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

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

| 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.
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>
<thead>
<tr>
<th>Article (Authors; Year)</th>
<th>Single-Subject Research Design</th>
<th>Quality Appraisal Scores and Rankings based on SSED Appraisal Tools (maximum quality score)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Certainty Framework</td>
</tr>
<tr>
<td>Crozier &amp; Tincani, 2005</td>
<td>Withdrawal (A-B-A-C)</td>
<td>“Inconclusive”</td>
</tr>
<tr>
<td>Ganz &amp; Sigafoos, 2005</td>
<td>Changing Criterion</td>
<td>“Suggestive”</td>
</tr>
<tr>
<td>Ozdemir, 2008</td>
<td>Multiple Baseline Across Participants</td>
<td>“Preponderant”</td>
</tr>
<tr>
<td>Tincani, 2004</td>
<td>Alternating Treatment</td>
<td>“Preponderant”</td>
</tr>
</tbody>
</table>

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.

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)
- $d$-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) \( \Rightarrow \) 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)

<table>
<thead>
<tr>
<th></th>
<th>IRD</th>
<th>NAP</th>
<th>PAND</th>
<th>PEM</th>
<th>PDO²</th>
<th>PND</th>
<th>PND-C</th>
<th>VA</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRD</td>
<td>1</td>
<td>.632*</td>
<td>.831*</td>
<td>.824*</td>
<td>.913*</td>
<td>.914*</td>
<td>.644*</td>
<td>.828*</td>
</tr>
<tr>
<td>NAP</td>
<td>.632*</td>
<td>1</td>
<td>.608*</td>
<td>.606*</td>
<td>.610*</td>
<td>.620*</td>
<td>.711*</td>
<td>.549*</td>
</tr>
<tr>
<td>PAND</td>
<td>.831*</td>
<td>.608*</td>
<td>1</td>
<td>.858*</td>
<td>.872*</td>
<td>.849*</td>
<td>.607*</td>
<td>.781*</td>
</tr>
<tr>
<td>PEM</td>
<td>.824*</td>
<td>.606*</td>
<td>.858*</td>
<td>1</td>
<td>.934*</td>
<td>.845*</td>
<td>.637*</td>
<td>.829*</td>
</tr>
<tr>
<td>PDO²</td>
<td>.913*</td>
<td>.610*</td>
<td>.872*</td>
<td>.943*</td>
<td>1</td>
<td>.941*</td>
<td>.679*</td>
<td>.879*</td>
</tr>
<tr>
<td>PND</td>
<td>.914*</td>
<td>.620*</td>
<td>.849*</td>
<td>.845*</td>
<td>.941*</td>
<td>1</td>
<td>.631*</td>
<td>.892*</td>
</tr>
<tr>
<td>PND-C</td>
<td>.644*</td>
<td>.711*</td>
<td>.607*</td>
<td>.673*</td>
<td>.679*</td>
<td>.631*</td>
<td>1</td>
<td>.631*</td>
</tr>
<tr>
<td>VA</td>
<td>.828*</td>
<td>.549*</td>
<td>.781*</td>
<td>.829*</td>
<td>.879*</td>
<td>.892*</td>
<td>.631*</td>
<td>1</td>
</tr>
</tbody>
</table>

* 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? \( \Rightarrow \) 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


- 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

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)

- 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)

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](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


References (cont.)


References (cont.)

References (cont.)


References (cont.)


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

Considerations for Single Case Researchers

To cite this paper:
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

16% 17%

33%

33%

33%
PIR

- 25%
- 33%
- 50%
- 33%
- 50%
WIR

16%
MTS

16%
IBS Distortions
Ledford, Ayres, Lane, & Lam (2015). JoSE

Interval Size as a Function of DPO

Estimated Data (PIR)

Actual duration

Percentage of Session

0.00 0.50 1.00 1.50 2.00 2.50 3.00 3.50

0 20 40 60 80 100
IBS Distortions
Ledford, Ayres, Lane, & Lam (2015). JoSE

Interval Size as a Function of DPO

Percentage of Session

Estimated Data (MTS)
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
## A-B-A-B Designs

<table>
<thead>
<tr>
<th></th>
<th>Duration</th>
<th>PIR</th>
<th>MTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00</td>
<td>--</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
<td>--</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>0.98</td>
<td>1.00</td>
<td>--</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>5</td>
<td>0.55</td>
<td>0.45</td>
<td>✓</td>
</tr>
<tr>
<td>6</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>7</td>
<td>1.00</td>
<td>0.95</td>
<td>✓</td>
</tr>
<tr>
<td>8</td>
<td>1.00</td>
<td>0.99</td>
<td>✓</td>
</tr>
<tr>
<td>9</td>
<td>0.44</td>
<td>0.44</td>
<td>✓</td>
</tr>
<tr>
<td>10</td>
<td>0.98</td>
<td>0.96</td>
<td>✓</td>
</tr>
<tr>
<td>11</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>12</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
</tbody>
</table>
# A-B-A-B Designs

<table>
<thead>
<tr>
<th></th>
<th>Duration</th>
<th>PIR</th>
<th>MTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00</td>
<td>--</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
<td>--</td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td>0.98</td>
<td>1.00</td>
<td>--</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>5</td>
<td>0.55</td>
<td>0.45</td>
<td>✓</td>
</tr>
<tr>
<td>6</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>7</td>
<td>1.00</td>
<td>0.95</td>
<td>✓</td>
</tr>
<tr>
<td>8</td>
<td>1.00</td>
<td>0.99</td>
<td>✓</td>
</tr>
<tr>
<td>9</td>
<td>0.44</td>
<td>0.44</td>
<td>✓</td>
</tr>
<tr>
<td>10</td>
<td>0.98</td>
<td>0.96</td>
<td>✓</td>
</tr>
<tr>
<td>11</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>12</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
</tbody>
</table>

**PIR**

- **↓** 25%
- **↑** 0%
- **✓** 75%

**MTS**

- **↓** 33%
- **↑** 0%
- **✓** 67%
# ATD Designs

<table>
<thead>
<tr>
<th>Duration</th>
<th>PIR</th>
<th>MTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00</td>
<td>0.50</td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>6</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>7</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>8</td>
<td>0.18</td>
<td>0.29</td>
</tr>
<tr>
<td>9</td>
<td>0.46</td>
<td>-0.11</td>
</tr>
<tr>
<td>10</td>
<td>0.63</td>
<td>0.25</td>
</tr>
<tr>
<td>11</td>
<td>0.38</td>
<td>0.21</td>
</tr>
<tr>
<td>12</td>
<td>0.36</td>
<td>0.44</td>
</tr>
<tr>
<td>13</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>14</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>15</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>16</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>17</td>
<td>0.30</td>
<td>0.50</td>
</tr>
<tr>
<td>18</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>19</td>
<td>0.88</td>
<td>1.00</td>
</tr>
<tr>
<td>20</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>21</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>22</td>
<td>0.44</td>
<td>0.50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Duration</th>
<th>PIR</th>
<th>MTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.36</td>
<td>0.44</td>
</tr>
<tr>
<td>13</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>14</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>15</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>16</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>17</td>
<td>0.30</td>
<td>0.50</td>
</tr>
<tr>
<td>18</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>19</td>
<td>0.88</td>
<td>1.00</td>
</tr>
<tr>
<td>20</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>21</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>22</td>
<td>0.44</td>
<td>0.50</td>
</tr>
</tbody>
</table>

- PIR: 14% ✓ 27% ➡
- MTS: 27% ➡ 14% ✓ 59%
<table>
<thead>
<tr>
<th></th>
<th>Duration</th>
<th>PIR</th>
<th>MTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00</td>
<td>0.50</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>5</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>6</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>7</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>8</td>
<td>0.18</td>
<td>0.29</td>
<td>✓</td>
</tr>
<tr>
<td>9</td>
<td>0.46</td>
<td>-0.11</td>
<td>✓</td>
</tr>
<tr>
<td>10</td>
<td>0.63</td>
<td>0.25</td>
<td>✓</td>
</tr>
<tr>
<td>11</td>
<td>0.38</td>
<td>0.21</td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Duration</th>
<th>PIR</th>
<th>MTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.36</td>
<td>0.44</td>
<td>✓</td>
</tr>
<tr>
<td>13</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>14</td>
<td>0.25</td>
<td>0.25</td>
<td>✓</td>
</tr>
<tr>
<td>15</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>16</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>17</td>
<td>0.30</td>
<td>0.50</td>
<td>✓</td>
</tr>
<tr>
<td>18</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>19</td>
<td>0.88</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>20</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>21</td>
<td>1.00</td>
<td>1.00</td>
<td>✓</td>
</tr>
<tr>
<td>22</td>
<td>0.44</td>
<td>0.50</td>
<td>✓</td>
</tr>
</tbody>
</table>
## ATD Designs

<table>
<thead>
<tr>
<th>PIR</th>
<th>Duration</th>
<th>PIR</th>
<th>MTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00</td>
<td>0.50</td>
<td>▶️</td>
</tr>
<tr>
<td>2</td>
<td>0.18</td>
<td>0.29</td>
<td>▶️</td>
</tr>
<tr>
<td>3</td>
<td>0.46</td>
<td>-0.11</td>
<td>▼️</td>
</tr>
<tr>
<td>4</td>
<td>0.63</td>
<td>0.25</td>
<td>▼️</td>
</tr>
<tr>
<td>5</td>
<td>0.38</td>
<td>0.21</td>
<td>▼️</td>
</tr>
<tr>
<td>6</td>
<td>0.36</td>
<td>0.44</td>
<td>▶️</td>
</tr>
<tr>
<td>7</td>
<td>0.25</td>
<td>0.25</td>
<td>✔️</td>
</tr>
<tr>
<td>8</td>
<td>0.30</td>
<td>0.50</td>
<td>▶️</td>
</tr>
<tr>
<td>9</td>
<td>0.88</td>
<td>1.00</td>
<td>▶️</td>
</tr>
<tr>
<td>10</td>
<td>0.44</td>
<td>0.50</td>
<td>▶️</td>
</tr>
</tbody>
</table>

**PIR**

- ▼️ 30%
- ▶️ 60%
- ✔️ 10%

**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


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:
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).
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).
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
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.
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).
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.
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., A₁ vs B₁ and A₂ vs B₂).
- 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).
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)

<table>
<thead>
<tr>
<th>Inclusion/Exclusion Criteria</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1st criterion</strong></td>
<td>.873</td>
</tr>
<tr>
<td><strong>2nd criterion</strong></td>
<td>1.000</td>
</tr>
<tr>
<td><strong>3rd criterion</strong></td>
<td>.978</td>
</tr>
<tr>
<td><strong>4th criterion</strong></td>
<td>1.000</td>
</tr>
<tr>
<td><strong>5th criterion</strong></td>
<td>1.000</td>
</tr>
<tr>
<td><strong>6th criterion</strong></td>
<td>.985</td>
</tr>
<tr>
<td><strong>7th criterion</strong></td>
<td>.986</td>
</tr>
<tr>
<td><strong>8th criterion</strong></td>
<td>.944</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Design Standards</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall standard</td>
<td>.850</td>
</tr>
<tr>
<td>DS#1: Independent variable</td>
<td>1.000</td>
</tr>
<tr>
<td>DS#2A: aIOA Collected</td>
<td>.803</td>
</tr>
<tr>
<td>DS#2B: IOA 20%</td>
<td>.900</td>
</tr>
<tr>
<td>DS#2C: Minimum quality thresholds of IOA</td>
<td>.722</td>
</tr>
<tr>
<td>DS#3: Replication effects</td>
<td>.827</td>
</tr>
<tr>
<td>DS#4: Number of data points</td>
<td>.712</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Moderators</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant age</td>
<td>1.000</td>
</tr>
<tr>
<td>Participant communication/language level</td>
<td>.744</td>
</tr>
<tr>
<td>Independent variable</td>
<td>.843</td>
</tr>
<tr>
<td>Dependent variable</td>
<td>.739</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Raw Data</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>.869</td>
</tr>
<tr>
<td>Intervention</td>
<td>.895</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5</td>
<td>NA</td>
<td>NA</td>
<td>26</td>
<td>NA</td>
<td>47</td>
<td>12</td>
</tr>
<tr>
<td>1</td>
<td>35</td>
<td>67</td>
<td>66</td>
<td>31</td>
<td>64</td>
<td>5</td>
<td>38</td>
</tr>
<tr>
<td>0</td>
<td>29</td>
<td>2</td>
<td>3</td>
<td>12</td>
<td>5</td>
<td>17</td>
<td>19</td>
</tr>
</tbody>
</table>

a IOA-Interobserver agreement
### Results: Number of studies, participants, analyses and Tau results - Age

<table>
<thead>
<tr>
<th>Age</th>
<th>Number of Studies</th>
<th>Number of Study Participants</th>
<th>Number of Analyses</th>
<th>Group Tau $[Cl_{95}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>aPRESCH</td>
<td>33</td>
<td>103</td>
<td>244</td>
<td>0.659904</td>
</tr>
<tr>
<td>bELEM</td>
<td>21</td>
<td>45</td>
<td>113</td>
<td>0.604178</td>
</tr>
<tr>
<td>cSEC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>dADULT</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>0.593097</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of Studies</th>
<th>Number of Study Participants</th>
<th>Number of Analyses</th>
<th>Group Tau $[Cl_{95}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>aSPSOMESPON</td>
<td>11</td>
<td>30</td>
<td>52</td>
<td>0.65077</td>
</tr>
<tr>
<td>bSPNOTSPON</td>
<td>14</td>
<td>42</td>
<td>100</td>
<td>0.594762</td>
</tr>
<tr>
<td>cNOSP</td>
<td>9</td>
<td>21</td>
<td>62</td>
<td>0.734505</td>
</tr>
</tbody>
</table>

$^a$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

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Number of Studies</th>
<th>Number of Study Participants</th>
<th>Number of Analyses</th>
<th>Group Tau [Cl₉₅]</th>
</tr>
</thead>
<tbody>
<tr>
<td>aADI</td>
<td>23</td>
<td>85</td>
<td>178</td>
<td>0.646182</td>
</tr>
<tr>
<td>bIWD</td>
<td>14</td>
<td>63</td>
<td>167</td>
<td>0.676521</td>
</tr>
</tbody>
</table>

aADI-adult-led didactic instructions, bIWD-individual with disabilities-led instructions

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

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Number of Studies</th>
<th>Number of Study Participants</th>
<th>Number of Analyses</th>
<th>Group Tau [Cl₉₅]</th>
</tr>
</thead>
<tbody>
<tr>
<td>aAAC</td>
<td>6</td>
<td>11</td>
<td>35</td>
<td>0.765988</td>
</tr>
<tr>
<td>bVOC</td>
<td>18</td>
<td>78</td>
<td>130</td>
<td>0.546551</td>
</tr>
<tr>
<td>cSOC</td>
<td>11</td>
<td>32</td>
<td>70</td>
<td>0.663016</td>
</tr>
<tr>
<td>dNOVOC</td>
<td>6</td>
<td>24</td>
<td>36</td>
<td>0.789988</td>
</tr>
</tbody>
</table>

aAAC-use of augmentative and alternative communication systems, bVOC-vocalization, verbalize target words, cSOC-social behaviors, joint attention, social play skills, social interpersonal skills, dNOVOC-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.
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


Questions?

Thank you!
Contact: Ee Rea Hong

irehong@human.tsukuba.ac.jp
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:

**Visual analysis**

- **Tau-U**: 0.8964
- **Hedges g (across study)**: 2.19

**Figure 5.** Percentage of opportunities the parents provided their children to use the cards during baseline, intervention, and follow-up.
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)
<table>
<thead>
<tr>
<th>Journal</th>
<th># studies included</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal of Applied Behavior Analysis</td>
<td>3</td>
</tr>
<tr>
<td>Research in Developmental Disabilities</td>
<td>3</td>
</tr>
<tr>
<td>Journal of Early Intervention</td>
<td>2</td>
</tr>
<tr>
<td>Early Education and Development</td>
<td>1</td>
</tr>
<tr>
<td>Journal of Emotional and Behavioral Disorders</td>
<td>1</td>
</tr>
<tr>
<td>Journal of Speech, Language, Hearing Research</td>
<td>1</td>
</tr>
<tr>
<td>Journal of Positive Behavior Interventions</td>
<td>1</td>
</tr>
<tr>
<td>Scholarly Inquiry for Nursing Practice: An International Journal</td>
<td>1</td>
</tr>
<tr>
<td>Targeted parent behavior</td>
<td># studies</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Naturalistic language intervention (i.e. EMT, imitation training, natural language paradigm)</td>
<td>6</td>
</tr>
<tr>
<td>Instructional intervention (i.e. DTT, generalized teaching strategy)</td>
<td>2</td>
</tr>
<tr>
<td>Applied behavior analysis (instructional prompts, general teaching strategies)</td>
<td>2</td>
</tr>
<tr>
<td>Arrangement of opportunities for AAC use</td>
<td>1</td>
</tr>
<tr>
<td>Joint attention intervention</td>
<td>1</td>
</tr>
<tr>
<td>Prompting pro-social sibling interactions</td>
<td>1</td>
</tr>
</tbody>
</table>
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
### Tau-U Calculator

- contrast
- chart
- clear all
- correct baseline

### Results

<table>
<thead>
<tr>
<th>id</th>
<th>Label</th>
<th>S</th>
<th>PAIRS</th>
<th>TAU</th>
<th>TAUb</th>
<th>VARs</th>
<th>SD</th>
<th>SDao</th>
<th>Z</th>
<th>P Value</th>
<th>CI 85%</th>
<th>CI 90%</th>
<th>CI 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>trend:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phase:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>corrected baseline:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>combined:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Weighted Average

<table>
<thead>
<tr>
<th>Label</th>
<th>Tau</th>
<th>Var-Tau</th>
<th>Z</th>
<th>P-Value</th>
<th>CI 85%</th>
<th>CI 90%</th>
<th>CI 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[www.singlecaseresearch.org](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

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**

<table>
<thead>
<tr>
<th>Level and/or Trend</th>
<th>Visual Analysis Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data are stable in few phases or conditions and few tiers</td>
</tr>
<tr>
<td>2</td>
<td>Data are moderately stable in most phases or conditions and most tiers</td>
</tr>
<tr>
<td>3</td>
<td>Data are moderately stable in all phases or conditions and all tiers.</td>
</tr>
<tr>
<td>4</td>
<td>An increase or reduction in variability as predicted is apparent in few phases and few tiers</td>
</tr>
<tr>
<td>5</td>
<td>An increase or reduction in variability as predicted is apparent in most phases and most tiers</td>
</tr>
<tr>
<td>6</td>
<td>An increase or reduction in variability as predicted is apparent in all phases and all tiers</td>
</tr>
<tr>
<td>Overlap</td>
<td>There is between 0-15% nonoverlapping data for all phase/condition comparisions and all tiers</td>
</tr>
<tr>
<td></td>
<td>There is approximately 40% nonoverlapping data for all phase/condition comparisions and all tiers</td>
</tr>
<tr>
<td></td>
<td>There is no more than 70% nonoverlapping data for all phase/condition comparisions and all tiers</td>
</tr>
<tr>
<td></td>
<td>There is 100% nonoverlapping data for all phase/condition comparisions and all tiers</td>
</tr>
<tr>
<td>Immediacy of effect</td>
<td>An effect is not observed in the predicted direction within the first three to five data points for any phase changes or tiers</td>
</tr>
<tr>
<td></td>
<td>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</td>
</tr>
<tr>
<td></td>
<td>A moderate to large effect is observed in the predicted direction within the first three to five data points for some phases change and most tiers</td>
</tr>
<tr>
<td></td>
<td>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</td>
</tr>
<tr>
<td>Consistency of similar phases or conditions</td>
<td>No similar phases contain consistency with regard to level, trend, and variability, including within case and across cases</td>
</tr>
<tr>
<td></td>
<td>Few or some similar phases contain consistency with regard to level, trend, and variability, including within case and across cases</td>
</tr>
<tr>
<td></td>
<td>Most similar phases contain consistency with regard to level, trend, and variability, including within case and across cases</td>
</tr>
<tr>
<td></td>
<td>All similar phases contain consistency with regard to level, trend, and variability, including within case and across cases</td>
</tr>
<tr>
<td>Overall impression of strength of functional relation</td>
<td></td>
</tr>
<tr>
<td>Averaged score</td>
<td></td>
</tr>
</tbody>
</table>
Correlation to traditional visual analysis

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

<table>
<thead>
<tr>
<th></th>
<th>M (SD)</th>
<th>Tau-U</th>
<th>Hedges’ g</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual analysis</td>
<td>4.34 (1.24)</td>
<td>.63**</td>
<td>.26</td>
</tr>
<tr>
<td>Tau-U</td>
<td>0.74 (0.17)</td>
<td></td>
<td>.50*</td>
</tr>
<tr>
<td>Hedges’ g</td>
<td>1.34 (0.83)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*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

<table>
<thead>
<tr>
<th></th>
<th>10th</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
<th>90th</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Visual analysis</strong></td>
<td>2.59</td>
<td>3.08</td>
<td>4.5</td>
<td>5.58</td>
<td>5.89</td>
</tr>
<tr>
<td><strong>Tau-U</strong></td>
<td>0.46</td>
<td>0.62</td>
<td>0.78</td>
<td>0.86</td>
<td>0.97</td>
</tr>
<tr>
<td><strong>Hedges’ g</strong></td>
<td>0.35</td>
<td>0.66</td>
<td>1.07</td>
<td>2.09</td>
<td>2.50</td>
</tr>
</tbody>
</table>
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