Quality Indicators for Single Case Designs

Rob Horner, Charlie Greenwood

www.uoecs.org

Specific Goals

- Define an emerging role for single case designs
  - Defining features of Single-case research
  - Programs of research
  - Grant applications

- Review standards for visual analysis of single case studies
  - WWC Design standards
  - WWC Evidence criteria
  - Training protocol for new scholars

- Introduce Repeated Acquisition Designs

Context

- Single-case methods developed and used within Applied Behavior Analysis

Shavelson & Towne, 2002
Emergence of "Evidence-based" practices

Claims that Visual Analysis is Unreliable

IES commitment to Rigorous Science in Education
A Moment in Time

- We have a window of time to define the contribution of Single-case research

**What is Needed?**

- Document how Single-case methods constitute a rigorous form of science
- Improve the training provided to new scholars
- Improve the review standards used for journal and grant applications (visual and statistical analysis)
- Define a standard for documenting evidence-based practices using single-case research methods.

Current Opportunities

- IES White Paper
- **What Works Clearinghouse**
  - Protocol for using Single-case Research in the identification of “promising” and “evidence-based” practices
- **Real Need:** Agreement within the Single-case research community about standards for design and analysis.

Design Examples

- Reversal/Withdrawal Designs
- Multiple Baseline Designs
- Alternating Treatment Designs
- Others:
  - Changing Criterion
  - Repeated Acquisition
  - Multiple Probe
The roles for single-case research in IES Goal Structure.

- Useful in the iterative development of interventions.
- Useful in design of any interventions focused on the “individual participant” (especially with low-incidence populations).
- Useful for pilot research to assess the effect size needed for RCTs.
- Useful for fine-grained analysis of “weak and non-responders”

Context

- Designing interventions

<table>
<thead>
<tr>
<th>Goal 2: Development</th>
<th>Goal 3: Efficacy</th>
<th>Goal 4: Effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop Intervention</td>
<td>Randomized Control Trials</td>
<td>Analysis of Weak and Non-Responders with Single-case Designs</td>
</tr>
</tbody>
</table>

Evaluation of Single-case Studies

- What Works Clearinghouse Standards
  - Design Standards
  - Evidence Criteria
  - Social Validity
Design Evaluation

**Meets Standard**
- IV manipulated directly
- Measure is reliable (IOA = .80 percent agreement; .60 Kappa)
- 20% of data points in each phase
- Design allows opportunity to assess basic effect at three different points in time. Controls for threat to Int Validity.
- Five data points per phase (or design equivalent)
- ATD (four comparison option)

**Meets with Reservation**
- All of above, except at least three data points per phase
- Non-concurrent Multiple Baseline

**Does not Meet Standard**
When Assessing **Design Standard**

- Does the design allow for the opportunity to assess experimental control?
  - Baseline
  - At least five data points per phase (3 w/reservation)
  - Opportunity to document at least 3 basic effects, each at a different point in time.

---

**Example Reversal Design**

1. Baseline
2. Each phase has at least 5 data points (3 w/reservation)
3. Design allows for assessment of "basic effect" at three different points in time.

---

**Does Not Meet Standard**

---
First Demonstration of Basic Effect

Second Demonstration of Basic Effect

Third Demonstration of Basic Effect

Meets Standard

Does Not Meet Standard

Meets Standard
Review Fisher studies:

To what extent does each graph allow documentation of experimental control?
Evaluate the Design

- Meets Design Standards
- Meets with Reservations
- Does Not Meet Design Standards

Effect Size Estimation
Social Validity Assessment

Evaluate the Evidence

- Strong Evidence
- Moderate Evidence
- No Evidence

Evidence Criteria

**Strong**

- Baseline
  - Documentation of research question “problem”
  - Documentation of predictable pattern (>5 data points)
- Each Phase of the Analysis
  - Documentation of predictable pattern (>5 data points)
- Basic effects
  - Documentation of predicted change in the DV when IV is manipulated
- Experimental Control
  - Three demonstrations of basic effect, each at a different point in time.
  - No demonstrations of intervention failure

Evidence Criteria

**Moderate**

- All of “Strong” criteria, with these exceptions:
  - Only 3-4 data points per phase
  - Three demonstrations of effect, but with additional demonstrations of failure-to-document effect.
  - Non-concurrent multiple baseline

No Evidence

- Misnomer
- Evidence does not meet Moderate level.
Visual Analysis of Single Case Designs

**Baseline**
- Document the "problem" requiring intervention
- Typically 5 or more data points

**Each Phase**
- Documents a clear pattern of responding
- Typically 5 or more data points

**Adjacent phases**
- Do data document a "basic effect"

**Whole study**
- Do the phases document experimental control (e.g. at least three demonstrations of a basic effect, each at a different point in time.

---

Assessing within phase "pattern" and between phase "basic effect"

- **Level**
- **Trend**
- **Variability**
- **Overlap**
- **Immediacy of Effect**
- **Consistency across similar phases**

Other: vertical analysis; intercept gap

---

Visual Analysis of Single Case Designs

- Change in Level
- Change in Trend
- Change in Variability
- Immediacy of Effect
- Overlap
- Consistency of Data Patterns across similar Phases

Parsonson & Baer, 1978; Kratochwill & Levin, 1992
Multiple Baseline Design

- Level, Trend, Variability, Overlap
- Immediacy of Effect
- Consistency across similar phases
- Stability in non-intervened series when effect demonstrated in one series
Magnitude of separation
- Greater the difference between two conditions, larger the demonstration of a functional relation

Consistency of separation
- Greater consistency of separation between two conditions (no overlap) larger the demonstration of a functional relation

Number of data points used to establish experimental control
- At least 4 comparisons. The more points documenting separation, the stronger the demonstration of experimental control.

Alternating Treatment/ Multi-element designs

Alice

Percentage of Trials with Scream, Push

FA sessions
Implications for Construction of Research Questions in Single-case Research

Increase the precision of Research Questions
- Define conceptual logic for research question
- Define research question with greater precision
- IV related to change in level, trend, variability?

“Is there a functional relation between Functional Communication Training and reduction in the level and variability of problem behavior?”

Visual Analysis Activity

Development of visual analysis graphs
- Bruce Wampold, Richard Freund, Rick Allen, Tom Kratochwill, Chris Swoboda
- Level, trend, variability, overlap
- Purpose is to emphasize interaction

Examine and score each graph
- Use ALL data in the graph
- Combine assessment of: Level, trend, variability, overlap, immediacy of effect, consistency of data patterns in similar phases.
- Use scoring metric:
  - 0 = no functional relation
  - 5 = publishable
  - 7 = strong functional relation

Compare scores with Excel file
- Median scores from five “expert” Single-case Researchers

Reversal Design

Is there a functional relation between the Intervention and reduction in the level of the problem behavior?

<table>
<thead>
<tr>
<th></th>
<th>No Exp Control</th>
<th>Publishable</th>
<th>Strong Exp Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>
ABAB 1

ABAB 2

ABAB 3
Multiple Baseline

Is there a functional relation between introducing the intervention and reduction in the level of the problem behavior?
Alternating Treatment Designs

Is there a difference in the percentage of 10 s intervals with social initiation under Condition B compared with Condition C?

ATD 1

ATD 2

Level of Experimental Control

No Exp Control 1 2 3 4
Publishable 5 6
Strong Exp Control 7
Five studies documenting experimental control

Conducted in at least three different locations by at least three different researchers

Across at least 20 different participants

Each study demonstrates an effect size of \( d \geq 0.50 \)??
Example of using Single-Case Research to document Evidence-based Practice

**Functional Communication Training** to reduce the level of problem behavior in school, home and community

- **At least Five Studies**
  - Bird, Dones, Moniz & Robinson (1989)
  - Carr & Durand (1986)
  - Durand & Carr (1987)
  - Wacker et al., (1990)

- **At least three settings /scholars**
- **Six locations, Research Groups**
- **At least 20 participants**
- 42 different participants document effect

Disseminating Evidence-based Practices

- **Evidence-based is not enough**
  - In addition to the features of the practice: define what outcomes, when/where used, by whom, with what target populations, at what fidelity?
  - The innovative practice needs to not only be evidence-based, but dramatically easier and better than what is already being used.
  - The practice should be defined conceptually as well as procedurally, to allow guidance for adaptation.

Statistical Analyses of Single-case Research

- **Documenting Experimental Control/ Statistical Significance:**
  - No statistical model currently replicates the conceptual logic used in visual analysis.
  - Consider statistical analysis of SCD as supplemental

- **Effect Size**
  - Not a variable traditionally addressed in Single-case Analyses.
  - It needs to be.
Calculating Effect Size within Single-Case Research

Core criteria for an effect size measure used with single-case research:

- Comparable to Cohen’s d
  - So results can be used in meta-analyses
- Assesses the “experimental effect” under analysis (e.g., examines ALL phases of a study, not just a AB comparison

Control for:
- Autocorrelation (serial dependency)
- Weighting of level, trend, variability, overlap, immediacy
- Outliers
- Type of data (e.g., counts, rates... from a non-normal distribution)

Statistical Analysis of SCD: Current Status

Emerging Approaches

- Percentage of non-overlapping data
  - Parker et al.,
  - Wolery et al.,
- Randomization Tests
  - Kratochwill & Levin
- Generalized Least Squares Regression
  - Swaminathan et al.,
  - SingSub
- Hierarchical Linear Modeling
  - Hedges, Shadish, & Rindskopf
  - d-Estimator

Statistical Analysis:

1. Done to confirm visual analysis
2. Done as part of meta-analyses

Generalized Least Squares Regression

Swaminathan et al., 2010

- Control for autocorrelation
- Establish regression line
- Build confidence intervals
- Compare observed with expected

SingSub

- Testing in progress
- Need adaptation for dealing with non-normal distributions.
Table 3: GLS Effect Sizes for Ross & Horner (2009), Target Students

<table>
<thead>
<tr>
<th>Student</th>
<th>Autocorrelation</th>
<th>MSE</th>
<th>Standardized Effect</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rob</td>
<td>-0.333</td>
<td>1.192</td>
<td>-0.961</td>
<td>-7.579</td>
<td>40</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Bruce</td>
<td>0.006</td>
<td>0.927</td>
<td>-8.663</td>
<td>-4.612</td>
<td>40</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Cindy</td>
<td>-0.254</td>
<td>0.951</td>
<td>-3.220</td>
<td>-5.044</td>
<td>42</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Scott</td>
<td>-0.272</td>
<td>0.736</td>
<td>-5.034</td>
<td>-8.308</td>
<td>43</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Anne</td>
<td>0.059</td>
<td>1.010</td>
<td>-4.450</td>
<td>-7.398</td>
<td>41</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Ken</td>
<td>-0.281</td>
<td>0.689</td>
<td>-3.900</td>
<td>-8.740</td>
<td>42</td>
<td>&lt;0.001*</td>
</tr>
</tbody>
</table>

**The d-Statistic**

Hedges, Shadish & Rindskopf

Here is the basic version of the d-estimator

\[
ES = \frac{D}{\sqrt{\frac{1}{S_y^2} + \frac{\sum_i S_i^2/m}{n(1-\rho)^2 - n(1-\rho^2) + 2\rho(1-\rho^2)}}}
\]

D-bar is the difference between baseline and treatment period means (averaged across people), \(S_y^2\) is the sample variance of the observations in the baseline period for the \(i\)th person, \(S_i^2\) is the sample variance between the baseline person means, \(\rho\) is the autocorrelation, \(n\) is the number of time points, \(m\) is the number of people. The parenthetical expression in the numerator is a correction for small sample bias.

**In Summary**

- Single case methods are an effective and efficient approach for documenting experimental effects.
- Need exists for more precise standards for training and using visual analysis, and combinations of visual analysis with statistical analysis.
  - Exp. Control: Three demonstrations of basic effect, each at a different point in time.
  - Research Questions: Define variable of expected effect (level, trend)
- We need agreement as a field on the standard(s) we will use to document evidence-based practices via single-case methods.
  - Define "Practices": Op Description, Who can use, For whom, with What outcome
  - Evidence base practices: 5 studies, 3 researcher/locations, \(>\) participants.
  - There are encouraging (but still emerging) approaches for statistical analysis that will improve meta-analysis options.