Meta-Analysis of Social Skills Interventions Using Single Subject Design

Jeffrey Chenier, M.A., Katherine Hunter, M.A., Aaron J. Fischer, M.A., Emily Patty, M.A., Kristen O’Leary, Haley York, Natalie Robichaux, Jeremy Liu, Thomas Dressler, Kelsey Hartman, and Frank Gresham, Ph.D.
Overview

• Introduction
  – Social Skills and Evidence-Based Practice
  – Group studies
  – Single subject studies
  – Findings from last 2 years

• Method

• Different Analyses and Results
  – Standardized Mean Difference
  – PAND and Phi coefficient
  – Visual Analysis

• Conclusions

• Discussant
Why Social Skills?!?

• Learned behaviors that enable positive interactions and allow for escape/avoidance of negative interactions

• Academic Enablers (DiPerna & Elliott, 2002)
  – 3rd grade social skills were a better predictor of academic achievement in 8th grade than 3rd grade academic achievement (Caprara et. al, 2000)

• Myriad of problems co-occurring with social skills deficits
  – Both externalizing and internalizing
Intervention in Schools

• No Child Left Behind Act (2001)
• Scientifically Based Research
  – Section 9101(37):
    Research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities and programs
What does this look like?

• What Works Clearinghouse (WWC, 2011): Group Designs
  – Meet Evidence Standards:
    • “well-designed and well-implemented randomized controlled trials”
  – Meet Evidence Standards with reservations:
    • “quasi-experiments with no severe design or implementation problems OR randomized clinical trials with severe design or implementation problems”
Aggregating Studies

• WWC (2011)
  – Use effect size indicators for study results
  – Allows the comparison of effects across studies
  – For studies in the same domain, average across studies

• Meta-Analysis
  – Objective method for summarizing a body of empirical findings (Strube & Hartmann, 1982)
  – Emphasizes the direction and magnitude of effects across studies for a particular intervention
Social Skills Meta-Analyses: Group Designs

• Does social skills training work?
  - Gresham, Cook, Crews, and Kern, 2004

<table>
<thead>
<tr>
<th>Meta-Analysis</th>
<th>n studies</th>
<th>ES g</th>
<th>ES r</th>
<th>BESD Control</th>
<th>BESD Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ang &amp; Hughes (2001)</td>
<td>38</td>
<td>.62</td>
<td>.30</td>
<td>35%</td>
<td>65%</td>
</tr>
<tr>
<td>Beelman et al. (1994)</td>
<td>49</td>
<td>.47</td>
<td>.23</td>
<td>38%</td>
<td>62%</td>
</tr>
<tr>
<td>Losel &amp; Beelman (2003)</td>
<td>84</td>
<td>.38</td>
<td>.19</td>
<td>40%</td>
<td>60%</td>
</tr>
<tr>
<td>Schneider (1992)</td>
<td>79</td>
<td>.89</td>
<td>.40</td>
<td>30%</td>
<td>70%</td>
</tr>
<tr>
<td>Schneider &amp; Byrne (1985)</td>
<td>51</td>
<td>.65</td>
<td>.31</td>
<td>34%</td>
<td>66%</td>
</tr>
<tr>
<td>Quinn et al. (1999)</td>
<td>35</td>
<td>.20</td>
<td>.10</td>
<td>45%</td>
<td>55%</td>
</tr>
</tbody>
</table>

Means

\[ M = .60 \] \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad M = .29 \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad M = 35\% \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad M = 65\% \]
Social Skills Meta-Analyses: Group Designs

• Does social skills training work?
  – Godbold et. al, 2010
    • 34 group design studies
      – Random Assignment with Equivalent Starting Groups
        » $g = 0.67$, $p<0.05$; BESD treatment = 82%
        » Significantly higher than quasi-experimental designs
What about Single Subject Designs?

• Definition of a single subject design:
  – One (or several) individuals serving as their own control being exposed to one (or several) conditions of an independent variable and comparing changes in a measurable dependent variable across those conditions (Kazdin, 1982)
Types of Single Case Designs

• Four primary designs:
  – Reversal-withdrawal (ABAB)
  – Multiple baseline
  – Changing criterion
  – Alternating treatments
Reversal/Withdrawal (ABAB)
Multiple Baseline

Subject A

Baseline | Introduction of Treatment
---------|--------------------------

Subject B

Baseline | Introduction of Treatment
---------|--------------------------

Subject C

Baseline | Introduction of Treatment
---------|--------------------------
Evidence in Single Subject Designs

  • IV must be systematically manipulated, with the researcher determining when and how the IV conditions change
  • Each outcome variable must be measured systematically over time by more than one assessor, and the study needs to have IOA calculated 20% of the time in each condition, and IOA percentage must meet minimum thresholds
    – 0.80 IOA or 0.60 Cohen’s Kappa
  • Study must include at least three attempts to demonstrate an intervention effect at three different points in time or with three different phase repetitions
    – Phase must have a minimum of three data points
Single Subject Effect Sizes?

- No agreed upon method or gold standard to calculate effect sizes from single-case design research (WWC, 2010)
  - Problems
    - How to quantify the effect?
      - How accurate is the effect?
    - How comparable are the effects across other SC designs?
    - How comparable are the effects compared to group design effect sizes?
Nonparametric Methods

– Percentage of Nonoverlapping Data (PND), Percentage of All Nonoverlapping Data (PAND), Percent Exceeding the Median (PEM)

• “Distributional properties of these measures are unknown, so standard errors and statistical tests are not formally justified.”
  – Additionally, trend is not addressed
• Because of the lack of statistical justification, only use if an approximate size of the effect is desired.
• Wolery et al. (2010) compared four nonparametric methods to visual inspection of effect and each method had its own host of issues, so much that they called for their abandonment.
Parametric Methods

– Regression Estimates
  • Advantages
    – Familiarity
    – Ability to model trends
    – Ability to attain an Effect Size from a single case
  • Disadvantages
    – Inability to deal with complex structures present in single case design
    – More complicated
Parametric Methods

– Multilevel Modeling (Ex. HLM)
  • Advantages
    – Ability to account for complexity of design
    – Ability to accommodate autocorrelation
    – Tests for change over time for level, trend, and level and trend
    – Accounts for initial level of target behavior
    – Usable on nearly every design
    – Accommodates missing data or unequal intervals between measurements
  • Disadvantages
    – Unfamiliarity
    – Technically challenging and time consuming
    – Different metric from group design Effect Sizes, therefore the estimate is not comparable
Quantitative Methods

– Differing methods to calculate a Standardized Mean Difference (SMD) statistic

  • Advantages
    – Encourages inclusion of SC designs in evaluating effects of interventions
    – Potentially gives another method in which to rank order interventions

  • Disadvantages
    – Not completely comparable to group design research
      » Pooled within-group variance not comparable to pooled within phase variance
    – Small n leads to imprecise estimates
    – Trend is not assessed
Summary of Effect Size Estimators for SC Design (WWC, 2010)

• Simply put, science is not there yet
• Nonparametric estimators should be reported with a parametric estimator (regression, HLM)
• Quantitative methods are not as statistically sound as they should be, but the base from which to build is present
### SC Meta Analyses Since 2000

<table>
<thead>
<tr>
<th>Meta Analysis</th>
<th>Disability Type</th>
<th>n studies</th>
<th>Intervention Type</th>
<th>ES statistic</th>
<th>ES</th>
<th>Degree of Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bellini &amp; Akullian, 2007</td>
<td>Autism Spectrum</td>
<td>15</td>
<td>Video Modeling</td>
<td>PND</td>
<td>81%</td>
<td>Effective</td>
</tr>
<tr>
<td>Bellini &amp; Akullian, 2008</td>
<td>Autism Spectrum</td>
<td>7</td>
<td>Video Self Monitoring</td>
<td>PND</td>
<td>77%</td>
<td>Questionable/Effective</td>
</tr>
<tr>
<td>Bellini et al., 2007</td>
<td>Autism</td>
<td>15</td>
<td>Child Specific</td>
<td>PND</td>
<td>71%</td>
<td>Questionable</td>
</tr>
<tr>
<td>Bellini et al., 2007</td>
<td>Autism</td>
<td>7</td>
<td>Collateral Skills</td>
<td>PND</td>
<td>75%</td>
<td>Questionable</td>
</tr>
<tr>
<td>Bellini et al., 2007</td>
<td>Autism</td>
<td>20</td>
<td>Comprehensive</td>
<td>PND</td>
<td>72%</td>
<td>Questionable</td>
</tr>
<tr>
<td>Bellini et al., 2007</td>
<td>Autism</td>
<td>10</td>
<td>Peer Mediated</td>
<td>PND</td>
<td>62%</td>
<td>Questionable</td>
</tr>
<tr>
<td>Bellini et al., 2007</td>
<td>Autism</td>
<td>55</td>
<td>Total</td>
<td>PND</td>
<td>70%</td>
<td>Questionable</td>
</tr>
<tr>
<td>Kokina &amp; Kern, 2010</td>
<td>Autism Spectrum</td>
<td>18</td>
<td>Social Stories</td>
<td>PND</td>
<td>60%</td>
<td>Questionable</td>
</tr>
<tr>
<td>Wang &amp; Spillane, 2009</td>
<td>Autism</td>
<td>2</td>
<td>CBT</td>
<td>d</td>
<td>.47-.1.24, .24-.59</td>
<td>Medium - Large, Small - Medium</td>
</tr>
<tr>
<td>Wang &amp; Spillane, 2009</td>
<td>Autism</td>
<td>1</td>
<td>CBT</td>
<td>PND</td>
<td>100%</td>
<td>Very Promising</td>
</tr>
<tr>
<td>Wang &amp; Spillane, 2009</td>
<td>Autism</td>
<td>9</td>
<td>Others</td>
<td>PND</td>
<td>80.77%</td>
<td>Effective</td>
</tr>
<tr>
<td>Wang &amp; Spillane, 2009</td>
<td>Autism</td>
<td>9</td>
<td>Peer Mediated</td>
<td>PND</td>
<td>60.69%</td>
<td>Questionable</td>
</tr>
<tr>
<td>Wang &amp; Spillane, 2009</td>
<td>Autism</td>
<td>6</td>
<td>Social Stories</td>
<td>PND</td>
<td>67.21%</td>
<td>Questionable</td>
</tr>
<tr>
<td>Wang &amp; Spillane, 2009</td>
<td>Autism</td>
<td>11</td>
<td>Video Modeling</td>
<td>PND</td>
<td>84.25%</td>
<td>Effective</td>
</tr>
<tr>
<td>Durlak &amp; Weissberg, 2010</td>
<td>Typical</td>
<td>68</td>
<td>After School Programs (Targeting Positive Social Behaviors)</td>
<td>g</td>
<td>0.19</td>
<td>Small (Significantly different from zero)</td>
</tr>
<tr>
<td>Schneider et al, 2008</td>
<td>Autism</td>
<td>19</td>
<td>Social Behavior</td>
<td>phi</td>
<td>0.72</td>
<td>Large</td>
</tr>
</tbody>
</table>
Summary

- 7 total studies
  - 6 focusing on children with Autism Spectrum Disorders, 1 for children who are typically developing

- Multiple interventions available

- Effectiveness
  - 3 very effective
  - 3 moderately effective
  - 1 not as effective (but still statistically significant)

- Statistic
  - 4: PND
  - 2: SMD (d or g)
  - 1: phi coefficient
Single Subject Meta-Analysis

• Chenier et al., 2010
  – Meta-Analysis of single subject social skills intervention studies from 2000-2009
  – Used standardized mean difference and found a large effect overall

• Current study
  – Extend last year’s study to include articles from 2010
  – Use multiple measures of effect size and compare results across methods
Method

- Literature Search, 2000-2010
  - Keyword Search in PsycINFO

- 6782 Articles Total

<table>
<thead>
<tr>
<th>Social Skills +</th>
<th>Competence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention</td>
</tr>
<tr>
<td></td>
<td>Training</td>
</tr>
</tbody>
</table>
Method: Coding 1

• Abstract Review
  – Primary Inclusion Criteria
    • Must be a single-subject design
    • No books, book chapters, reviews, meta-analyses, or dissertations
  – Secondary Inclusion Criteria
    • Social skills intervention or targets a social skill
    • Focuses on ages 3-21 (through high school)
    • Does NOT target drugs, alcohol, or sexual offenders
  – IOA of 99% (on approximately 22% of articles)
  – 341 studies remaining
Method: Coding 2

• Inclusion Criteria
  – Full article in English
  – Article includes single-subject graphs
    • AB graphs alone are not sufficient but all other types qualify
  – Does study fit our Social Skills definition? (Gresham, Van, and Cook, 2006)
    – Facilitates initiating and maintain positive social relationships
    – Contributes to peer acceptance and friendship development
    – Results in satisfactory school adjustment
    – Allows individuals to cope with and adapt to the demands of the social environment
  – Not part of a larger treatment package
Method: Coding 2

- Coded 225 unique studies
- IOA of 89% (on approximately 36% of studies)
- 82 studies initially passed to Coding 3
- Upon further inspection, only 73 studies were eligible for inclusion in the final analyses
Method: Coding 3

- Participant demographics
  - Number
  - Age Range
  - Gender
  - Race
  - Disability Type

- Dependent Variables
  - Number
  - Setting
  - How measured
  - Design
  - Efficacy
  - IOA
  - Study Quality (Horner, 2005)
    - Three replication across or within
    - Treatment Integrity
    - DV operationally defined
    - IOA (above .8, over 20%)

73 studies; IOA of 100% on 21% of studies
## Demographics

<table>
<thead>
<tr>
<th>Participant Age</th>
<th>Ethnicity</th>
<th>Disability Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ages 0-2</td>
<td>Caucasian</td>
<td>Typically Developing</td>
<td>61</td>
</tr>
<tr>
<td>Ages 3-5</td>
<td>African American</td>
<td>ADHD</td>
<td>69</td>
</tr>
<tr>
<td>Ages 6-9</td>
<td>Hispanic</td>
<td>ODD</td>
<td>11</td>
</tr>
<tr>
<td>Ages 10-12</td>
<td>Asian American</td>
<td>Asperger's</td>
<td>10</td>
</tr>
<tr>
<td>Ages 13-15</td>
<td>Vietnamese</td>
<td>Autism</td>
<td>1</td>
</tr>
<tr>
<td>Ages 16-18</td>
<td>Native American</td>
<td>PDD</td>
<td>1</td>
</tr>
<tr>
<td>Ages 19-21</td>
<td>Other</td>
<td>EBD</td>
<td>1</td>
</tr>
</tbody>
</table>

**Number of Participants: 309**

Males: 226; Females: 62

*As reported by the authors: not all demographics were reported in every article

**Anderson-Butcher, 2003 analyzed a whole school, thus these data are not included
Analyses

• 3 different analyses
  – Standardized Mean Difference Effect Size
  – Percent of All Non-Overlapping Data
  – Visual Analysis

• For each one:
  – Overall effect size across all studies
  – Broken down by treatment (IV)
  – Broken down by measured outcome (DV)
  – Broken down by study quality
Analyses: Treatments (IVs)

- Computer Based Programs; n = 4
- Manualized Social Skills Training; n = 16
- Peer-Mediated Interventions; n = 15
- Reinforcement-Based Training; n = 4
- Self-Monitoring; n = 7
- Social Stories; n = 8
- Unmanualized Social Skills Training; n = 17
- Video Modeling; n = 9
Analyses: Outcome Measures (DVs)

- Academic Behaviors (i.e. AET, following instructions, opportunities to respond); n = 13
- Adaptive/Functional Skills (i.e. joint attention, sequence of steps, purchasing items); n = 5
- Conversation Skills (i.e. eye contact, requests, appropriate responses, turn-taking); n = 28
- Emotion/Empathy Skills; n = 4
- Initiation; n = 13
- Negative Behavior (i.e. disruptive behavior, aggressive behavior); n = 9
- Negative Interactions; n = 9
- Neutral Interactions; n = 3
- Play Skills (i.e. reciprocal play, playing games, sportsmanship); n = 18
- Positive Interactions; n = 25
- Rating Scales (i.e. CEI, Number of correct skills); n = 6
- Social Skills; n = 5
Analyses: Study Quality (Horner, 2005)

• Each article given one point for:
  – Having three replications across or within graphs
  – Reporting Treatment Integrity (94% of articles said they took treatment integrity data but only 73% reported the data)
  – Operational definition of the DV
  – Reporting IOA above .80 for at least 20% of studies
* Took out operational definition of IV criteria due to almost all studies doing this

• Quality

  4 : n = 35
  3 : n = 29
  2 : n = 7
  1 : n = 2
First Analysis

• Quantitative Analysis
• Standardized Mean Difference: Hedge’s g
Method

• Data Extraction
  – UnGraph (Biosoft, 2004)
    • Extracts numerical data from graphs and puts it into Microsoft Excel
  – High reliability and validity in collecting data from single subject graphs (Shadish et al., 2009)
Method

• Included Articles
  – 50 of 73 original articles were able to be analyzed
Method

• UnGraph Procedures
  – Extract graph from article PDF
  – Save graph as a separate file
  – Open file in UnGraph
Method

- UnGraph Procedures
  - Extract graph from article PDF
  - Save graph as a separate file
  - Open file in UnGraph
  - Select coordinates to construct a grid over graph
Method

• **UnGraph Procedures**
  – Extract graph from article PDF
  – Save graph as a separate file
  – Open file in UnGraph
  – Select coordinates to construct a grid over graph
  – **Select each individual data point on graph**
Method

• UnGraph Procedures
  – Extract graph from article PDF
  – Save graph as a separate file
  – Open file in UnGraph
  – Select coordinates to construct a grid over graph
  – Select each individual data point on graph
  – Save data in an Excel file (unedited)
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Session #</td>
<td>BL</td>
<td>Session #</td>
<td>intervention</td>
</tr>
<tr>
<td>0.9662162</td>
<td>0</td>
<td>18.96622</td>
<td>94.7826087</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>20</td>
<td>101.3043478</td>
</tr>
<tr>
<td>2.9121622</td>
<td>55.217391</td>
<td>20.97297</td>
<td>100.8695652</td>
</tr>
<tr>
<td>4.0067568</td>
<td>23.478261</td>
<td>22.00676</td>
<td>101.3043478</td>
</tr>
<tr>
<td>4.9797297</td>
<td>6.5217391</td>
<td>22.91892</td>
<td>101.3043478</td>
</tr>
<tr>
<td>5.9527027</td>
<td>6.9565217</td>
<td>23.89189</td>
<td>101.3043478</td>
</tr>
<tr>
<td>6.9256757</td>
<td>5.6521739</td>
<td>24.92568</td>
<td>101.3043478</td>
</tr>
<tr>
<td>7.8378378</td>
<td>69.565217</td>
<td>25.95946</td>
<td>96.08695652</td>
</tr>
<tr>
<td>8.9324324</td>
<td>64.347826</td>
<td>26.99324</td>
<td>93.04347826</td>
</tr>
<tr>
<td>9.8445946</td>
<td>13.913043</td>
<td>27.96622</td>
<td>100.4347826</td>
</tr>
<tr>
<td>10.817568</td>
<td>0.4347826</td>
<td>29.06081</td>
<td>98.26086957</td>
</tr>
<tr>
<td>11.851351</td>
<td>24.347826</td>
<td>30.03378</td>
<td>100.8695652</td>
</tr>
<tr>
<td>13.006757</td>
<td>51.304348</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.918919</td>
<td>8.2608696</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.952703</td>
<td>10.869565</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.925676</td>
<td>4.3478261</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.898649</td>
<td>3.9130435</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Method

• UnGraph Procedures
  – Extract graph from article PDF
  – Save graph as a separate file
  – Open file in UnGraph
  – Select coordinates to construct a grid over graph
  – Select each individual data point on graph
  – Save data in an Excel file (unedited)
  – Round data to account for human error (edited)
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Session #</strong></td>
<td><strong>BL</strong></td>
<td><strong>Session #</strong></td>
<td><strong>Intervention</strong></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>19</td>
<td>95</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>20</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>55</td>
<td>21</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>22</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>23</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>24</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>70</td>
<td>26</td>
<td>96</td>
</tr>
<tr>
<td>9</td>
<td>64</td>
<td>27</td>
<td>93</td>
</tr>
<tr>
<td>10</td>
<td>14</td>
<td>28</td>
<td>100</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>29</td>
<td>98</td>
</tr>
<tr>
<td>12</td>
<td>24</td>
<td>30</td>
<td>100</td>
</tr>
<tr>
<td>13</td>
<td>51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Method

• UnGraph Procedures
  – Extract graph from article PDF
  – Save graph as a separate file
  – Open file in UnGraph
  – Select coordinates to construct a grid over graph
  – Select each individual data point on graph
  – Save data in an Excel file (unedited)
  – Round data to account for human error (edited)
  – Calculate effect sizes
Method

• Effect Size Calculation (Shadish, 2007)
  - \( G = (\overline{M_t} - \overline{M_b}) / s_b \)
  
  \( \overline{M_t} = \text{Mean of Treatment} \)
  \( \overline{M_b} = \text{Mean of Baseline} \)
  \( s_b = \text{Standard Deviation of Baseline} \)

  - Currently the best quantitative method available, but not absolutely accurate

  - Interpretation of SMD (Cohen, 1988)
    • Effect size of .2 is considered a small effect
    • Effect size of .5 is considered a moderate effect
    • Effect Size of .8 is considered a large effect
Method

1. Calculate the Mean of Baseline & Treatment for Each Individual Graph

*Calculated in Excel
Method

2. Calculate Baseline and Treatment *Mean of Means* Across the Graphs

3. Calculate Standard Deviation of Baseline from the Mean of Means

4. Calculate Effect Size of Target Behavior

*Calculated in Excel*
Results SMD: Overall Effect Size

- Overall Effect Size for Social Skills
  - Calculated by aggregating yielded effect sizes across treatment condition (Independent Variables)
  - Overall Effect Size = 8.36
## Results: SMD by Treatment Type

<table>
<thead>
<tr>
<th>Treatment Type (IV)</th>
<th>Effect Size Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manualized SST</td>
<td>12.96</td>
</tr>
<tr>
<td>Video Modeling</td>
<td>11.19</td>
</tr>
<tr>
<td>Peer Mediated</td>
<td>8.49</td>
</tr>
<tr>
<td>Unmanualized SST</td>
<td>7.77</td>
</tr>
<tr>
<td>Self Monitoring</td>
<td>5.91</td>
</tr>
<tr>
<td>Computer-Based</td>
<td>5.83</td>
</tr>
<tr>
<td>Social Stories</td>
<td>4.60</td>
</tr>
<tr>
<td>Reinforcement-Based</td>
<td>2.99</td>
</tr>
</tbody>
</table>
## Results: SMD by Dependent Variable

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Effect Size Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initiation</td>
<td>18.92</td>
</tr>
<tr>
<td>Adaptive/Functional Skills</td>
<td>17.30</td>
</tr>
<tr>
<td>Positive Interaction</td>
<td>13.75</td>
</tr>
<tr>
<td>Play Skills</td>
<td>11.31</td>
</tr>
<tr>
<td>Conversation Skills</td>
<td>8.81</td>
</tr>
<tr>
<td>Rating Scale</td>
<td>5.88</td>
</tr>
<tr>
<td>Emotion/Empathy</td>
<td>4.31</td>
</tr>
<tr>
<td>Social Skills</td>
<td>3.30</td>
</tr>
<tr>
<td>Academic Behaviors</td>
<td>3.21</td>
</tr>
<tr>
<td>Negative behavior</td>
<td>1.21</td>
</tr>
<tr>
<td>Negative Interaction</td>
<td>1.19</td>
</tr>
<tr>
<td>Neutral Interaction</td>
<td>0.09</td>
</tr>
</tbody>
</table>
## Results: SMD by Study Quality

<table>
<thead>
<tr>
<th>Study Quality</th>
<th>Effect Size Mean</th>
<th># of Studies Included</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.43</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>5.07</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>6.06</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>10.40</td>
<td>24</td>
</tr>
</tbody>
</table>
Advantages of SMD

• Able to compare treatment effectiveness across studies with effect size calculation
• More objective than visual analysis
• Don’t need level training that visual analysis requires
Disadvantages of SMD

- Effect size estimator not entirely accurate
- No correction for small sample size
- Could not include certain articles due to basal/ceiling responding
- Excessive amount of time needed to extract every data point in every graph
  - 550 total graphs (from 73 studies)
  - Each graph takes about 20-25 min. to extract data and calculate SMD
  - Total time about 230 hours…or 9 ½ days
- Have to Purchase UnGraph
  - Only 1 license per purchase
Future Directions

• Continue to analyze single subject data using further developed calculations
• Journals could require reporting of effect sizes or submission of raw data
Second Analysis

- Nonparametric Method
- Percentage of All Non-overlapping Data (PAND)
- Phi Coefficient
Percent All Non-overlapping Data (Parker, Hagan-Burke, & Vannest, 2007)

- School psychology requires more reliable ways to determine the efficacy of an intervention, especially with single case research
- PAND: statistic designed for single case research which calculates data non-overlap and an accompanying effect size
Calculating PAND
(Parker, Hagan-Burke, & Vannest, 2007)

• Count the smallest number of data points from either phase which, if eliminated, would create completely non-overlapping data between the two phases

• Multiple solutions may exist
An Example
(Parker & Vannest, 2009)

• Find the least number of points whose removal would create non-overlapping data
• Two points are identified
An Example
(Parker & Vannest, 2009)

- PAND: all *remaining* non-overlapping points divided by the total number of points
  - PAND: 19/21 = 90%
Calculating PAND
(Parker & Vannest, 2009)

• PAND differs from Percent Non-overlapping Data (PND)
  – Uses all data points across both phases
  – Not as influenced by outliers
PAND Analyses

• Articles:
  – Included for analyses: 73

• Inter-observer agreement (IOA)
  – IOA calculated for 25% of articles
  – Average IOA = 84%
Converting PAND
(Parker, Hagan-Burke, & Vannest, 2007)

- PAND can be converted to a phi coefficient
  - \( \phi = (PAND \times 2) - 1 \)
  - Ranging from -1 to +1
  - Values equivalent to Pearson’s \( r \)
  - Interpreting \( \phi \) (Cohen 1988, 1992)
    - Strong = .50
    - Moderate = .30
    - Weak = .10
Converting *phi*  
(Parker, Hagan-Burke, & Vannest, 2007)

- Pearson’s *phi* can be converted to Cohen’s *d*  

\[
d = \frac{2\Phi}{\sqrt{1 - \Phi^2}}.
\]

- Effect size measure
- Interpreting *d* (Cohen 1988)
  - Strong = .80
  - Moderate = .50
  - Weak = .20
Method

Baseline

Intervention

Generalization

Brian

Gaven

Number of Sessions
Method

- Baseline
- Intervention
- Generalization

Time graph showing changes over sessions for Brian and Gaven.
Method
PAND Results

• Overall effect: \( \phi = 0.80; \ d = 2.67 \)

• Analyzed effects by:
  – IV type
  – DV type
  – Study quality (1-4 rating)
    • Three replications
    • Treatment integrity
    • DV operationally defined
    • IOA meets criteria
## PAND Results IV

<table>
<thead>
<tr>
<th>IV Category</th>
<th>Overall $\phi$</th>
<th>Overall $d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Based (n = 4)</td>
<td>0.89</td>
<td>3.87</td>
</tr>
<tr>
<td>Video Modeling (n = 9)</td>
<td>0.88</td>
<td>3.67</td>
</tr>
<tr>
<td>Peer Mediated (n = 15)</td>
<td>0.87</td>
<td>3.31</td>
</tr>
<tr>
<td>Unmanualized SST (n=17)</td>
<td>0.87</td>
<td>3.31</td>
</tr>
<tr>
<td>Social Stories (n = 8)</td>
<td>0.79</td>
<td>2.55</td>
</tr>
<tr>
<td>Reinforcement (n = 4)</td>
<td>0.74</td>
<td>2.21</td>
</tr>
<tr>
<td>Self-Monitoring (n = 8)</td>
<td>0.72</td>
<td>2.09</td>
</tr>
<tr>
<td>Manualized SST (n = 16)</td>
<td>0.69</td>
<td>1.92</td>
</tr>
<tr>
<td>DV Category</td>
<td>Overall $\phi$</td>
<td>Overall $d$</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>----------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Play Skills ($n = 18$)</td>
<td>0.88</td>
<td>3.67</td>
</tr>
<tr>
<td>Rating Scale ($n = 6$)</td>
<td>0.87</td>
<td>3.31</td>
</tr>
<tr>
<td>Adaptive/Functional Skills ($n = 6$)</td>
<td>0.86</td>
<td>3.37</td>
</tr>
<tr>
<td>Conversation Skills ($n = 28$)</td>
<td>0.85</td>
<td>3.21</td>
</tr>
<tr>
<td>Initiation ($n = 13$)</td>
<td>0.84</td>
<td>3.11</td>
</tr>
<tr>
<td>Social Skills ($n = 5$)</td>
<td>0.82</td>
<td>2.88</td>
</tr>
<tr>
<td>Positive Interaction ($n = 25$)</td>
<td>0.80</td>
<td>2.67</td>
</tr>
<tr>
<td>Academic Behavior ($n = 14$)</td>
<td>0.79</td>
<td>2.55</td>
</tr>
<tr>
<td>Emotion/Empathy ($n = 4$)</td>
<td>0.78</td>
<td>2.52</td>
</tr>
<tr>
<td>Negative Behavior ($n = 9$)</td>
<td>0.71</td>
<td>2.00</td>
</tr>
<tr>
<td>Negative Interaction ($n = 9$)</td>
<td>0.63</td>
<td>1.64</td>
</tr>
<tr>
<td>Neutral Interaction ($n = 3$)</td>
<td>0.60</td>
<td>1.50</td>
</tr>
</tbody>
</table>
## PAND Results – Study Quality

<table>
<thead>
<tr>
<th>Quality Category</th>
<th>Overall $\phi$</th>
<th>Overall $d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (n= 2)</td>
<td>0.85</td>
<td>3.21</td>
</tr>
<tr>
<td>2 (n= 7)</td>
<td>0.79</td>
<td>2.55</td>
</tr>
<tr>
<td>3 (n= 29)</td>
<td>0.8</td>
<td>2.67</td>
</tr>
<tr>
<td>4 (n= 35)</td>
<td>0.8</td>
<td>2.67</td>
</tr>
</tbody>
</table>
Advantages of PAND
(Parker, Hagan-Burke, & Vannest, 2007; Parker & Vannest, 2009)

• Uses all data points across both phases
• Not as influenced by outliers as ANOVA procedures or PND
• PAND gives both an ES and an exact measure of data non-overlap
  – Useful when no effect size is reported
• May be translated to $\phi$ or $d$ to determine effect
Disadvantages of PAND
(Parker, Hagan-Burke, & Vannest, 2007; Parker & Vannest, 2009)

- Low statistical power for small studies
  - Not recommended with less than 20 data points
- Possible human error in hand calculations from graphs
  - Crowded or complex graphs
- Only measures simple mean level shifts
  - Does not control for positive baseline trend
  - Large effect size alone does not mean change was due to the intervention
Disadvantages of PAND
(Parker, Hagan-Burke, & Vannest, 2007; Parker & Vannest, 2009)

• Insensitivity at upper end of the scale
  – When there is no data overlap between two phases, PAND is 100%, regardless of the distance between the data

• PAND measures overlap only
  – Lacks sensitivity of an ANOVA procedure in measuring distance (i.e. degree of change)
Third Analysis

• Visual Analysis
Visual Analysis

• Systematic interpretation of treatment programs through visual inspection of graphed data (Cooper, Heron, & Heward, 2007)
  – Can be done between and within treatment phases
  – No formalized rules, but some common criteria
Visual Analysis Criteria

• Each phase of treatment rated on 7 criteria:
  (WWC, 2010; Cooper et al., 2007)
  – Level
  – Trend
  – Variability
  – Immediacy of Effect
  – Overlap
  – Consistency
  – Number of Data Points
Visual Analysis Scoring

• All phases received either a “0” or a “1” score for each criterion.
  – Phase scores ranged from 0-7

• Phase scores were averaged to compute effects by DV, IV, and study quality
Level

• The mean score of data points within a phase (WWC, 2010)
  – 0 points if the mean score did not change in the desired direction from baseline to treatment
  – 1 point if the mean score changed in the desired direction from baseline to treatment
Trend

- The overall direction of the data path (Cooper et al., 2007)
  - 0 points for no trend, a trend identical to baseline, or a trend in the wrong direction
  - 1 point for a trend in the desired direction
Variability

- The level of deviation of data points around the best-fitting line (WWC, 2010)
  - 0 points for high variability
  - 1 point for low variability
Immediacy of Effect

- The magnitude of change (level, trend, or variability) between phases (Cooper et al., 2007)
  - 0 points for absent or delayed effects
  - 1 point for rapid changes in level, trend, or variability
Overlap

- The degree to which data points in intervention overlap with data points in baseline (WWC, 2010)
  - 0 points for considerable overlap
  - 1 point for no or low percentage of overlap
Consistency

- The similarity of data patterns from phases within the same conditions (WWC, 2010)
  - 0 points for low consistency
  - 1 point for high consistency
Number of Data Points

- Number of data points within phase (Cooper et al., 2007)
  - 0 points for less than 3 data points
  - 1 point for 3 or more data points
Visual Analysis

• Articles:
  – Included for analysis: 73 studies

• Inter-observer agreement (IOA)
  – Calculated on 22 studies (30%)
  – Average IOA: $r = 86.83\%$
Visual Analysis Results

• Average score for all studies = 4.32
• Averages by criterion

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>0.82</td>
</tr>
<tr>
<td>Trend</td>
<td>0.39</td>
</tr>
<tr>
<td>Variability</td>
<td>0.18</td>
</tr>
<tr>
<td>Immediacy of Effect</td>
<td>0.65</td>
</tr>
<tr>
<td>Overlap</td>
<td>0.58</td>
</tr>
<tr>
<td>Consistency</td>
<td>0.58</td>
</tr>
<tr>
<td>Number of Data Points</td>
<td>0.86</td>
</tr>
</tbody>
</table>
## Independent Variable

<table>
<thead>
<tr>
<th>Treatment Type</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmanualized SST ($n = 17$)</td>
<td>4.74</td>
</tr>
<tr>
<td>Video Modeling ($n = 9$)</td>
<td>4.53</td>
</tr>
<tr>
<td>Computer-Based Program ($n = 4$)</td>
<td>4.43</td>
</tr>
<tr>
<td>Peer-Mediated Training ($n = 15$)</td>
<td>4.41</td>
</tr>
<tr>
<td>Manualized SST ($n = 16$)</td>
<td>4.23</td>
</tr>
<tr>
<td>Social Stories ($n = 8$)</td>
<td>4.11</td>
</tr>
<tr>
<td>Self-Monitoring ($n = 7$)</td>
<td>3.87</td>
</tr>
<tr>
<td>Reinforcement-Based Training ($n = 4$)</td>
<td>3.78</td>
</tr>
</tbody>
</table>
## Dependent Variable

<table>
<thead>
<tr>
<th>DV Category</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Play Skills (n = 18)</td>
<td>4.99</td>
</tr>
<tr>
<td>Negative Behavior (n = 9)</td>
<td>4.91</td>
</tr>
<tr>
<td>Adaptive/Functional Skills (n = 6)</td>
<td>4.63</td>
</tr>
<tr>
<td>Positive Interaction (n = 26)</td>
<td>4.62</td>
</tr>
<tr>
<td>Social Skills (n = 5)</td>
<td>4.60</td>
</tr>
<tr>
<td>Rating Scales (n = 6)</td>
<td>4.54</td>
</tr>
<tr>
<td>Initiations (n = 13)</td>
<td>4.38</td>
</tr>
<tr>
<td>Conversation Skills (n = 28)</td>
<td>4.37</td>
</tr>
<tr>
<td>Academic Behavior (n = 14)</td>
<td>4.32</td>
</tr>
<tr>
<td>Emotion/Empathy (n = 4)</td>
<td>3.79</td>
</tr>
<tr>
<td>Negative Interaction (n = 9)</td>
<td>3.64</td>
</tr>
<tr>
<td>Neutral Interaction (n = 3)</td>
<td>2.5</td>
</tr>
</tbody>
</table>
## Study Quality

<table>
<thead>
<tr>
<th>Study Quality</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ((n = 2))</td>
<td>3.25</td>
</tr>
<tr>
<td>2 ((n = 7))</td>
<td>4.55</td>
</tr>
<tr>
<td>3 ((n = 29))</td>
<td>4.38</td>
</tr>
<tr>
<td>4 ((n = 35))</td>
<td>4.29</td>
</tr>
</tbody>
</table>
Advantages of Visual Analysis

• Less complex calculations
  – Averages rather than SMD and Phi

• Common practice
  – Traditionally accepted as the “gold standard” of single-case data interpretation

• Greater study inclusion
  – No need to exclude studies with zero baseline
Disadvantages of Visual Analysis

- **Subjective**
  - Not all graphs are easy to interpret
  - IOA may suffer as a result

- **Not Standardized**
  - Difficult to categorize level of effectiveness
  - Cannot compare directly to other measures of effect size

- **All criteria equally weighted**
  - Some criteria may be more relevant than others
Combined Results - Correlations

- SMD and VA correlate at $r = .39$
- SMD and Phi correlate at $r = .40$
- Phi and VA correlate at $r = .81$

<table>
<thead>
<tr>
<th></th>
<th>SMD Pearson Correlation</th>
<th>VA Pearson Correlation</th>
<th>Phi Pearson Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMD</td>
<td>1</td>
<td>.394</td>
<td>.400</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.005</td>
<td>.004</td>
</tr>
<tr>
<td>VA</td>
<td>.394</td>
<td>1</td>
<td>.810**</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.005</td>
<td>.000</td>
</tr>
<tr>
<td>Phi</td>
<td>.400</td>
<td>.810**</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.004</td>
<td>.000</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).  
a. Listwise N=49
## Combined Results - Treatment

<table>
<thead>
<tr>
<th>SMD</th>
<th>PAND</th>
<th>VA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manualized SST</td>
<td>Computer-Based</td>
<td>Unmanualized SST</td>
</tr>
<tr>
<td>Video Modeling</td>
<td>Video Modeling</td>
<td>Video Modeling</td>
</tr>
<tr>
<td>Peer-Mediated</td>
<td>Peer-Mediated</td>
<td>Computer-Based</td>
</tr>
<tr>
<td>Unmanualized SST</td>
<td>Unmanualized SST</td>
<td>Peer-Mediated</td>
</tr>
<tr>
<td>Self-Monitoring</td>
<td>Social Stories</td>
<td>Manualized SST</td>
</tr>
<tr>
<td>Computer-Based</td>
<td>Reinforcement-Based</td>
<td>Social Stories</td>
</tr>
<tr>
<td>Social Stories</td>
<td>Self-Monitoring</td>
<td>Self-Monitoring</td>
</tr>
<tr>
<td>Reinforcement-Based</td>
<td>Manualized SST</td>
<td>Reinforcement-Based</td>
</tr>
</tbody>
</table>
## Combined Results

### DV

<table>
<thead>
<tr>
<th>SMD</th>
<th>PAND</th>
<th>VA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initiation</td>
<td>Play Skills</td>
<td>Play Skills</td>
</tr>
<tr>
<td>Adaptive Behavior</td>
<td>Rating Scales</td>
<td>Negative Behavior</td>
</tr>
<tr>
<td>Positive Interaction</td>
<td>Adaptive Behavior</td>
<td>Adaptive Behavior</td>
</tr>
<tr>
<td>Play Skills</td>
<td>Conversation Skills</td>
<td>Positive Interaction</td>
</tr>
<tr>
<td>Conversation Skills</td>
<td>Initiation</td>
<td>Rating Scales</td>
</tr>
<tr>
<td>Rating Scales</td>
<td>Social Skills</td>
<td>Social Skills</td>
</tr>
<tr>
<td>Emotion/Empathy</td>
<td>Positive Interaction</td>
<td>Initiation</td>
</tr>
<tr>
<td>Social Skills</td>
<td>Academic Behaviors</td>
<td>Conversation Skills</td>
</tr>
<tr>
<td>Academic Behaviors</td>
<td>Emotion/Empathy</td>
<td>Academic Behaviors</td>
</tr>
<tr>
<td>Negative Behavior</td>
<td>Negative Behavior</td>
<td>Emotion/Empathy</td>
</tr>
<tr>
<td>Negative Interaction</td>
<td>Negative Interaction</td>
<td>Negative Interaction</td>
</tr>
<tr>
<td>Neutral Interaction</td>
<td>Neutral Interaction</td>
<td>Neutral Interaction</td>
</tr>
</tbody>
</table>
# Combined Results – Study Quality

<table>
<thead>
<tr>
<th>SMD</th>
<th>PAND</th>
<th>VA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Combined Results and Implications

• PAND and Visual Analysis correlated the highest
• Even so, every analysis found differing results for IV, DV, and Study Quality analyses
  – Implications for treatment
• More research is needed to determine which method is the most accurate
  – Method used changes obtained results
  – Disadvantages of certain methods (i.e. time requirement for SMD calculation) may be considered differently depending on method accuracy
Limitations

- A regression model or multilevel model was not used, thus those methods could not be compared
- File-drawer problem: unpublished studies were not included, which may increase the effect
- Direct measurement of behavior is subject to drastic increases or decreases, inflating effects
- Inability to compare all graphs across all methods due to measurement limitations
- Calculations subject to human error
- Currently, no way to determine which method most accurately represents the effect of treatment so comparison of results is difficult
- Additionally, many behaviors studied were not investigated in group design, rendering it inaccurate to compare these results to group design methods
Future Directions (Meta-Analysis)

• Include parametric measures
  – Someone design a feasible multilevel model program for SPSS

• Determine how to estimate and compare accuracy of various methods

• Empirically validate criteria of evaluating effects for various methods

• Develop standardized, comparable (to group design) methodologies for the synthesis of single-subject studies
Future Directions (Researchers/Practitioners)

• Report all relevant demographic information
• Operationally define all relevant dependent variables
• Advocate for and use only articles that use sound methodology
• Take and report treatment integrity measures
Selected References


• Questions or comments?
Email Katherine Hunter at kathunter110@gmail.com