)
  - Increase sample size (kicked out many due to not meeting assumptions for all ES estimate s).
- Variety of dependent variables
- Tau-U considers Phase A trend, IRD does not
- Data on generalization and maintenance not assessed
- R-IRD-did not run confidence intervals-false sense of precision (Parker, Vannest, & Brown, 2009)
- Cascading logic model (need to include child outcome data)
- Compare group design effect size with Hedges' g scores for single-case research