Moving from Single-Case to Conventional Group Intervention Designs, and Vice Versa

Joel R. Levin
University of Arizona
Moving from Single-Case to Conventional Group Intervention Designs, and Vice Versa

Argument: Through various randomization strategies, it is possible to design single-case intervention studies that possess the same or similar scientific credibility characteristics to those of conventional randomized group intervention studies.

With the additional inclusion of a sufficient number of replication components, who’s to say that a superbly implemented randomized single-case intervention study is less “valuable” than a superbly implemented conventional randomized group intervention study?
Let me state from the outset that I am an unapologetic, dyed-in-the-wool, tried-and-true conventional group intervention researcher who professes (to the extreme) the principles and practices of carefully controlled randomized experimental designs (i.e., “true” experiments, controlling for as many potential confounding variables as is superhumanly possible). Nonetheless, I have elected to play devil’s advocate in the “case” that follows…
Here’s an instrumentally relevant problem for you to ponder…

Two different intervention studies are proposed to an education research funding agency. Both studies are comparing the effectiveness of a promising mathematics instructional intervention with that of a promising reading instructional intervention.

A. A conventional group RCT study: 4 randomly selected first-grade low SES classrooms, randomly assigned to two instructional conditions (Math, Reading) for 8 weekly units of targeted-skill instruction; standardized math and reading outcome measures are assessed on a pretest and posttest using hierarchical linear modeling (HLM).
B. A single-case “RCT” study: 4 randomly selected low SES first-grade students, with each student to receive both math (A) and reading (B) weekly targeted-skill interventions in a randomized crossover design format, based on a randomly selected crossover point for each student (out of 5 “acceptable” ones), with standardized math and reading outcome-measure assessments taken every week for 15 weeks and analyzed by a valid randomization statistical test.
In addition, suppose it can be demonstrated that: (1) the major “intervention effect” research question is associated with respectable statistical power in each of the studies; but (2) only one of the studies can be funded.

Which one should be funded, the conventional group study or the single-case study?

On what basis?
The Rich Hypothesis (Which Ironically is the “Poor” Hypothesis)

In a 2013 *New York Times* article, Motoko Rich marshals research evidence – especially that coming from a study by economist Jonah Rockoff at Columbia University – to support the hypothesis that for lower-SES students, in particular, “teachers [and presumably curricular innovations] have bigger impacts on math test scores than on English test scores.”

Rationale

• Differing childhood reading and math background experience (in terms of time, quality, etc.) as a function of social class.

• In addition to the well-known finding that greater frequency and quality of early childhood reading opportunities occur in higher SES families (e.g., Hart & Risley, 1992), Rich notes that: “Reading also requires background knowledge of cultural, historical and social references. Math is a more universal language of equations and rules.”

“Your mother or father doesn’t come up and tuck you in at night and read you equations. But [some kinds of] parents do read kids bedtime stories, and [some kinds of] kids do engage in discussions around literacy, and [some kinds of] kids are exposed to literacy in all walks of life outside of school.”

*Parenthetical comments added by the presenter to bolster (or even bias?) the rationale for the hypothetical intervention study to follow.*
Discriminant Validity (Campbell & Fiske, 1959): An intervention differentially affects Outcomes X and Y. In particular, the intervention produces a desired effect on Outcome X but not on Outcome Y.

Selective Facilitation/Transfer-Appropriate Processing (Morris, Bransford, & Franks (1977): Interventions A and B differentially affect Outcomes X and Y. In particular, Intervention A produces a desired effect on Outcome X but not on Outcome Y; in contrast, Intervention B produces a desired effect on Outcome Y but not on Outcome X.


Components of the Critical Statistical Test

Comparative Intervention Effect on the Math Test (Y): Math Instruction – Reading Instruction should be positive and large: $(\text{Math} - \text{Reading})_Y > 0$

Comparative Intervention Effect on the Reading Test (X): Math Instruction – Reading Instruction should be negligible or only slightly negative: $(\text{Math} - \text{Reading})_X \approx 0$

Consequently, the instructional difference between Y and X should be positive and large:

$(\text{Math} - \text{Reading})_Y - (\text{Math} - \text{Reading})_X > 0$

These differences are calculated for each student at each of the potential intervention start points to form the basis of the subsequent randomization test.
Predicted Differential Effect of Two Instructional Interventions (Math and Reading) on Math and Reading Performance (Crossover Design Data Simplified for Ease of Interpretation)
Levin et al.’s (2014) Single-Case Randomized Intervention Start-Point, Random-Order Crossover Design

<table>
<thead>
<tr>
<th>Week</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student 1</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>Student 2</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Student 3</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Student 4</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
</tbody>
</table>

Note: Students are randomly assigned (in equal numbers) to either a Math-Reading (AB) or a Reading-Math (BA) intervention order. For each of the 15 weeks, assessments are made on both standardized reading and math measures. The crossover occurs on a student-by-student basis, starting on a week that is randomly selected from 5 “acceptable” ones (anywhere from the 6th through the 10th week).
Analysis to Be Conducted Via the ExPRT Randomization Test Package

A single-case crossover design application is combined with the Levin-Wampold (1999) “comparative intervention effectiveness” randomization test procedure in Gafurov and Levin’s (2015) ExPRT program to assess the hypothesis of primary interest (namely, the intervention by test type interaction).

─ the two outcome measures, reading (X) and math (Y) test scores, are entered as paired observations for each student; and

─ for each student at each time period, the paired outcomes are entered in their naturally occurring intervention order, either Math-Reading (AB) or Reading-Math (BA)
**ExPRT Setup for Levin & Wampold’s (1999) Replicated Simultaneous Start-Point Model**

<table>
<thead>
<tr>
<th>Week</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student 1&lt;sub&gt;X&lt;/sub&gt;</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>Student 1&lt;sub&gt;Y&lt;/sub&gt;</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>Student 2&lt;sub&gt;X&lt;/sub&gt;</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Student 2&lt;sub&gt;Y&lt;/sub&gt;</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Student 3&lt;sub&gt;X&lt;/sub&gt;</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Student 3&lt;sub&gt;Y&lt;/sub&gt;</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Student 4&lt;sub&gt;X&lt;/sub&gt;</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>Student 4&lt;sub&gt;Y&lt;/sub&gt;</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
</tbody>
</table>
Power curves for unrestricted and restricted AB crossover designs. The rejection rate for the comparative-effectiveness hypothesis is shown as a function $d$ and $N$, for a 15-observations design, 5 potential start points, and an autocorrelation of .30.
Power Analysis for the Comparative Intervention Effectiveness Randomization Test in a “True” Crossover Design

With 2 students randomly assigned to each of the two intervention crossover orders, 15 outcome observations, 5 potential intervention start points, and a posited autocorrelation of .30, a researcher would have a 73% chance of detecting a differential instructional intervention effect amounting to a Busk-Serlin “no assumptions” $d$ of 1.00 and a 93% chance of detecting an intervention effect of $d = 1.50$. [These powers would decrease with increases in the magnitude of the autocorrelation.]
<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</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>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Average: 18.206704  Count: 10  Sum: 182.06704
| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U |
| 1st pot | # pot | interv | stpnts | units | max | # pnts | act interv | stpnts | Data: 1:Original, 2:Standardized | Tails (1,2) | 1A>B or T1>T2, 2B>A or T2>T1 | Sig. | n pnts, ordr perms, tot output (yes/no) | 1:Mean, 2:Slope, 3:Variance(B/A), 4:Variance(B-A) | Missing Code | If Pairs Test: 1:Gen, 2:Comp, 3:Comp(Rnd XY) | Design: 1:Std, 2:Cross | If Std: 1:Fixed, 2:Rand | If Cross: 1:Cond, 2:Time | Actual Order | Run | 1st Then Plot |
| 2 | 6 | 5 | 4 | 15 | 8 | 1 | 0.05 | 1 | 2 Sig. p<0.001 | 625 | no | Rank = 1 of 3750 | Rank = 1 of 3750 | 1 | 2 | 2 | 2 | 1 | AB | Run | Plot |
| 3 | 6 | 5 | 7 | | | | | | | | | | | | | | | | | | |
| 4 | 6 | 5 | 10 | | | | | | | | | | | | | | | | | | |
| 5 | 6 | 5 | 6 | | | | | | | | | | | | | | | | | | |

*time elapsed - 1.300781 sec*
Not Just “Hypothetical” Anymore

Pictorial Mnemonic-Strategy Interventions for Children with Special Needs: Illustration of a Multiply Randomized Single-Case Crossover Design

Yooyeun Hwang, Joel R. Levin, Evan W. Johnson

Overview of the Study

- 8 fifth graders with cognitive disabilities
- Two learning tasks (inventions/dates and vocabulary acquisition), along with a filler task
- 28 items for each task, divided into seven 20-30 minute weekly-session lists of 4 items apiece
- Two different learning strategies for each task, mnemonic and “own preferred method” control
- Children used one type of strategy for one task and the other type of strategy for the other task for 3 or 4 weeks and then “crossed over” to the opposite strategies for 4 or 3 weeks
- With one of two experimenters, children administered the two learning tasks (in counterbalanced order), then the filler test, and then tested on the two focal lists
Design and Procedure Randomization

• Randomly assigning children (in equal numbers) to the two crossover design orders
• Randomly assigning a strategy crossover point to each child (following either Session 3 or 4)
• Constructing seven 4-item lists (randomly selected from 28 items) for both the inventions and vocabulary materials
• Randomly ordering the items within lists
• Randomly assigning lists to children
• Randomly assigning children (in equal numbers) to experimenters
• Randomly determining task study/test order within each session
Statistical Analysis Randomization

- Incorporating the design’s restricted randomization scheme into the randomization-test analyses
- Incorporating each child’s randomly selected intervention crossover point into the randomization-test analyses
- Incorporating the strategy-order factor into the focal “composite” randomization-test analysis
Embedding a Single-Case Intervention Study Within a “Group” Intervention Study

- Objective: Kratochwill et al.’s (2004) experimental investigation of a community-based program, Family and Schools Together (FAST), was designed to improve the behavioral and academic functioning of Native American students.

- Design: In seven different schools, between 10 and 18 students were matched on the basis of several behavioral characteristics and randomly assigned to program and nonprogram conditions (thereby creating between 5 and 9 pairs per school), with teachers ostensibly “blind” to conditions.

- One focal outcome measure of the study was on students’ arithmetic skills and so on different assessment occasions throughout the school year, all students were provided with several pages of arithmetic problems to solve.

Embedding a Single-Case Intervention Study Within a “Group” Intervention Study

• What if some small number of pairs (in this example, 7 pairs) were selected to participate in a single-case microexperiment to assess the differential effect of an instructional intervention on program and nonprogram students?
• Suppose that on a single assessment occasion arithmetic problems were randomly distributed over 8 pages (representing 8 outcome observations).
• Further suppose that students were required to complete the first 2 pages according to whatever basic arithmetic procedures they typically used.
• Then, following a randomly selected intervention point (out of, say, 5 designated ones), suppose that some specialized arithmetic-skill instruction were provided to both program and nonprogram students.
• Finally, students were required to complete the remaining pages of problems using the new procedures they had just been taught.
### Levin & Wampold’s (1999) Replicated Simultaneous Start-Point Model

<table>
<thead>
<tr>
<th>Time Period</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair 1X</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A*</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>Pair 1Y</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A*</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>Pair 2X</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A*</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>Pair 2Y</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A*</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
</tbody>
</table>

**Note:** Potential intervention start points are between Time Periods 5 and 17 inclusive.

*Randomly selected intervention start point for each pair of units

Bottom Line

Who’s to say which is worth more to science and society:

A pound of conventional group intervention randomized trials research?

OR

Sixteen ounces of unconventional single-case intervention randomized trials research?
A Few Ethical Issues...

Original thoughts on this topic*:

“We know a thing or two because we’ve seen a thing or two.”

*Actually plagiarized from J. K. Simmons’ Farmers’ Auto Insurance TV ad, just to provide a preview of an unforgivable ethical violation.
A Few Ethics Issues...

• “Random” assignment and “do overs”
• Outliers and nonresponders
  – *a priori* decisions and rules
  – scoring response protocols (“blind” scoring)
• Data “massaging”
• “Multiplicity,” directional tests, and changing “predictions”
• Response-guided decisions
  – “close but no cigar rationalizations”
  – the “multiple judges” solution, and problem
• Professional research/writing expectations and attributions
  – the “data thief” and credit where credit is due

…As Manifested in Single-Case Intervention Research

Design
- Participant randomization
- Intervention randomization
- Intervention start-point randomization

Procedural
- Research bias
- Adequate controls

Analysis
- Negative results → publication bias
- Selective reporting
- Multiple analyses

Scoring
- “Blind”/Masked