Journal of Modern Applied Statistical Methods

Volume 14 | Issue 1 Article 22

5-1-2015

Algorithms for Assessing Intervention Effects in Single-Case Studies

Chao-Ying Joanne Peng Dr.
Indiana University-Bloomington, peng@indiana.edu

Li-Ting Chen
Indiana University - Bloomington, litingc@unr.edu

Recommended Citation
Peng, Chao-Ying Joanne Dr. and Chen, Li-Ting (2015) "Algorithms for Assessing Intervention Effects in Single-Case Studies," *Journal of Modern Applied Statistical Methods*: Vol. 14 : Iss. 1 , Article 22.
DOI: 10.22237/jmasm/1430454060

Algorithms for Assessing Intervention Effects in Single-Case Studies

Cover Page Footnote
This research was supported in part by two Maris M. Proffitt and Mary Higgins Proffitt Endowment Grants of Indiana University awarded to C.-Y. J. Peng while L.-T. Chen served as a research assistant on these two grants.
Address correspondence to Joanne Peng, Department of Counseling and Educational Psychology, Indiana University, Bloomington, IN 47405. E-mail: peng@indiana.edu.

Journal of Modern Applied Statistical Methods
May 2015, Vol. 14, No. 1, 276-307.
Copyright © 2015 JMASM, Inc.
ISSN 1538 − 9472

JMASM Algorithms and Code
Algorithms for Assessing Intervention Effects in Single-Case Studies

Chao-Ying Joanne Peng
Indiana University at Bloomington
Bloomington, IN

Li-Ting Chen
Indiana University at Bloomington
Bloomington, IN

Free web-based resources or popular software to assess six data features recommended by the *What Works Clearinghouse: Procedures and Standards Handbook* (IES, 2013 February) to determine intervention effects in a single-case study (Lambert, Cartledge, Heward, & Lo, 2006) are demonstrated. Lambert et al. (2006) employed a reversal (or ABAB) design and visual inspection to investigate the effectiveness of the report-card treatment in reducing disruptive behaviors in students. In our demonstration, we assessed each of the six data features separately; then integrated six assessments into one comprehensive analysis of the intervention effect. A simple approach to the determination of intervention effects illustrates how researchers and practitioners can be empowered to interpret data comprehensively and formulate evidence-based conclusions logically from well-designed and well-executed single-case studies.

Keywords: algorithm, intervention effect, single-case studies, level, trend, variability, immediacy, overlap, effect size, Spearman rank correlation, Page test, confidence interval

Introduction

Horner et al. (2005) defined a single-case design (SCD) as a "rigorous, scientific methodology used to define basic principles of behavior and establish evidence-based practice." (p. 165). SCDs are particularly important to clinical studies in which detailed information about aspects of a few participants' behavior is gathered over an extended period of time in order to determine effects of an intervention. Yet determining intervention effects in SCD studies presents unique challenges due to the small sample size, the correlated nature of outcome measures, and the difficulty of applying statistical methods to SCD data. Visual inspection has been traditionally used by researchers and practitioners to assess an

Dr. Peng is an Adjunct Professor of Statistics. Email her at peng@indiana.edu. Dr. Chen is a Research Associate. Email her at litchen@indiana.edu.

intervention effect. Indeed, according to the Institute of Education Sciences' publication, *What Works Clearinghouse: Procedures and Standards Handbook* (IES, 2013 February, hereafter abbreviated as the *WWC Handbook*), "Single-case researchers traditionally have relied on visual analysis of the data to determine (a) whether evidence of a relation between an independent variable and an outcome variable exists, and (b) the strength or magnitude of that relation." (p. E.5).

The subjectivity associated with visual analysis and its lack of a theoretical framework for testing a scientific hypothesis have hampered the generalizability of SCD findings. The *WWC Handbook* actually recommends the examination of six data features both within and between phases in order to determine the effectiveness of an intervention effect. The six data features include: level/level change, trend, variability, immediacy of the effect, overlap, and consistency of data in similar phases. These six features should be assessed collectively to determine if (1) the observed pattern of data in the intervention phase is indeed due to the intervention effects and (2) the observed pattern of data in the intervention phase is different from the predicted pattern of data, predicated from data collected in the baseline phase. The *WWC Handbook* further recommends that a measure of the strength of the relation between an independent variable and an outcome be computed and reported to accompany the assessment of that relation.

Given the importance of the WWC's initiative "to be a central and trusted source of scientific evidence for what works in education." (IES, 2013 February, p. 1) and the intended purpose of the *WWC Handbook* to provide "a detailed description of the standards and procedures of the WWC" (IES, 2013 February, p. 2), it is imperative that researchers and practitioners be empowered to evaluate evidence of intervention effects in any SCD study according the WWC standards and recommendations. In this paper, we demonstrate how to assess each of these six features in a real world data set (Lambert et al., 2006). In our demonstration, we assessed each of the six data features separately first. We subsequently integrated six assessments into one comprehensive analysis of the intervention effect. These assessments were conducted using free or commercially available software. The computing algorithms for these assessments appear in Appendices A to C. We conclude this paper by discussing relative advantages of our simple and straightforward approach, compared to visual analysis or complex statistical modeling and methods.

The Lambert data set

The Lambert data set was first reported and analyzed in *Journal of Positive Behavior Interventions* by Lambert et al. (2006). In Lambert et al. (2006)'s study, nine students from two classrooms were observed in baseline (the single-student responding or SSR) phase and the treatment (the response card or RC) phase for their disruptive behaviors during the teacher' instruction. A disruptive behavior, such as engaging in a conversation, provoking others, laughing or touching others, was recorded in 10 intervals of a study session (p. 89 of Lambert et al., 2006). The study employed a reversal (or an ABAB) design with two baseline phases (SSR1 and SSR2), each followed by an intervention phase (RC1 or RC2). The number of intervals in which a disruptive behavior was recorded was the outcome or the dependent measure. Figure 1 presents the findings reproduced from pp. 93-94 of the Lambert et al. (2006) article with permission. Using visual analyses, Lambert et al. (2006) concluded that the use of report cards was successful in decreasing disruptive behaviors for these nine students.

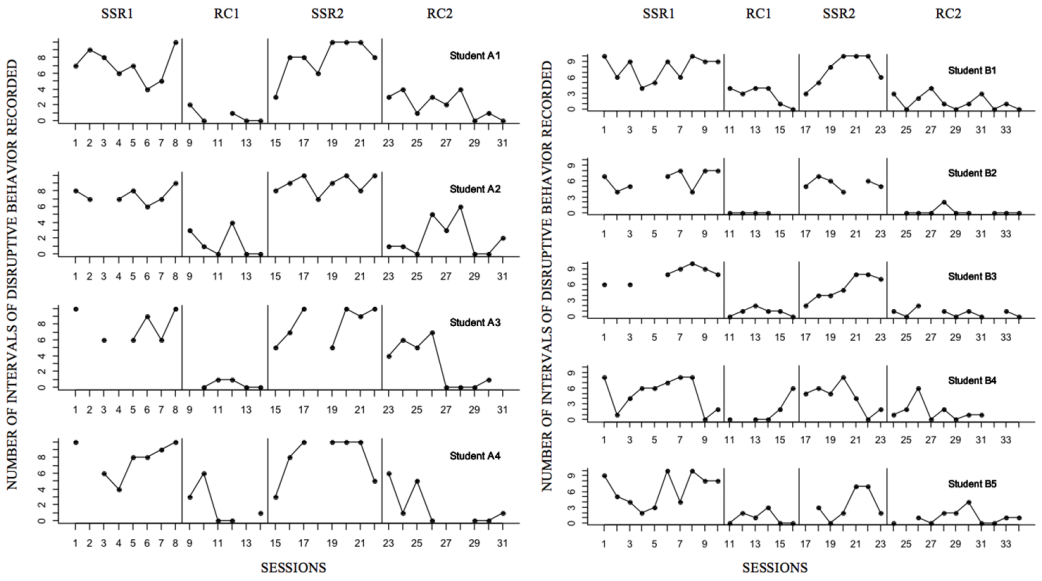

Figure 1. Number of intervals of disruptive behaviors during single-student responding (SSR) and response card (RC) condition. Adapted from "Effects of Response Cards on Disruptive Behavior and Academic Responding During Math Lessons by Fourth-Grade Urban Students," by Lambert et al., 2006, *Journal of Positive Behavior Interventions*, *8*, pp. 93-94, Copyright 2006 by Sage Publications. Used with permission.

Notice that there are breaks in Figure 1 due to student absences (p. 93 of Lambert et al., 2006). These breaks were ignored in the reanalysis of this data set by the special issue of *Journal of School Psychology* (Shadish, 2014). In this paper, we treat these breaks as missing data in order to retain the initial structure of this data set. Because there were different numbers of sessions implemented in the two baseline phases (SSR1 and SSR2) and the intervention phases (RC1 and RC2) in Classrooms A and B, we decided to analyze the two classroom data sets separately. Data collected from four students (A1 to A4) in Classroom A are hereafter referred to as the Lambert-A data set. B1 to B5 students' data from Classroom B are referred to as the Lambert-B data set. Both Lambert-A and -B data sets were systematically analyzed using SAS (Appendix A), a free web-based calculator (Appendix B; Vannest, Parker, & Gonen, 2011), and SPSS (Appendix C).

Assessment of level/level change

The *WWC Handbook* defines "level" as the mean score for data within a phase (2013, p. E.6). A level change between phases therefore indicates a change in the outcome measure due to the intervention. To assess the level and level change, we applied six paired-samples *t*-tests to means obtained from adjacent phases in Lambert-A and -B data sets (Table 1). The SAS computing codes for assessing levels and level changes are shown in Part A of Appendix A. The *t*-statistics and their corresponding *p*-values were further verified by two free web-sites located at http://www.statdistributions.com/chisquare/ and http://www.graphpad.com/quickcalcs/ttest1.cfm, respectively.

According to Table 1 results, the three paired-samples *t*-tests for Lambert-A data ranged from 18.57 to −16.99 with $df = 3$ (or 4−1). For Lambert-B data, the three paired-samples *t*-tests ranged from 8.52 to −6.70 with $df = 4$ (or 5−1). All six paired-samples *t*-tests were statistically significant at $\alpha = .05$ (one-tailed), suggesting that there was a level change between phases for both data sets. And the level changes supported the effectiveness of the intervention, namely, the report card treatment.

Table 1. Means, *SD*s, *t*-tests of differences between phases in Lambert-A and –B data sets

	SSR1-RC1		RC1-SSR2		SSR2-RC2	
	Set A	**Set B**	**Set A**	**Set B**	**Set A**	**Set B**
Mean[a]	6.45	5.46	−7.26	−4.01	6.19	4.21
SD[b]	0.69	1.43	0.85	1.34	0.70	1.62
m[c]	4	5	4	5	4	5
t-test[d]	18.57 (*df*=3)	8.52 (*df*=4)	−16.99 (*df*=3)	−6.70 (*df*=4)	17.81 (*df*=3)	5.82 (*df*=4)
p-value	0.00015	0.0005	0.0002	0.0013	0.0002	0.00215

Note. [a] Means are computed as an average of individuals' difference score over sessions between the two adjacent phases. Missing scores are left as missing.
[b] *SD*s are computed as the square root of the variance of individuals' difference scores. Missing scores are left as missing.
[c] *m* = number of participants.
[d] *t*-test of adjacent phases, *df* = *m*−1.

Assessment of trend

"*Trend* refers to the slope of the best-fitting straight line for the data within a phase," according to The *WWC Handbook* (2013, p. E.6). Because a best-fitting straight line is a narrow definition for trends, we elected to assess monotonic trends in the Lambert data set using the Page test. A monotonic trend can be either increasing or decreasing. It is more general than a linear trend because a monotonic trend incorporates different slopes throughout a data pattern to reflect an upward (or increasing), or a downward (or decreasing), trend in data. Marascuilo and McSweeney (1977) and Page (1963) recommended the Page test for testing monotonic changes over time in SCD. The type of measurement required by the Page test is ranks of data or ranked data. Marascuilo and Busk (1988) and Busk and Marascuilo (1992) effectively applied the Page test to assess trends in the simple AB design, the multiple-baseline AB designs and replicated ABAB designs across participants. Recently, Peng and Chen (2014) proposed a measure of effect sizes (ES) and its confidence interval (CI) to accompany the Page test. Both the ES and its CI are derived from the Page test statistic to further determine an increasing, or a decreasing, trend in data.

To assess trends in the Lambert data set, we conducted six Page tests, computed six corresponding ES measures and their CIs. These results appear in Tables 2-7. SAS computing codes for assessing trends in Lambert-A data are shown in Part B of Appendix A.

Six Page tests of trends

The Page test was applied to three adjacent phases (SSR1-RC1, RC1-SSR2, SSR2-RC2) in both Lambert A and B data sets. A total of six Page tests were performed. According to Lambert et al. (2006), the RC intervention should minimize a student's disruptive behavior. Therefore, for two of the three adjacent phrases (i.e., SSR1-RC1 and SSR2-RC2), we proposed to test the null hypothesis of no trend against the alternative of a monotonic decreasing trend. For the RC1-SSR2 adjacent phrases, the null hypothesis is the same as before; yet the alternative hypothesis states that there is a monotonic increasing trend. Thus, all alternative hypotheses were directional. For demonstration purposes, we describe the Page test of the SSR1-RC1 phases from the Lambert-A data first (Table 2). The results of the other two adjacent phases from Set A are presented in Tables 3 and 4. Parallel analyses of the Lambert-B data appear in Tables 5-7.

For data obtained from the SSR1-RC1 phases in Lambert-A data, the following null and alternative hypotheses are specified, in (1) and (2), respectively:

$$H_0 : \bar{R}_1 = \bar{R}_2 = \ldots = \bar{R}_{14} \tag{1}$$

$$H_1 : \bar{R}_1 \geq \bar{R}_2 \geq \ldots \geq \bar{R}_{14}, \text{ with at least one strict inequality.} \tag{2}$$

Note that the null and alternative hypotheses specify mean ranks of students' scores only. Furthermore, the rejection of H_0 requires no more than one inequality in the ranked data, a decline in this case. In order to apply the Page test to test H_0 in (1), the raw data in the upper panel of Table 2 were converted to ranks for each student, shown in the middle panel of Table 2. Ranks are assigned from high to low within each student. If scores were tied, we broke the tie by averaging the two corresponding ranks, such as assigning the rank of 10.5 to the two 7s for Student A1 in both Sessions 1 and 5 during the SSR1 phase. Missing data were treated conservatively in the sense of supporting the null hypothesis, instead of the alternative hypothesis. Thus, if the H_0 of no trend can be rejected at $\alpha = .05$ with this conservative approach, it can be rejected at the same or a lower α level, if the missing data were replaced by a score in support of the alternative hypothesis. Thus, for Student A1 in Session 11 in the RC1 phase (upper panel of Table 2), we treated the missing score, shown as a period (.), with a score of 2, appearing in parenthesis. The score of 2 was the highest score of Student A1 in the RC1 phase. Replacing the missing score by 2 supported the null hypothesis of no trend, more

than other scores taken from Student A1 for this phase. This replacement led to a rank of 5.5, shown in parenthesis, in the middle panel of Table 2. Likewise, for Student A2 in Session 3 in the SSR1 phase, we treated the missing score with 6, in parenthesis. The score of 6 was the lowest score of Student A2 in the SSR1 phase. Other missing data were treated similarly in either the SSR1 or the RC1 phase.

Table 2. Number of intervals of disruptive behaviors and their ranks in 8 sessions (1 to 8) of the SSR1 phase and 6 sessions (9 to 14) of the RC1 phase of Class A (Lambert et al., 2006)

		SSR1								RC1					
	Session	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
	A1	7	9	8	6	7	4	5	10	2	0	.(2)	1	0	0
	A2	8	7	.(6)	7	8	6	7	9	3	1	0	4	0	0
	A3	10	.(6)	6	.(6)	6	9	6	10	.(1)	0	1	1	0	0
	A4	10	.(4)	6	4	8	8	9	10	3	6	0	0	.(6)	1
	Mean	8.75	6.5	6.5	5.75	7.25	6.75	6.75	9.75	2.25	1.75	0.75	1.5	1.5	0.25
	SD	1.5	2.08	1	1.26	0.96	2.22	1.71	0.5	0.96	2.87	0.96	1.73	3	0.5
		SSR1 Ranks								**RC1 Ranks**					
	Session	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
	A1	10.5	13	12	9	10.5	7	8	14	5.5	2	(5.5)	4	2	2
	A2	12.5	10	(7.5)	10	12.5	7.5	10	14	5	4	2	6	2	2
	A3	13.5	(9)	9	(9)	9	12	9	13.5	(5)	2	5	5	2	2
	A4	13.5	(5.5)	8	5.5	10.5	10.5	12	13.5	4	8	1.5	1.5	(8)	3
	Total Rank $\left(\sum_{i=1}^{m=4}\bar{R}_j\right)$	50	37.5	36.5	33.5	42.5	37	39	55	19.5	16	14	16.5	14	9
	Expected Rank (Y_j)	14	13	12	11	10	9	8	7	6	5	4	3	2	1

H₀	$H_0: \bar{R}_1 = \bar{R}_2 = \ldots = \bar{R}_{14}$	***m, n***	4, 14	**Standardized *L* (or *z*)**	z = 5.06[c]
H₁	$H_1: \bar{R}_1 \geq \bar{R}_2 \geq \ldots \geq \bar{R}_n$, w / ≥ 1 strict inequality	***χ²(df=1)***	25.60 [b]	***z*-upper**	7.02[d]
Page *L*	*L* = 3788.5[a]	***p*-value**	< .0001	***z*-lower**	3.10[d]

Note: Missing data are denoted as a period (.). Its rank is based on the score shown in parenthesis next to the period (.). The score in the parenthesis is assigned a rank, also shown in parenthesis, based on the lowest score of SSR1 phase or the highest score of the RC1 phase, in support of the H_0. Tied scores are assigned the average rank of the corresponding ranks.

[a] $$L = 3788.5 = \sum_{j=1}^{n=14}\left[Y_j \times \left(\sum_{i=1}^{m=4}\bar{R}_j\right)\right] = \begin{bmatrix} 14\times(50)+13\times(37.5)+12\times(36.5)+11\times(33.5) \\ +10\times(42.5)+9\times(37)+8\times(39)+7\times(55)+6\times(19.5) \\ +5\times(16)+4\times(14)+3\times(16.5)+2\times(14)+1\times(9) \end{bmatrix}.$$

[b] $$25.60 = \chi_L^2 = \frac{\left[12L - 3mn(n+1)^2\right]^2}{mn^2(n^2-1)(n+1)} = \frac{\left[12\times 3788.5 - 3\times 4\times 14\times(14+1)^2\right]^2}{4\times 14^2\times(14^2-1)\times(14+1)} = \frac{[45462-37800]^2}{4\times 196\times 195\times 15} = \frac{58706244}{2293200} = 25.60014129.$$

[c] $$5.06 = z = \sqrt{\chi_L^2} = \sqrt{25.60014129} = 5.059658.$$

[d] 95% CI for Standardized L = $z \pm 1.96 = 5.06 \pm 1.96 = [3.10, 7.02]$.

Table 3. Number of intervals of disruptive behaviors and their ranks in 8 sessions (15 to 22) of the SSR2 phase and 9 sessions (23 to 31) of the RC2 phase of Class A (Lambert et al., 2006)

	SSR2							RC2									
Session	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**
A1	8	8	8	6	10	10	10	8	3	4	1	3	2	4	0	1	0
A2	8	9	10	7	9	10	8	10	1	1	0	5	3	6	0	0	2
A3	5	7	10	.(5)	5	10	9	10	4	6	5	7	0	0	0	0	. (7)
A4	3	8	10	.(3)	10	10	10	5	6	1	5	0	. (6)	. (6)	0	0	1
Mean	6.00	8.00	9.50	5.25	8.50	10.00	9.25	8.25	3.50	3.00	2.75	3.75	2.75	4.00	0.00	0.25	2.50
SD	2.45	0.82	1.00	1.71	2.38	0	0.96	2.36	2.08	2.45	2.63	2.99	2.50	2.83	0.00	0.50	3.11
	SSR2 Ranks							**RC2 Ranks**									
Session	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**
A1	12.5	12.5	12.5	10	16	16	16	12.5	6.5	8.5	3.5	6.5	5	8.5	1.5	3.5	1.5
A2	11.5	13.5	16	10	13.5	16	11.5	16	4.5	4.5	2	8	7	9	2	2	6
A3	7.5	12	16	(7.5)	7.5	16	14	16	5	10	7.5	12	2.5	2.5	2.5	2.5	(12)
A4	6.5	13	15.5	(6.5)	15.5	15.5	15.5	8.5	11	4.5	8.5	2	(11)	(11)	2	2	4.5
Total Rank	38	51	60	34	52.5	63.5	57	53	27	27.5	21.5	28.5	25.5	31	8	10	24
Expected Rank	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1

H₀	$H_0 : \bar{R}_1 = \bar{R}_2 = \ldots = \bar{R}_n$	***m, n***	4, 17	**Standardized *L* (or *z*)**	z2 = 5.08[c]
H₁	$H_1 : \bar{R}_1 \geq \bar{R}_2 \geq \ldots \geq \bar{R}_n$, w / ≥1strict inequality	***χ²* (*df*=1)**	25.77 [b]	***z*-upper**	7.04[d]
Page *L*	*L2* = 6543.5[a]	***p*-value**	< .0001	***z*-lower**	3.12[d]

***Note*:** Missing data are denoted as a period (.). Its rank is based on the score shown in parenthesis next to the period (.). The score in the parenthesis is assigned a rank, also shown in parenthesis, based on the lowest score of SSR2 phase or the highest score of the RC2 phase, in support of the H_0. Tied scores are assigned the average rank of the corresponding ranks.

[a] $$L2 = 6543.5 = \sum_{j=1}^{n=17}\left[Y_j \times \left(\sum_{i=1}^{m=4}\bar{R}_j\right)\right] = \begin{bmatrix} 17\times(38)+16\times(51)+15\times(60)+14\times(34)+13\times(52.5) \\ +12\times(63.5)+11\times(57)+10\times(53)+9\times(27)+8\times(27.5) \\ +7\times(21.5)+6\times(28.5)+6\times(28.5)+5\times(25.5) \\ +4\times(31)+3\times(8)+2\times(10)+1\times(24) \end{bmatrix}$$

[b] $$25.77 = \chi^2_{L2} = \frac{\left[12L - 3mn(n+1)^2\right]^2}{mn^2\left(n^2-1\right)(n+1)} = \frac{\left[12\times 6543.5 - 3\times 4\times 17\times(17+1)^2\right]^2}{4\times 17^2\times\left(17^2-1\right)\times(17+1)} = \frac{\left[78522-66096\right]^2}{4\times 289\times 288\times 18} = \frac{154405476}{5992704} = 25.76557694.$$

[c] $$z2 = 5.08 = \sqrt{\chi^2_{L2}} = \sqrt{25.76557694} = 5.075980392.$$

[d] $$95\%\ \text{CI for StandardizedL2} = z2 \pm 1.96 = 5.08 \pm 1.96 = [3.12, 7.04].$$

Table 4. Number of intervals of disruptive behaviors and their ranks of 6 sessions (9 to 14) of the RC1 phase and 8 sessions (15-22) of the SSR2 phase of Class A (Lambert et al., 2006)

	RC1						SSR2							
Session	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
A1	2	0	.(2)	1	0	0	8	8	8	6	10	10	10	8
A2	3	1	0	4	0	0	8	9	10	7	9	10	8	10
A3	.(1)	0	1	1	0	0	5	7	10	. (5)	5	10	9	10
A4	3	6	0	0	.(6)	1	3	8	10	. (3)	10	10	10	5
Mean	2.25	1.75	0.75	1.50	1.50	0.25	6.00	8.00	9.50	5.25	8.50	10.00	9.25	8.25
SD	0.96	2.87	0.96	1.73	3.00	0.50	2.45	0.82	1.00	1.71	2.38	0.00	0.96	2.36
	RC1 Ranks						**SSR2 Ranks**							
Session	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
A1	5.5	2	(5.5)	4	2	2	9.5	9.5	9.5	7	13	13	13	9.5
A2	5	4	2	6	2	2	8.5	10.5	13	7	10.5	13	8.5	13
A3	(5)	2	5	5	2	2	8	10	13	(8)	8	13	11	13
A4	5	8.5	1.5	1.5	(8.5)	3	5	10	12.5	(5)	12.5	12.5	12.5	7
Total Rank	20.5	16.5	14	16.5	14.5	9	31	40	48	27	44	51.5	45	42.5
Expected Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14

H₀	$H_0: \bar{R}_1 = \bar{R}_2 = \ldots = \bar{R}_n$	***m, n***	4, 14	**Standardized *L* (or *z*)**	z3 = 5.22[c]
H₁	$H_1: \bar{R}_1 \geq \bar{R}_2 \geq \ldots \geq \bar{R}_n$, w / ≥1strict inequality	***χ²* (*df*=1)**	27.27 [b]	***z*-upper**	7.18[d]
Page *L*	*L3* = 3809[a]	***p*-value**	< .0001	***z*-lower**	3.26[d]

***Note*:** Missing data are denoted as a period (.). Its rank is based on the score shown in parenthesis next to the period (.). The score in the parenthesis is assigned a rank, also shown in parenthesis, based on the lowest score of SSR2 phase or the highest score of the RC1 phase, in support of the H_0. Tied scores are assigned the average rank of the corresponding ranks.

[a] $$L3 = 3809 = \sum_{j=1}^{n=14}\left[Y_j \times \left(\sum_{i=1}^{m=4}\bar{R}_j\right)\right] = \begin{bmatrix} 1\times(20.5)+2\times(16.5)+3\times(14)+4\times(16.5)+5\times(14.5) \\ +6\times(9)+7\times(31)+8\times(40)+9\times(48)+10\times(27) \\ +11\times(44)+12\times(51.5)+13\times(45)+14\times(42.5) \end{bmatrix}.$$

[b] $$27.27 = \chi^2_{L3} = \frac{\left[12L - 3mn(n+1)^2\right]^2}{mn^2\left(n^2-1\right)(n+1)} = \frac{\left[12\times 3809 - 3\times 4\times 14(14+1)^2\right]^2}{4\times 14^2\times\left(14^2-1\right)\times(14+1)} = \frac{\left[45708-37800\right]^2}{4\times 196\times 195\times 15} = \frac{62536464}{2293200} = 27.27039246.$$

[c] $$z3 = 5.22 = \sqrt{\chi^2_{L3}} = \sqrt{27.27039246} = 5.222106133.$$

[d] $$95\%\ \text{CI for Standardized}L3 = z3 \pm 1.96 = 5.22 \pm 1.96 = [3.26, 7.18].$$

Table 5. Number of intervals of disruptive behaviors and their ranks in 10 sessions (1 to 10) of the SSR1 phase and 6 sessions (11 to 16) of the RC1 phase of Class B (Lambert et al., 2006)

	SSR1										RC1					
Session	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
B1	10	6	9	4	5	9	6	10	9	9	4	3	4	4	1	0
B2	7	4	5	.(4)	.(4)	7	8	4	8	8	0	0	0	0	.(0)	.(0)
B3	6	.(6)	6	.(6)	.(6)	8	9	10	9	8	0	1	2	1	1	0
B4	8	1	4	6	6	7	8	8	0	2	0	.(6)	0	0	2	6
B5	9	5	4	2	3	10	4	10	8	8	0	2	1	3	0	0
Mean	8.00	4.40	5.60	4.40	4.80	8.20	7.00	8.40	6.80	7.00	.80	2.40	1.40	1.60	.80	1.20
SD	1.58	2.07	2.07	1.67	1.30	1.30	2.00	2.61	3.83	2.83	1.79	2.30	1.67	1.82	0.84	2.68

	SSR1 Ranks										RC1 Ranks					
Session	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
B1	15.5	9.5	12.5	5.5	8	12.5	9.5	15.5	12.5	12.5	5.5	3	5.5	5.5	2	1
B2	12.5	8.5	11	(8.5)	(8.5)	12.5	15.0	8.5	15	15	3.5	3	3.5	3.5	(3.5)	(3.5)
B3	9	(9)	9	(9)	(9)	12.5	14.5	16	14.5	12.5	1.5	4	6	4	4	1.5
B4	15	5	8	10.5	10.5	13	15	15	2.5	6.5	2.5	(10.5)	2.5	2.5	6.5	10.5
B5	14	11	9.5	5.5	7.5	15.5	9.5	15.5	12.5	12.5	2	5.5	4	7.5	2	2
Total Rank	66	43	50	39	43.5	66	63.5	70.5	57	59	15	26.5	21.5	23	18	18.5
Expected Rank	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1

H₀	$H_0: \bar{R}_1 = \bar{R}_2 = \ldots = \bar{R}_n$	***m, n***	5, 16	**Standardized *L* (or *z*)**	z4 = 4.82[c]
H₁	$H_1: \bar{R}_1 \geq \bar{R}_2 \geq \ldots \geq \bar{R}_n$, w / ≥1 strict inequality	χ^2 **(*df*=1)**	23.25 [b]	***z*-upper**	6.78[d]
Page *L*	*L4* = 6726.5[a]	***p*-value**	< .0001	***z*-lower**	2.86[d]

Note: Missing data are denoted as a period (.). Its rank is based on the score shown in parenthesis next to the period (.). The score in the parenthesis is assigned a rank, also shown in parenthesis, based on the lowest score of SSR1 phase or the highest score of the RC1 phase, in support of the H_0. Tied scores are assigned the average rank of the corresponding ranks.

[a] $$L4 = 6726.5 = \sum_{j=1}^{n=16}\left[Y_j \times \left(\sum_{i=1}^{m=4}\bar{R}_j\right)\right] = \begin{bmatrix} 16\times(66)+15\times(43)+14\times(50)+13\times(39)+12\times(43.5)+11\times(66) \\ +10\times(63.5)+9\times(70.5)+8\times(57)+7\times(59)+6\times(15) \\ +5\times(26.5)+4\times(21.5)+3\times(23)+2\times(18)+1\times(18.5) \end{bmatrix}$$

[b] $$23.25 = \chi^2_{L4} = \frac{\left[12L - 3mn(n+1)^2\right]^2}{mn^2\left(n^2-1\right)(n+1)}$$

$$= \frac{\left[12\times 6726.5 - 3\times 5\times 16\times(16+1)^2\right]^2}{5\times 16^2\times\left(16^2-1\right)\times(16+1)} = \frac{[80718-69360]^2}{5\times 256\times 255\times 17} = \frac{129004164}{5548800} = 23.24902033.$$

[c] $$z4 = 4.82 = \sqrt{\chi^2_{L4}} = \sqrt{23.24902033} = 4.821723792.$$

[d] $$95\%\ \text{CI for Standardized}L4 = z4 \pm 1.96 = 4.82 \pm 1.96 = [2.86, 6.78].$$

Table 6. Number of intervals of disruptive behaviors and their ranks in 7 sessions (17 to 23) of the SSR2 phase and 11 sessions (24 to 34) of the RC2 phase of Class B (Lambert et al., 2006)

	SSR2							RC2										
Session	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**
B1	3	5	8	10	10	10	6	3	0	2	4	1	0	1	3	0	1	0
B2	5	7	6	4	.(4)	6	5	.(2)	0	0	0	2	0	0	.(2)	0	0	0
B3	2	4	4	5	8	8	7	1	0	3	.(3)	1	0	1	0	.(3)	1	0
B4	5	6	5	8	4	0	2	1	2	6	0	2	0	1	1	.(6)	.(6)	.(6)
B5	.(0)	3	0	2	7	7	2	0	.(4)	1	0	2	2	4	0	0	1	1
Mean	3.00	5.00	4.60	5.80	6.60	6.20	4.40	1.40	1.20	2.40	1.40	1.60	0.40	1.40	1.20	1.80	1.80	1.40
SD	2.12	1.58	2.97	3.19	2.61	3.77	2.30	1.14	1.79	2.30	1.95	0.55	0.89	1.52	1.30	2.68	2.39	2.61
	SSR2 Ranks							**RC2 Ranks**										
Session	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**
B1	10	13	15	17	17	17	14	10	2.5	8	12	6.0	2.5	6	10	2.5	6	2.5
B2	14.5	18	16.5	12.5	(12.5)	16.5	14.5	(10)	4.5	4.5	4.5	10	4.5	4.5	(10)	4.5	4.5	4.5
B3	9	13.5	13.5	15	17.5	17.5	16	6.5	2.5	11	(11)	6.5	2.5	6.5	2.5	(11)	6.5	2.5
B4	11.5	15	11.5	18	10	2	8	5.0	8	15	2	8	2	5	5	(15)	(15)	(15)
B5	(3.5)	14.0	3.5	11.5	17.5	17.5	11.5	3.5	(15.5)	8	3.5	11.5	11.5	15.5	3.5	3.5	8	8
Total Rank	48.5	73.5	60	74	74.5	70.5	64	35	33	46.5	33	42	23	37.5	31	36.5	40	32.5
Expected Rank	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1

H₀	$H_0 : \bar{R}_1 = \bar{R}_2 = \ldots = \bar{R}_n$	***m, n***	5, 18	**Standardized *L* (or *z*)**	z5 = 4.42[c]
H₁	$H_1 : \bar{R}_1 \geq \bar{R}_2 \geq \ldots \geq \bar{R}_n$, w / ≥1strict inequality	***χ²* (*df*=1)**	19.51 [b]	***z*-upper**	6.38[d]
Page *L*	*L5* = 9283[a]	***p*-value**	< .0001	***z*-lower**	2.46[d]

Note: Missing data are denoted as a period (.). Its rank is based on the score shown in parenthesis next to the period (.). The score in the parenthesis is assigned a rank, also shown in parenthesis, based on the lowest score of SSR2 phase or the highest score of the RC2 phase, in support of the H_0. Tied scores are assigned the average rank of the corresponding ranks.

[a] $$L5 = 9283 = \sum_{j=1}^{n=18}\left[Y_j \times \left(\sum_{i=1}^{m=5}\bar{R}_j\right)\right] = \begin{bmatrix} 18\times(48.5)+17\times(73.5)+16\times(60)+15\times(74)+14\times(74.5)+13\times(70.5) \\ +12\times(64)+11\times(35)+10\times(33)+9\times(46.5)+8\times(33)+7\times(42) \\ +6\times(23)+5\times(37.5)+4\times(31)+3\times(36.5)+2\times(40)+1\times(32.5) \end{bmatrix}.$$

[b] $$19.51 = \chi^2_{L5} = \frac{\left[12L - 3mn(n+1)^2\right]^2}{mn^2(n^2-1)(n+1)}$$

$$= \frac{\left[12\times 9283 - 3\times 5\times 18\times(18+1)^2\right]^2}{5\times 18^2\times(18^2-1)\times(18+1)} = \frac{\left[111396-97470\right]^2}{5\times 324\times 323\times 19} = \frac{193933476}{9941940} = 19.50660294.$$

[c] $$z5 = 4.42 = \sqrt{\chi^2_{L5}} = \sqrt{19.50660294} = 4.416628005.$$

[d] $$95\%\text{ CI for Standardized}L5 = z5 \pm 1.96 = 4.42 \pm 1.96 = [2.46, 6.38].$$

Table 7. Number of intervals of disruptive behaviors and their ranks of 6 sessions (11 to 16) of the RC1 phase and 7 sessions (17-23) of the SSR2 phase of Class B (Lambert et al., 2006)

	RC1						SSR2						
Session	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**
B1	4	3	4	4	1	0	3	5	8	10	10	10	6
B2	0	0	0	0	. (0)	.(0)	5	7	6	4	.(4)	6	5
B3	0	1	2	1	1	0	2	4	4	5	8	8	7
B4	0	.(6)	0	0	2	6	5	6	5	8	4	0	2
B5	0	2	1	3	0	0	.(0)	3	0	2	7	7	2
Mean	.80	2.40	1.40	1.60	.80	1.20	3.00	5.00	4.60	5.80	6.60	6.20	4.40
SD	1.79	2.30	1.67	1.82	.84	2.68	2.12	1.58	2.97	3.19	2.61	3.77	2.30
	RC1 Ranks						**SSR2 Ranks**						
Session	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**
B1	6	3.5	6	6	2	1	3.5	8	10	12	12.	12	9
B2	3.5	3.5	3.5	3.5	(3.5)	(3.5)	9.5	13	11.5	7.5	(7.5)	11.5	9.5
B3	1.5	4	6.5	4	4	1.5	6.5	8.5	8.5	10	12.5	12.5	11
B4	2.5	(11)	2.5	2.5	5.5	11	8.5	11	8.5	13	7	2.5	5.5
B5	3	8	6	10.5	3	3	(3)	10.5	3	8	12.5	12.5	8
Total Rank	16.5	30	24.5	26.5	18	20	31	51	41.5	50.5	51.5	51	43
Expected Rank	1	2	3	4	5	6	7	8	9	10	11	12	13

H₀	$H_0: \bar{R}_1 = \bar{R}_2 = \ldots = \bar{R}_n$	***m, n***	5, 13	**Standardized *L* (or *z*)**	z6 = 4.44[c]
H₁	$H_1: \bar{R}_1 \geq \bar{R}_2 \geq \ldots \geq \bar{R}_n$, w / ≥1strict inequality	**χ^2 (*df*=1)**	19.74 [b]	***z*-upper**	6.40[d]
Page *L*	*L6* =3707[a]	***p*-value**	< .0001	***z*-lower**	2.48[d]

Note: Missing data are denoted as a period (.). Its rank is based on the score shown in parenthesis next to the period (.). The score in the parenthesis is assigned a rank, also shown in parenthesis, based on the lowest score of SSR2 phase or the highest score of the RC1 phase, in support of the H_0. Tied scores are assigned the average rank of the corresponding ranks.

[a] $$L6 = 3707 = \sum_{j=1}^{n=13}\left[Y_j\left(\sum_{i=1}^{m=5}\bar{R}_j\right)\right] = \begin{bmatrix} 1\times(16.5)+2\times(30)+3\times(24.5)+4\times(26.5) \\ +5\times(18)+6\times(20)+7\times(31)+8\times(51)+9\times(41.5) \\ +10\times(50.5)+11\times(51.5)+12\times(51)+13\times(43) \end{bmatrix}.$$

[b] $$19.74 = \chi^2_{L6} = \frac{\left[12L - 3mn(n+1)^2\right]^2}{mn^2\left(n^2-1\right)(n+1)}$$

$$= \frac{\left[12\times 3707 - 3\times 5\times 13\times(13+1)^2\right]^2}{5\times 13^2\times\left(13^2-1\right)\times(13+1)} = \frac{[44484-38220]^2}{5\times 169\times 168\times 14} = \frac{39237696}{1987440} = 19.74283299.$$

[c] $$z6 = 4.44 = \sqrt{\chi^2_{L6}} = \sqrt{19.74283299} = 4.443290784.$$

[d] $$95\%\ \text{CI for Standardized}L6 = z6 \pm 1.96 = 4.44 \pm 1.96 = [2.48, 6.40].$$

Next, we computed the total rank for each of the 14 sessions. The total ranks $\left(\sum_{i=1}^{m=4} R_{ij}\right)$ were subsequently weighted by their expected ranks (Y_j), suggested by

H_1. The product of the total rank weighted by its expected rank was subsequently summed over all 14 sessions into the Page statistic, L, according to (3) below:

$$\begin{aligned} L &= \sum_{j=1}^{n=14}\left[Y_j \times \left(\sum_{i=1}^{m=4} R_{ij} \right) \right] \\ &= \begin{bmatrix} 14\times(50)+13\times(37.5)+12\times(36.5)+11\times(33.5) \\ +10\times(42.5)+9\times(37)+8\times(39)+7\times(55)+6\times(19.5) \\ +5\times(16)+4\times(14)+3\times(16.5)+2\times(14)+1\times(9) \end{bmatrix} \\ &= 3788.5 \end{aligned} \tag{3}$$

where, n = the number of sessions, m = the number of students/participants, Y_j = the expected rank of the j^{th} session, and R_{ij} = the observed rank of the i^{th} student's score in the j^{th} session.

The exact significance level of the L statistic can be obtained from Page (1963), if n ranges from 3 to 10 and m ranges from 2 to 50. Given the present values of $n = 14$ and $m = 4$, the significance level can be approximated by a chi-square distribution with $df = 1$, according to (4) below (Page, 1963, p. 224):

$$\begin{aligned} \chi^2_L &= \frac{\left[12L - 3mn(n+1)^2 \right]^2}{mn^2\left(n^2-1\right)(n+1)} \\ &= \frac{\left[12\times 3788.5 - 3\times 4\times 14\times(14+1)^2 \right]^2}{4\times 14^2\times\left(14^2-1\right)\times(14+1)} \\ &= \frac{\left[45462-37800\right]^2}{4\times 196\times 195\times 15} = \frac{58706244}{2293200} = 25.60 \end{aligned} \tag{4}$$

The above chi-square statistic is statistically significant at $p < .0001$ leading to a rejection of H_0 of no trend at $\alpha = .05$ (one-tailed), specified in (1) above. We therefore concluded that there was a monotonic decreasing trend across these 14 sessions, as specified in H_1 of (2).

The large-sample approximation to the sampling distribution of Page's L statistic yields acceptable Type I error rates for a directional Page test, as long as $n > 11$ for $\alpha = .05$, or $n > 18$ for $\alpha = .01$, according to Fahoome (2002). An acceptable Type I error rate was defined in Fahoome (2002) as within 10% of the

nominal α rate, in reference to Bradley (1978)'s work. Page (1963) also suggested that the large-sample chi-square approximation be used under one of three conditions: (1) for $m > 20$ with any n, (2) for $m > 12$ and $n \geq 4$, or (3) for any m when $n \geq 9$. Because $m = 4$ and $n = 14$, the Page test result and its statistical significance level were judged to be acceptable, according to Bradley (1978), Fahoome (2002), and Page (1963).

Summary of six Page tests of trends. The Page test was applied similarly to two other adjacent phases from Lambert-A data and to the three adjacent phases from Lambert-B data. Results of these Page tests are summarized in Tables 3 to 7, including their corresponding H_0s and H_1s. All six Page tests shown in Tables 2 to 7 were statistically significant at $p < .0001$, rejecting all H_0s at $\alpha = .05$ (one-tailed) and confirming a trend as specified in the corresponding H_1s. For data in the SSR1-RC1 and the SSR2-RC2 adjacent phases, the Page test results of L, $L2$, $L4$, and $L5$ suggested a monotonic decreasing trend from the baseline phase (i.e., SSR) to the intervention phase (i.e., RC) in both Lambert-A and -B data sets. For data in the RC1-SSR2 adjacent phases, the Page test results of $L3$ and $L6$ suggested a monotonic increasing trend from the intervention phase (i.e., RC1) to the baseline phase (i.e., SSR2) again for both A and B data sets.

Six ES measures derived from Page's *L*

The L statistic defined in (3) is conceptually and algebraically equivalent to the average Spearman rank correlation coefficient (ρ) between Students' ranked scores (i.e., the frequency of disruptive behaviors) and the expected ranks according to a monotonic decreasing or increasing trend (Page, 1963; van de Wiel & Di Bucchianico, 2001). It is an unstandardized ES measure of a monotonic trend in data. To convert L into a standardized ES, one divides Page's L (i.e., the average ρ) by its standard deviation (Lyerly, 1952; Page, 1963, p. 227) to yield a standardized normal z, as in (5):

$$\frac{\rho}{SD_\rho} = \rho \times \sqrt{m \times (n-1)} = z = \sqrt{\chi_L^2} = \sqrt{25.60} = 5.06 \qquad (5)$$

where χ_L^2 is defined in (4) above. This normalized z statistic is similar to Cohen's d, in the sense of being scale-free and ranging from negative to positive values without bounds. They differ, however, in their assumptions. Cohen's d and its population parameter δ assume normality and equal variances for underlying

populations (Cohen, 1988), whereas the standardized *L*, or the normalized *z* in (5), does not, because the latter is based on ranks of the data.

CI for the standardized ES derived from Page's *L*

Since the standardized *L*, or *z* from (5), follows a standard normal distribution (e.g., Fahoome, 2002; Lyerly, 1952), a nondirectional 95% CI for the standardized *L* can be constructed using (6) below:

$$95\%\ \text{CI for Standardized } L = z \pm 1.96 = 5.06 \pm 1.96 = [3.10, 7.02] \qquad (6)$$

Because the upper and the lower limits of the 95% CI are both positive, the 95% CI supports the earlier rejection of the H_0 of no trend at $\alpha = .05$, in favor of a monotonic decreasing trend across the 14 sessions from the SSR1-RC1 phases of the Lambert-A data.

Summary of six ESs and six CIs. The standardized ESs (or *z*s) and their corresponding CIs further confirmed the rejection of the H_0 of no trend and in favor of the H_1 of a monotonic trend. Taken together, the six Page test results, their corresponding ESs and CIs provided multiple evidence for monotonic decreasing trends in students' disruptive behaviors due to the intervention.

Assessment of variability

According to the *WWC Handbook* (2013), "*Variability* refers to the range or standard deviation of data about the best-fitting straight line." (p. E.6). Even though we did not fit a straight regression line to the Lambert data, the variability of scores was assessed within and between phases using SAS—see Part A of Appendix A; results are presented in Table 1. In five out of six instances, the intervention phases (RC1 and RC2) yielded less variability than their corresponding baseline phases, namely, SSR1 and SSR2 respectively. The only exception occurred in Lambert-A data set between SSR2 and RC2. We did not test the differences in variability because these statistical tests (e.g., Levene's *F*' test) are not robust under nonnormal conditions, which might be the case for the Lambert data.

Assessment of immediacy of the effect

According to the *WWC Handbook* (2013), "*Immediacy of the effect* refers to the change in level between the last three data points in one phase and the first three data points of the next. The more rapid (or immediate) the effect, the more convincing the inference that change in the outcome measure was due to manipulation of the independent variable." (p. E.6). Applying this definition to Figure 1 using the visual analysis, we determined that data patterns in the intervention phases (i.e., RC1 and RC2) exhibited an immediate decreasing effect on disruptive behaviors, compared to data patterns in the baseline phases (i.e., SSR1 and SSR2). Even though the last three data points of Student B4's from the SSR2 phase, compared to the first three data points of the RC2 phase, suggested an exception, the overall profile of this student's data supported a decline in disruptive behavior during the intervention phase. Thus, we concluded that there was an immediacy effect due to the intervention in both A and B data sets.

Assessment of overlap

According to the *WWC Handbook* (2013), "*Overlap* refers to the proportion of data from one phase that overlaps with data from the previous phase. The smaller the proportion of overlapping data points (or conversely, the larger the separation), the more compelling the demonstration of an effect." (p. E.6). To assess this data feature, we computed the degree of nonoverlap for all data pairs (NAP) in adjacent phases for each student (Table 8). NAP is defined as the number of pairs of data showing no overlap between a baseline phase and an intervention phase, divided by the total number of pairs (Parker & Vannest, 2009). Each NAP corresponds to two adjacent phases, such as SSR1 and RC1. Values of NAP range from 0 to 1. A value of 0 indicates that all data points in phase A (e.g., SSR1) are greater than the points in phase B (e.g., RC1). In contrast, a value of 1 indicates that all data points in phase A (e.g., RC1) are smaller than the points in phase B (e.g., SSR2). According to Table 8, all NAP results were statistically significant at $\alpha = .05$ (two-tailed), except for two students (B4 and B5) in two adjacent phases (RC1-SSR2, and SSR2-RC2). We therefore concluded that there was a statistically significant lack of overlap in students' outcome measures between phases, supporting the effectiveness of the intervention in decreasing disruptive behaviors. The NAPs and their corresponding statistical significance were computed using a free web-based calculator from http://www.singlecaseresearch.org/. The web-based calculator was developed by

Vannest, Parker, and Gonen (2011) and its functionalities are shown in Appendix B. The NAP results were subsequently verified by SPSS, shown in Appendix C.

Table 8. Nonoverlap of All Pairs (NAP) between phases in Lambert-A and -B data sets

	SSR1-RC1		RC1-SSR2		SSR2-RC2	
	NAP [a]	*p*-value [b]	NAP	*p*-value	NAP	*p*-value
Student A1	0.0000	0.0034	1.0000	0.0034	0.0417	0.0015
Student A2	0.0000	0.0027	1.0000	0.0019	0.0000	0.0005
Student A3	0.0000	0.0062	1.0000	0.0045	0.0982	0.0092
Student A4	0.0429	0.0094	0.9286	0.0149	0.0714	0.0073
Student B1	0.0250	0.0020	0.9167	0.0124	0.0260	0.0009
Student B2	0.0000	0.0066	1.0000	0.0105	0.0000	0.0015
Student B3	0.0000	0.0027	0.9881	0.0034	0.0079	0.0010
Student B4	0.1800	0.0500	0.7571	0.1439	0.2232	0.0728
Student B5	0.0333	0.0024	0.7778	0.1093	0.2167	0.0652

Note: Missing scores are left as missing.
[a] NAPs were computed using a web-based calculator developed by Vannest, Parker, and Gonen (2011)—see Appendix B, and verified by SPSS's Receiver Operator Characteristics module—see Appendix C.
[b] *p*-values were obtained from the web-based calculator developed by Vannest, Parker, and Gonen (2011)—see Appendix B, and verified by SPSS's Receiver Operator Characteristics module and its option called Area Under the Curve (AUC)—see Appendix C

Assessment of consistency of data in similar phases

According to the *WWC Handbook* (2013, p. E.6), "*Consistency of data in similar phases* involves looking at data from all phases within the same condition… and examining the extent to which there is consistency in the data patterns from phases with the same conditions. The greater the consistency, the more likely the data represent a causal relation." To determine the consistency of data, we employed the visual analysis of the Lambert-A and –B data sets and determined that data patterns were similar in the same phase between these two sets. Furthermore, we applied four independent-samples t-tests to each phase between means of sets A and B, whether it was baseline or intervention (Table 9). According to Table 9, the *t*-test was not statistically significant for any phase at $\alpha = .05$ (two-tailed with $df = 7 = 4+5-2$). These statistically insignificant t-test results suggested that the mean scores obtained from sets A and B were not statistically significantly different from each other. Thus, we concluded that there was consistency of data patterns within similar phases for both data sets. SAS programming codes for assessing consistency in the Lambert-A data are shown in Part C of Appendix A.

Table 9. Means, *SD*s, *t*-tests of differences within phases in Lambert-A and -B data sets

	SSR1		RC1		SSR2		RC2	
	Set A	**Set B**	**Set A**	**Set B**	**Set A**	**Set B**	**Set A**	**Set B**
Mean[a]	7.53	6.68	1.08	1.22	8.34	5.23	2.15	1.02
SD[b]	$\sqrt{3.48}=1.87$	$\sqrt{5.76}=2.40$	$\sqrt{2.67}=1.63$	$\sqrt{2.41}=1.55$	$\sqrt{4.23}=2.06$	$\sqrt{5.91}=2.43$	$\sqrt{5.83}=2.42$	$\sqrt{1.76}=1.33$
m[c]	4	5	4	5	4	5	4	5
n[d]	8	10	6	6	8	7	9	11
|*t*|[e]	0.5824 (*SE*=1.468)		0.1317 (*SE*=1.063)		2.0345 (*SE*=1.529)		0.8978 (*SE*=1.259)	
***p*-value**	0.579		0.899		0.081		0.4	

[a] Means are computed as an average of individuals' mean score over sessions within each phase. Missing scores are left as missing.
[b] *SD*s are computed as the square root of the averaged variance of individuals' variances of scores within each phase. Missing scores are left as missing.
[c] *m* = number of participants or students.
[d] *n* = number of sessions.
[e] two-tailed *t*-test of Set A vs. Set B with *df* = 7.

Conclusions based on six assessments

The analyses summarized in Tables 1-9 and interpreted above collectively examined all data features recommended by the *WWC Handbook* (2013) for documenting an intervention effect. These assessments led to the same conclusion, as Lambert et al. (2006) did based on visual analysis alone. Next, we discuss the simplicity and rationality of the demonstrated approach, compared to visual analysis or complex statistical modeling and methods for determining intervention effects.

Discussion

In this paper, we demonstrated how to use free web-based resources or popular software to assess six data features recommended by the *WWC Handbook* (IES, 2013 February) to determine intervention effects in a single-case study (Lambert et al., 2006). The six data features are level and level change between phases, trend, variability, immediacy of the effect, overlap, and consistency of data in similar phases. Lambert et al. (2006) employed a reversal (or ABAB) design to collect data on the effectiveness of the report-card intervention in reducing students' disruptive behaviors in classrooms. The intervention was judged to be effective by Lambert et al. (2006) based on visual inspection alone. Our approach was to assess each of the six data features separately; then integrate six assessments into one comprehensive analysis of the intervention effect.

Among the six data features, the assessment of trends is probably most discussed but least agreed upon in the literature. To assess trends in the Lambert data, we employed the Page test and computed its ES and CI, proposed by Peng and Chen (2014). The Page test has been shown in the literature to be applicable to a variety of SCD contexts, such as, the simple AB designs, multiple-baseline AB designs, or replicated ABAB designs. They are equally applicable to one participant as well as to multiple participants, to one study as well as to multiple studies in a meta-analytic framework. The versatile Page test requires only ranked data. It can be computed and interpreted even when data have no variance (namely, there is uniformity in scores), display ceiling or floor effects, or are incomplete (Peng & Chen, 2014). Likewise, its proposed ES and CI are interpretable as they are direct derivatives from Page's *L* statistic. The proposed ES is a meaningful measure of intervention effects and its precision is expressed by the CI (Peng & Chen, 2014). Both ES and CI can be computed simply using SAS algorithms shown in Appendix A. The reporting of ES and its precision, expressed as CI, have been required or highly recommended by refereed journals and professional organizations, such as the American Psychological Association (APA) and American Educational Research Association (AERA) (AERA, 2006; APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008; APA, 2010; Peng, Chen, Chiang, & Chiang, 2013).

The Lambert et al. (2006) data were recently reanalyzed in five articles published in a special issue of *Journal of School Psychology* (Shadish, 2014) to demonstrate alternative ways of analyzing and reporting SCD data, beyond the initial visual analysis. Each article published in that special issue employed complex statistical models (such as, the hierarchical linear modeling) and/or methods (such as, the Bayesian approach). These complex models and methods are often difficult to conceptualize or implement by practitioners not specially trained for these methodologies. In our demonstration, we assessed each of the six data features separately; then integrated six assessments into one comprehensive analysis. The separate assessments and the final integration were carried out using tools free from the Internet, or from the popular statistical software, such as SAS and SPSS. Thus, our approach to the determination of intervention effects is both simple and comprehensive. It illustrates how researchers, clinicians, teachers, parents, or policy makers can be empowered to interpret data efficiently and formulate evidence-based conclusions logically from well-designed and well-executed single-case studies.

Acknowledgements

This research was supported in part by two Maris M. Proffitt and Mary Higgins Proffitt Endowment Grants of Indiana University awarded to C.-Y. J. Peng while L.-T. Chen served as a research assistant on these two grants.

References

American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*, *35*, 33-40. doi:10.3102/0013189X035006033

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

APA Publications and Communications Board Working Group on Journal Articles Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, *63*, 839-851. doi:10.1037/0003-066X.63.9.839

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*, 144-152. doi:10.1111/j.2044-8317.1978.tb00581.x

Busk, P. L., & Marascuilo, L. A. (1988). Autocorrelation in single-subject research: A counterargument to the myth of no autocorrelation. *Behavioral Assessment*, *10*, 229-242.

Busk, P. L., & Marascuilo, L. A. (1992). Statistical analysis in single-case research: Issues, procedures, and recommendations, with applications to multiple behaviors. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 159-186). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Fahoome, G. (2002). Twenty nonparametric statistics and their large sample approximations. *Journal of Modern Applied Statistical Methods*, *1*(2), 248-268. Retrieved from http://digitalcommons.wayne.edu/jmasm/vol1/iss2/35

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*(2), 165-179. Retrieved from http://journals.cec.sped.org/ec/

Institute of Education Sciences (2013 February). *What Works Clearinghouse: Procedures and standards handbook* (Version 3.0). Retrieved from http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19

Lambert, M. C., Cartledge, G., Heward, W. L., & Lo, Y.-Y. (2006). Effects of response cards on disruptive behavior and academic responding during math lessons by fourth-grade urban students. *Journal of Positive Behavior Interventions*, *8*, 88-99. doi:10.1177/10983007060080020701

Lyerly, S. B. (1952). The average Spearman rank correlation coefficient. *Psychometrika*, *17*(4), 421-428.

Marascuilo, L. A., & Busk, P. L. (1988). Combining statistics for multiple-baseline AB and replicated ABAB designs across subjects. *Behavioral Assessment*, *10*, 1-28.

Marascuilo, L. A., & McSweeney, M. (1977). *Nonparametric and distribution-free methods for the social sciences*. Monterey, CA: Brooks/Cole Publishing Company.

Page, E. B. (1963). Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Journal of the American Statistical Association*, *58*, 243-254. doi:10.2307/2282965

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy*, *40*, 357-367. doi:10.1016/j.beth.2008.10.006

Peng, C.-Y. J. & Chen, L.-T. (2014). *Assessing intervention effects in single-case studies using the Page test*. Manuscript retrieved from https://oncourse.iu.edu/access/content/user/peng/CI for page test.condensed.Peng_11_Chen_6_+tablefigure.7-15-2014.pdf (July 19th, 2014).

Peng, C.-Y. J., Chen, L.-T., Chiang, H.-M., & Chiang, Y.-C. (2013). The impact of APA and AERA guidelines on effect size reporting. *Educational Psychology Review, 25*, 157-209. doi:10.1007/s10648-013-9218-2

SAS Institute Inc. (2014). *SAS/STAT 13.1 user's guide*. Cary, NC: SAS Institute Inc.

Shadish, W. R. (Ed.) (2014). Analysis and meta-analysis of single-case designs [Special issue]. *Journal of School Psychology*, *52*(2).

van de Wiel, M. A., & Di Bucchianico, A. (2001). Fast computation of the exact null distribution of Spearman's ρ and Page's L statistic for samples with and without ties. *Journal of Statistical Planning and Inference*, *92*, 133-145. doi:10.1016/S0378-3758(00)00166-X

Vannest, K. J., Parker, R. I., & Gonen, O. (2011). *Single Case Research: web based calculators for SCR analysis*. (Version 1.0) [Web-based application]. College Station, TX: Texas A&M University. Retrieved Thursday 10th July 2014. Available from http://www.singlecaseresearch.org

Appendix A: SAS Program for Assessing Level/Level Change, Trends, Variability, and Consistency in Lambert-A Data

```
*-------------------------------------------------------------------------------------------
*Data came from Lambert et al., (2006) with two Classrooms, A and B.
*Class A data are analyzed in this program. Class B data can be analyzed similarly.
*Each class has two baselines, SSR1 and SSR2, each followed by an intervention: RC1 and RC2.
*Class A has 4 participants, A1 to A4 and 31 sessions: 1-8 in SSR1, 9-14 in RC1, 15-22 in SSR2,
* sessions 23-31 in RC2.
*Class B has 5 participants, B1 to B5 and 34 sessions: 1-10 in SSR1, 11-16 in RC1, 17-23 in SSR2,
* sessions 24-34 in RC2.
* 3 Page tests, their Chi-square tests, and p-values are computed in this program, for class A.
*
*-------------------------------------------------------------------------------------------;

OPTIONS LS=80 PAGENO=1;
TITLE 'Lambert A Data analyzed using Page test';

DATA A;                         /*Classroom A data of 4 students*/
     INPUT id $ score1-score31;

*Class A has 4 participants, A1 to A4 and 31 sessions: 1-8 in SSR1, 9-14 in RC1, 15-22 in SSR2, 23-31
in RC2;

     minssr1=min (OF score1-score8);
     maxrc1=max (OF score9-score14);
     minssr2 = min (OF score15-score22);
     maxrc2=max (OF score23-score31);

*Compute the mean of each student for each phase--------------------------------------;

     meanssr1=mean(OF score1-score8);
     meanrc1=mean(OF score9-score14);
     meanssr2=mean(OF score15-score22);
     meanrc2=mean(OF score23-score31);

*Compute differences of adjacent phases----------------------------------------------;

     diff_ssr1_rc1=meanssr1-meanrc1;
     diff_rc1_ssr2=meanrc1-meanssr2;
     diff_ssr2_rc2=meanssr2-meanrc2;

*Compute the variance of each student for each phase---------------------------------;

     varssr1=VAR (OF score1-score8);
     varrc1=VAR (OF score9-score14);
     varssr2=VAR (OF score15-score22);
     varrc2=VAR (OF score23-score31);

* Create new variables for single imputation missing data-----------------------------;

     ARRAY score{*} score1-score31;
     ARRAY new{*} new1-new31;

     DO i = 1 to 31 by 1;
        new{i} = score{i};
     END;

DATALINES;
A1 7  9  8  6  7  4  5 10  2  0  .  1  0  0  8  8  8  6 10  10 10   8  3  4  1  3  2  4  0  1  0
A2 8  7  .  7  8  6  7  9  3  1  0  4  0  0  8  9 10  7  9  10  8  10  1  1  0  5  3  6  0  0  2
A3 10 .  6  .  6  9  6 10  .  0  1  1  0  0  5  7 10  .  5  10  9  10  4  6  5  7  0  0  0  0  .
A4 10 .  6  4  8  8  9 10  3  6  0  0  .  1  3  8 10  . 10  10 10   5  6  1  5  0  .  .  0  0  1
;

* Compute descriptive stat. in the data set-------------------------------------------;
```

```
* Part A ------------------------------------------------------------;

PROC MEANS DATA=A; RUN;

* Replace missing scores in each phase by the min or max of that phase for each participant from
Lambert-A data set -------------;

DATA A1; SET A;
         ARRAY score{*} score1-score31;
         ARRAY new{*} new1-new31;
         Do i = 1 to 31 by 1;
            session = i;
                IF 1<= session <= 8 THEN phase = 'SSR1';
                ELSE IF 9 <=session <=14 THEN phase = 'RC1';
                ELSE IF 15<= session<=22 THEN phase = 'SSR2';
                ELSE IF 23<=session <=31 THEN phase = 'RC2';

                IF phase = 'SSR1' AND new{i}=. THEN new{i}=minssr1;
                ELSE IF phase = 'RC1' AND  new{i}=. THEN new{i}=maxrc1;
                    ELSE IF phase = 'SSR2' AND new{i}=. THEN new{i}=minssr2;
                ELSE IF phase = 'RC2' AND  new{i}=. THEN new{i}=maxrc2;

         END;
         KEEP id new1-new31;


*Create three data sets for two adjacent phases for Lambert A data set -------------------------;


DATA A_SSR1_RC1; set A1; KEEP id new1-new14;
DATA A_SSR2_RC2; set A1; KEEP id new15-new31;
DATA A_RC1_SSR2; set A1; KEEP id new9-new22;

*Rank data in SSR1-RC1 phases from SAS data set A_SSR1_RC1 of Lambert A data set ---------------;

PROC TRANSPOSE DATA=A_SSR1_RC1 OUT=Table1;
/* transpose the data matrix in order to rank scores*/
     ID id;
RUN;

PROC RANK DATA=Table1 OUT=Table1;
     VAR A1-A4;

PROC TRANSPOSE DATA=Table1 OUT=Table1 PREFIX=rank;   /* transpose the ranked data back */

* Compute total ranks for 14 sessions in SSR1-RC1 phases from SAS data set A_SSR1_RC1 of Lambert A
data set ----------;

PROC MEANS DATA=Table1;                          /* compute the total of rank1 to rank14 */
     VAR rank1-rank14;
     OUTPUT OUT=Table1 SUM=sum1-sum14;

PROC PRINT DATA = Table1; RUN;

* Part B-------------------------------------------------------------------;

* Compute Page L, chi-square, z and CI of z for SSR1-RC1 phases from Lambert A data set -------;

*Page test for SSR1-RC1 phases in Lambert A data set----------------------------------------;

DATA L_1; SET Table1;

L1 =
14*sum1+13*sum2+12*sum3+11*sum4+10*sum5+9*sum6+8*sum7+7*sum8+6*sum9+5*sum10+4*sum11+3*sum12+2*sum13+1*
sum14;   /*Page L stat */
m = 4;
n = 14;
n1= n+1;
p = (n1)**2;        /* n+1 squared */
```

```
q = n**2;           /* n squared */
q1 = n**2 - 1;      /* n squared -1 */

Chi1= ((12*L1 - 3*m*n*p)**2)/((m*q)*q1*n1);
chi_p1 = probchi(Chi1,1);
z1 = sqrt(chi1);    /* 95% of z CI for L1  */

z1_lower = z1-1.96;  /* Lower bound of z    */
z1_upper = z1+1.96;  /* Upper bound of z    */

PROC PRINT DATA=L_1; RUN;

*Page test for SSR2-RC2 phases in Lambert A data set----------------------------------------;

*Rank data in SSR2-RC2 phases from SAS data set A_SSR2_RC2 of Lambert A data set -------------;

PROC TRANSPOSE DATA=A_SSR2_RC2 OUT=Table2;
/* transpose the data matrix in order to rank scores*/
     ID id;
RUN;

PROC RANK DATA=Table2 OUT=Table2;
     VAR A1-A4;RUN;

PROC TRANSPOSE DATA=Table2 OUT=Table2 PREFIX=rank;   /* transpose the ranked data back */

* Compute total ranks for 17 sessions in SSR2-RC2 phases from SAS data set A_SSR2_RC2 of Lambert A
data set ----------;

PROC MEANS DATA=Table2;                          /* compute the total of rank23 to rank31 */
     VAR rank1-rank17;
     OUTPUT OUT=Table2 SUM=sum15-sum31;

PROC PRINT DATA = Table2; RUN;

* Compute Page L, chi-square, z and CI of z for SSR2-RC2 phases from Lambert A data set --------;

DATA L_2; SET Table2;

L2 =
17*sum15+16*sum16+15*sum17+14*sum18+13*sum19+12*sum20+11*sum21+10*sum22+9*sum23+8*sum24+7*sum25+6*sum2
6+5*sum27+4*sum28+3*sum29+2*sum30+1*sum31;   /*Page L stat */
m = 4;
n = 17;
n1= n+1;
p = (n1)**2;        /* n+1 squared */
q = n**2;           /* n squared */
q1 = n**2 - 1;      /* n squared -1 */

Chi2= ((12*L2 - 3*m*n*p)**2)/((m*q)*q1*n1);
chi_p2 = probchi(Chi2,1);
z2 = sqrt(chi2);    /* 95% of z CI for L2  */

z2_lower = z2-1.96;  /* Lower bound of z2    */
z2_upper = z2+1.96;  /* Upper bound of z2    */

PROC PRINT DATA=L_2; RUN;

*Page test for RC1-SSR2 phases in Lambert A data set----------------------------------------;

*Rank data in RC1-SSR2 phases from SAS data set A_RC1_SSR2 of Lambert A data set -------------;

PROC TRANSPOSE DATA=A_RC1_SSR2 OUT=Table3;
/* transpose the data matrix in order to rank scores*/
     ID id;
RUN;

PROC RANK DATA=Table3 OUT=Table3;
     VAR A1-A4;
```

```
PROC TRANSPOSE DATA=Table3 OUT=Table3 PREFIX=rank;   /* transpose the ranked data back */

* Compute total ranks for 14 sessions in RC1_SSR2 phases from SAS data set A_RC1_SSR2 of Lambert A
data set ----------;

PROC MEANS DATA=Table3;                          /* compute the total of rank9 to rank22 */
     VAR rank1-rank14;
     OUTPUT OUT=Table3 SUM=sum9-sum22; RUN;

PROC PRINT DATA = Table3; RUN;

* Compute Page L, chi-square, z and CI of z for SSR2-RC2 phases from Lambert A data set --------;

DATA L_3; SET Table3;

L3 =
1*sum9+2*sum10+3*sum11+4*sum12+5*sum13+6*sum14+7*sum15+8*sum16+9*sum17+10*sum18+11*sum19+12*sum20+13*s
um21+14*sum22;     /*Page L stat */
m = 4;
n = 14;
n1= n+1;
p = (n1)**2;        /* n+1 squared */
q = n**2;           /* n squared */
q1 = n**2 - 1;      /* n squared -1 */

Chi3= ((12*L3 - 3*m*n*p)**2)/((m*q)*q1*n1);
chi_p3 = probchi(Chi3,1);
z3 = sqrt(chi3);    /* 95% of z CI for L3  */

z3_lower = z3-1.96;  /* Lower bound of z3    */
z3_upper = z3+1.96;  /* Upper bound of z3    */


PROC PRINT DATA=L_3; RUN;
```

Appendix B: Assessing Overlap in Lambert-A Data Using a Web-Based Calculator (Vannest, Parker, & Gonen, 2011) at http://singlecaseresearch.org

Figure B1. Web-based calculator for single-case studies developed by Vannest et al. (2011)

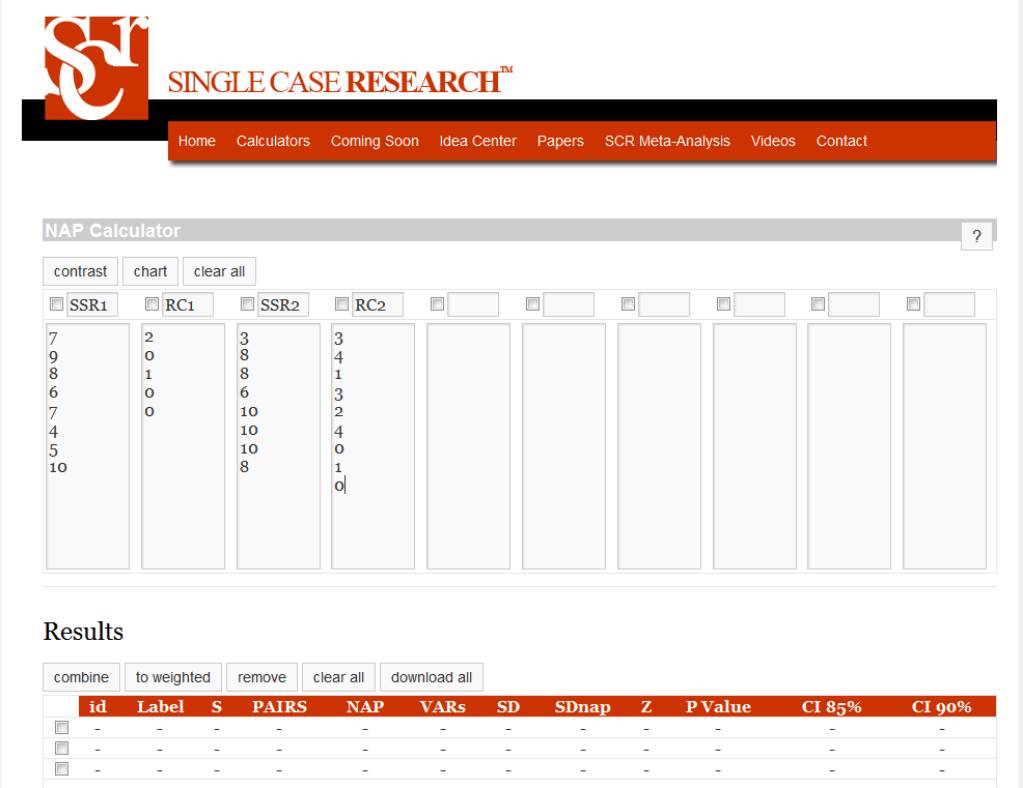

Figure B2. Data entry for student A1 of Lambert-A data set in web-based calculator for single-case studies developed by Vannest et al. (2011)

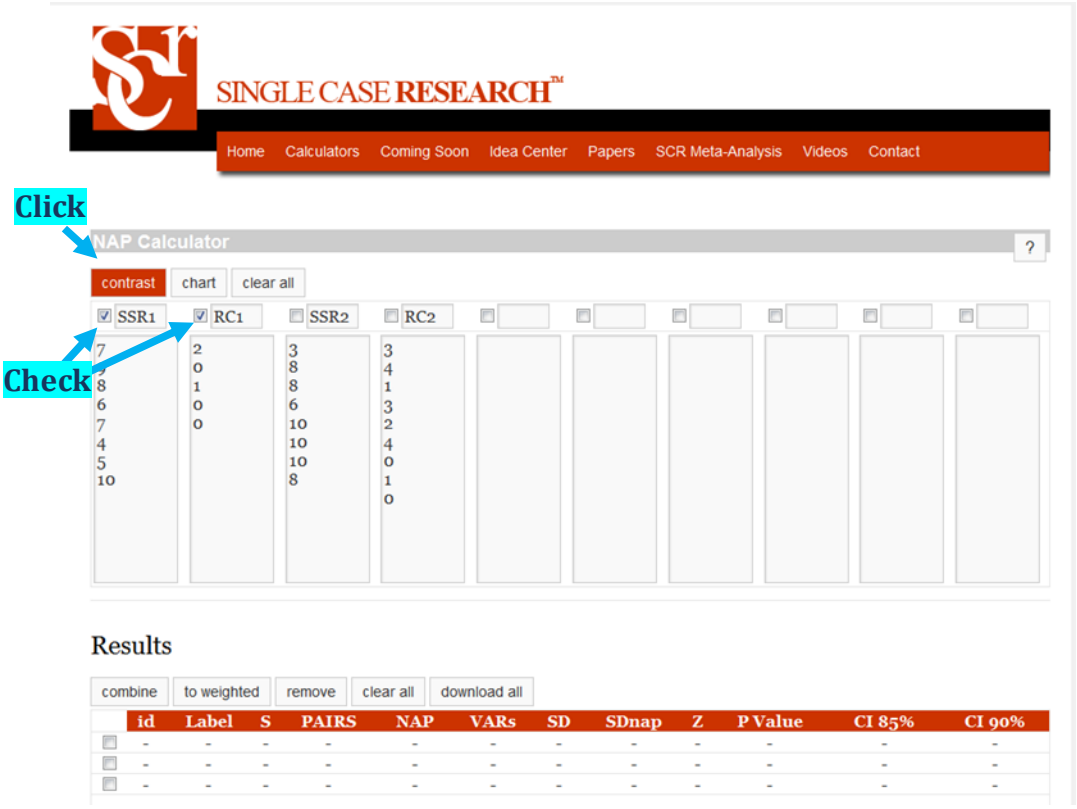

Figure B3. Compute NAP of SSR1-RC1 for student A1 of Lambert-A data set in web-based calculator for single-case studies developed by Vannest et al. (2011)

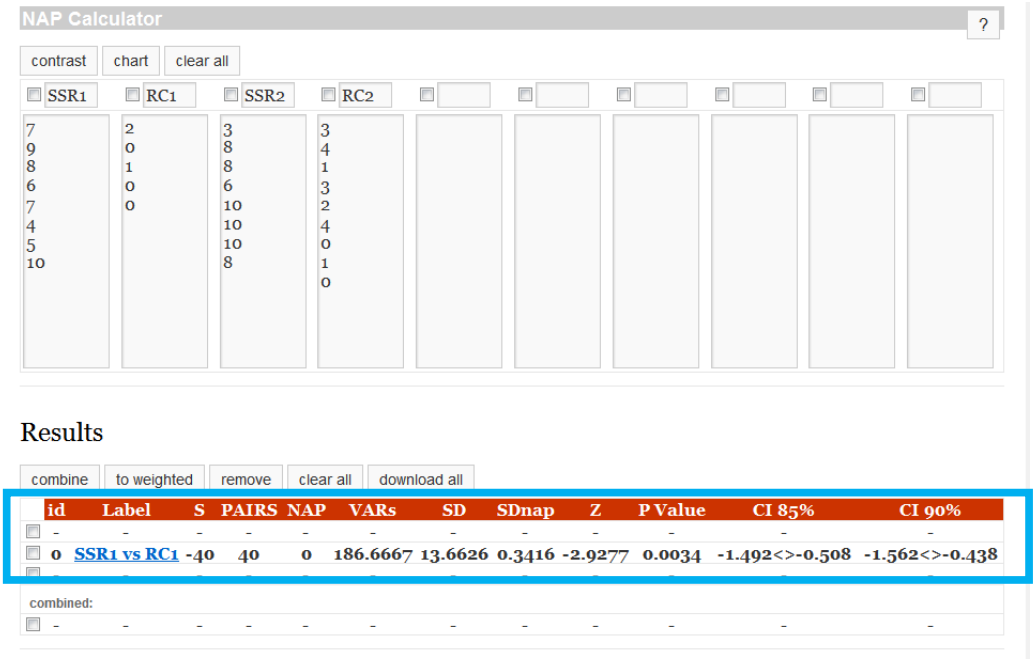

Figure B4. Obtain NAP of SSR1-RC1 for student A1 of Lambert-A data set from web-based calculator for single-case studies developed by Vannest et al. (2011)

Appendix C: Assessing Overlap in Lambert-A Data Using SPSS (Version 21)

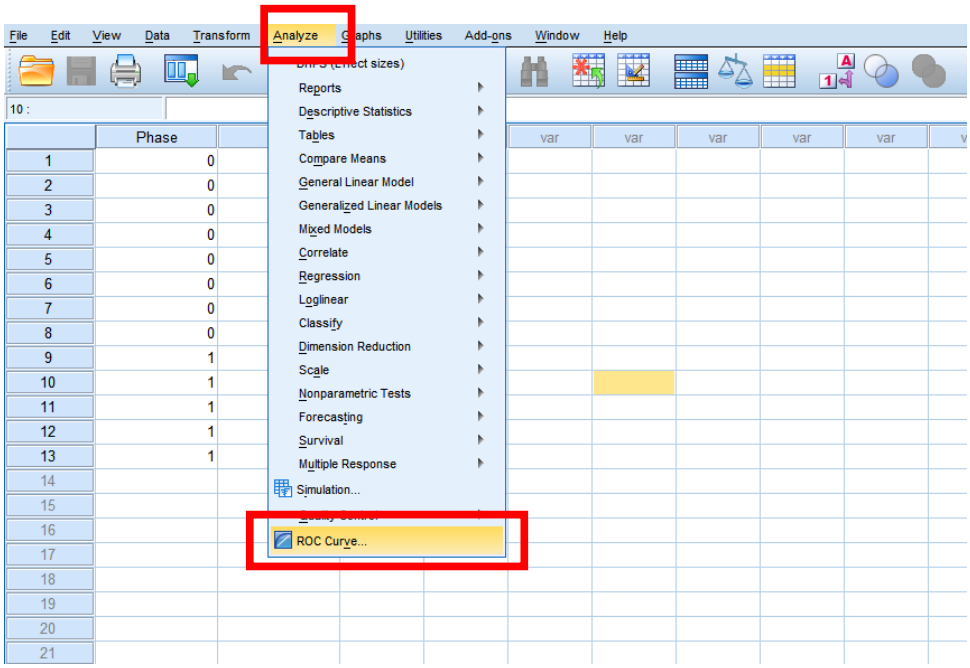

Figure C1. Compute NAP using SPSS Receiver Operator Characteristics module

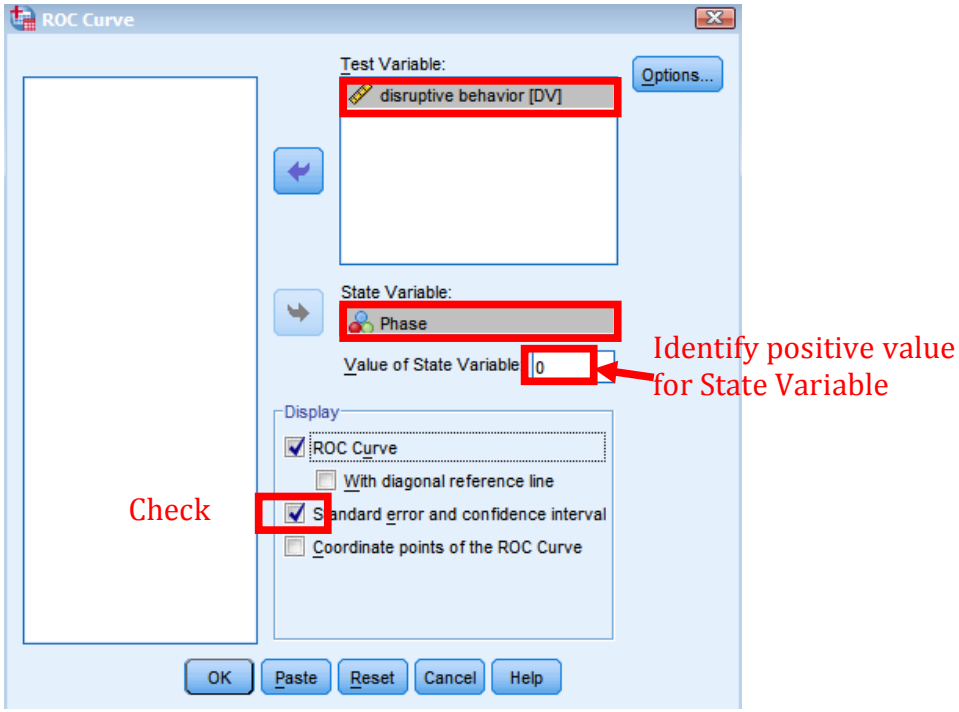

Figure C2. Dialogue window after selecting ROC Curve

Area Under the Curve

Test Result Variable(s): disruptive behavior

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
NAP 1.000	.000	***p*-value** .003	1.000	1.000

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Figure C3. Obtain NAP of SSR1-RC1 for student A1 of Lambert-A data set from SPSS 21.0
