Bottom-Up Analysis of Single-Case
Research Designs

Richard I. Parker & Kimberly J. Vannest

VOLUME 17 NUMBER 1

JOURNAL OF BEHAVIORAL EDUCATION

JANUARY 2008

Available online
www.springerlink.com

🖾 Springer

Springer

Bottom-Up Analysis of Single-Case Research Designs

Richard I. Parker · Kimberly J. Vannest

Abstract This paper defines and promotes the qualities of a "bottom-up" approach to single-case research (SCR) data analysis. Although "top-down" models, for example, multi-level or hierarchical linear models, are gaining momentum and have much to offer, interventionists should be cautious about analyses that are not easily understood, are not governed by a "wide lens" visual analysis, do not yield intuitive results, and remove the analysis process from the interventionist, who alone has intimate understanding of the design logic and resulting data patterns. "Bottom-up" analysis possesses benefits which fit well with SCR, including applicability to designs with few data points and few phases, customization of analyses based on design and data idiosyncrasies, conformation with visual analysis, and directly meaningful effect sizes. Examples are provided to illustrate these benefits of bottom-up analyses.

Keywords Single-case research · Methods · Visual analysis · Effect size

Introduction

The current special series discusses meta-analyzing single-case research (SCR) designs and focuses on computing effect sizes. The purpose of this article is to distinguish a "bottom-up" approach to analysis of data from SCR designs from the "top-down" approach typified by hierarchical or multi-level modeling (HLM, MLM) (van den Noorgate and Onghena 2007, 2008), randomization (Kratochwill and Levin 2010), and complex multi-series regression, either ordinary least squares (OLS) (Allison and Gorman 1993; Huitema and McKean 2000) or generalized least squares (GLS) (Maggin et al. 2011). Both approaches can be used to interpret data

R. I. Parker (✉) · K. J. Vannest
Texas A & M University, College Station, TX, USA
e-mail: rip_parker@msn.com

Springer

from a single study and to compile data into a meta-analysis. The term "bottom-up" refers to an analytic strategy that proceeds from visually guided selection of individual phase contrasts (the "bottom") to combining them to form a single (or a few) omnibus effect size representing the entire design (the "top").

Our view is that both bottom-up and top-down approaches are viable and currently merit focused attention. This commentary highlights only the former, while acknowledging the elegance and power of the latter, which yield an omnibus, design-wide effect for an entire design through a single complex model.

The Case for Bottom-Up

Although top-down models such as MLM or HLM are gaining momentum and have much to offer, behavior analysts should be cautious about top-down approaches for four reasons. First, MLM and HLM analyses are not easily understood and may require additional training to calculate or even to consume. Second, top-down analyses are not governed by a "wide lens" visual analysis. In other words, top-down analyses might provide sophisticated options for interpreting effects, but they are not consistent with the traditional visual analyses and may not adequately capture the entire effect. Third, top-down approaches do not yield intuitive results and may result in behavior analysts reporting a score that they interpret by comparing to a potentially arbitrarily determined criterion rather than thoroughly understanding and interpreting the data. Fourth, top-down approaches remove the analysis process from the behavior analyst, who alone has intimate understanding of the design logic and resulting data patterns.

Bottom-up analysis possesses benefits that fit well with SCR, including applicability to designs with few data points and few phases, customization of analyses based on design and data idiosyncrasies, conformation with visual analysis, and directly meaningful effect sizes (Faith et al. 1996; Parker and Vannest 2009). We propose that there are four advantages to the bottom-up approach discussed herein. First, bottom-up approaches are applicable to limited designs with short data series and few phases. Second, bottom-up approaches make it possible to detail the analysis to the idiosyncrasies of a design, including mid-study design changes and unanticipated response patterns. Third, bottom-up analyses rely on simple nonparametric analyses that yield directly meaningful effect sizes that are consistent with the tradition of visual analyses. For example, percentage of nonoverlapping data (PND; Scruggs and Mastropieri 2001) is a commonly used effect size for SCR designs (Scruggs and Mastropieri 2001), likely because it is the estimate of effect that is most directly aligned with visual analyses and easily interpreted. PND has fundamental difficulties that we thoroughly discuss elsewhere (e.g., Parker et al. 2007) and cannot be used as an estimate of effect, but bottom-up approaches are also consistent with the concept of data nonoverlap that aligns with visual analyses and results in data that are proportional estimates of effect. Fourth, bottom-up approaches keep the behavior analyst in control of the numerous decisions needed to yield valid effect sizes. The last of these four points is the most general, and supported by the other three.

When the "bottom-up" analysis employs a distribution-free effect size, such as nonoverlap of all pairs (NAP; Parker and Vannest 2009; also Burns et al. current issue and Petersen-Brown et al. current issue) or TauU (Parker et al. 2011a, b), then three additional advantages accrue: (a) the data are directly interpretable (as nonoverlap) results from phase contrasts and for overall design, (b) the calculations are easy, even by hand with short data series, and (c) behavior analysts avoid the need for advanced statistical retraining. These additional three advantages further help the behavior analyst remain in control of data analysis and interpretation.

These advantages also highlight potential pitfalls in the application of complex omnibus models to SCR designs. Imagine a worst-case scenario: (a) omnibus or top-down analytic models are built for a handful of common SCR design types, without regard for peculiarities of each study, (b) the model is expressed in sophisticated formulae not understood by the behavior analyst, (c) the behavior analyst, who alone understands the intricacies of the design and resulting data patterns, is unable to fully communicate this knowledge to a statistician, *or* (d) the intricacies of the design and data patterns are successfully communicated, but the adjusted model has uncertain validity (raises questions about data dependencies, sample size, and other assumptions), (e) the behavior analyst cannot query the appropriateness of the statistical model, as it is not visually accessible, and (f) the statistician provides the interventionist with results that are not well understood without advanced statistical training. A scenario containing all of these pitfalls together would be bleak indeed, but in reality, we have seen one or more of these problems often enough to be concerned. The statistical sophistication of omnibus models seems to be outpacing two-way communication with interventionists and behavior analysts, who are the only people with intimate understanding of the SCR design and data.

Data Analysis and Design Strength

The focus of this paper is SCR data analysis, the type that yields effect size (ES) summaries of the amount of client behavior change. The presumption that an ES also communicates the effectiveness of the intervention entails a causal inference, which is separate from the ES, and may not be warranted. This causal inference, which is the heart of conclusion validity (or internal validity), is legitimately made only with a strong design (i.e., a design that strategically rules out other reasonable explanations for the behavior improvement). Moreover, unlike group research, initial specification of a strong design does not ensure that the resulting design will be strong. Regardless of the strength of a specified design "on the shelf", the obtained data must be visually examined for unexpected patterns and for consistency of results across multiple (usually A vs B) phase contrasts. Visual scrutiny of data patterns within and across phases may lead to the conclusion that despite appropriate planning, the resulting design has degraded from strong to weak, and is unable to support a causal inference. It can happen that phase contrasts that were earmarked for ES calculation may be unusable, because of unexpected data patterns. Thus, analyses prescribed by design-type or template may need to be revised in light of actual data obtained. It is unfortunately true that any phase

contrast can yield an ES, but the ES may not reflect improvement due to the intervention.

An Example

Figure 1 shows a complex design, combining a multiple baseline design (MBD) across subjects with an ABA reversal design in the top three tiers—Adam, Bob, and Carol. A fourth tier (Dan) was intended as a control only. Added to the graph are two nonstandard symbols. Short horizontal dashed lines represent the segment of a baseline which serves as "concurrent baseline control." The second symbol is a downward arrow showing the intervention onset point controlled by the "concurrent baseline control" below it.

The design was planned to have strong conclusion validity (one onset and reversal in each of three tiers, and each intervention onset [arrows] having a concurrent baseline control underneath it). However, visual analysis of data patterns shows that these well-laid plans were thwarted. There is no reversal noted from B to A_2, at least not for Adam and Bob, so a reversal design analysis is not warranted. Instead, an analysis plan must be individually constructed for these data.

Our visual analysis turns next to the MBD across subjects, composed of three A_1B contrasts with a control series. Having discarded the reversal design logic, does strong MBD design logic still survive? Yes, because (a) the three phase B onset arrows all have concurrent baselines (dotted lines) for comparison, (b) none of those control segments show undesired improvement, which would signal that a potentially confounding event caused client improvement, and (c) changes from A_1 to B for Adam, Bob, and Carol are all in the intended direction. Therefore, our design and data patterns permit calculation of an ES for each of the three AB

Fig. 1 Example of a complex multiple baseline design across subjects
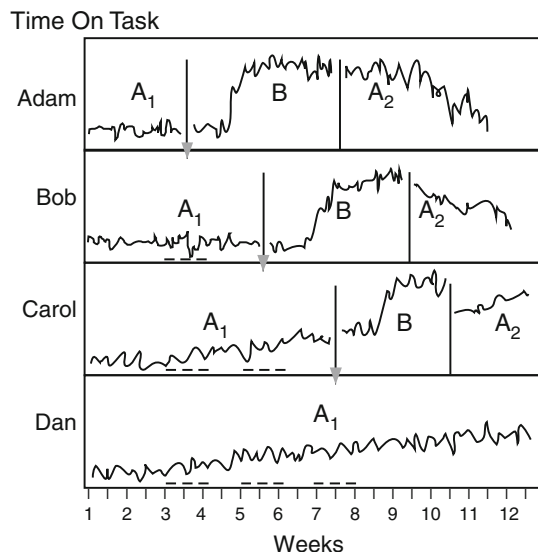
contrasts, and combining them for an omnibus, design-wide ES, but visual scrutiny has more to reveal.

A visual scan of the three B phases identifies two data patterns that have implication for analysis. First, in Carol's series, positive baseline trend exists, which would confound our inference that the performance change from A_1 to B can be explained by the treatment alone. Supporting the conclusion of Carol's phase A trendedness is the trendedness of Dan's control series. Baseline or pre-existing improvement trend is a dreaded confound and must be controlled statistically. Visual analysis also detects a second important data pattern with implications for analysis: the phase B lag (by three or four data points) in improvement, which is consistent across three tiers. This apparent "acquisition period" does not well represent amount of change due to the intervention. A fair estimate of the impact of treatment on phase B performance level (i.e., ES) would exclude the short lag periods or incorporate them in a trend or change-point model.

From the preceding visual analysis, a bottom-up analytic model is finally built. It is composed of three A versus B contrasts, from Adam, Bob, and Carol, each of which has a concurrent baseline control, so we claim high conclusion validity for the omnibus result. The statistical model for each contrast can be a level difference only, with two provisos. First, we need to set a rule (and report to readers) to exclude phase B data in the first three observations (acquisition period). Second, the baseline trend must be controlled in Carol's baseline. The distribution-free TauU is suitable for these tasks.

After the three A versus B contrasts are calculated, leading to a TauU ES for Adam, Bob, and Carol, the three ES are combined. The most defensible combination method is to take their average, after weighting each one by the inverse (reciprocal) of its variance ($Wt_{bob} = 1/Var_{bob}$) (Hedges and Olkin 1985; Sanchez-Meca and Marin-Martinez 2010). So Bob's weighted ES is $ES_{bob}*Wt_{bob}$. This weights more heavily the ES with the lowest bounce in its series, which also tend to be longer series. The three weighted ES are added together, and their sum is divided by the sum of the weights. The result is an averaged, overall, or omnibus TauU ES. The final step is calculation of a standard error ($SE_{omnibus}$) of this omnibus ES. The most defensible calculation is to take the root of the inverse of the sum of the weights $SE_{omnibus} = sqrt(1/(Wt_{adam} + Wt_{bob} + Wt_{carol}))$ (Borenstein et al. 2007). With this SE, a p value and confidence intervals can be calculated for the omnibus ES.

In summary, visual analysis is required to guide the ES calculation process. First, visual analysis can show which design logic has survived data collection. In the example, a combined MBD and reversal design was planned, but the reversal design was nullified by results. Visual analysis also confirmed consistency of results across Adam, Bob, and Carol, which is required for a valid MBD design that permits conclusion validity. Visual analysis detected which phase contrasts were legitimate and could be combined together for an omnibus ES. It also told us what level of conclusion validity that omnibus ES would have (strong in our example). Visual analysis also determined that a level shift analysis was sufficient for Adam and Bob, but not for Carol, who required a unique analytic strategy to control baseline trend.

The example design and data demonstrate a desirable sequence for bottom-up analysis of SCR data guided by visual analysis: (a) implementation of a design that

was planned to possess medium or strong conclusion validity, (b) broad lens visual analysis of response patterns across phases to determine whether the planned design was supported, nullified, or supported in a modified form, (c) selection of particular phase contrasts, following the logic of the supported design, (d) visual scanning to note consistency of responding across the chosen phase contrasts, and (e) narrowly focused visual analysis of response patterns within phases to determine potential influence by external events and to choose a fair ES calculation method. Following these five steps (driven by visual analysis), an ES may be calculated, which has a known degree of conclusion validity. These final calculations are described next.

Effect Size Calculation

Perhaps the most flexible ES index for bottom-up analysis is TauU (Parker et al. 2011a, b). First, it is nonparametric, distribution-free, and suitable for data with any distribution shape and for any type of scale. Second, it has strong statistical power (at least 91–95 % that of OLS regression), so is suitable for even short data series. Third, it permits statistical control of potentially confounding baseline trend, if it exists. Fourth, it is congruent with traditional visual analysis, as it is based on data nonoverlap between phases.

TauU is easily calculated by hand for very short data series, but for designs such as our example, a Kendall's Rank Correlation module should be used. Among our favorites is the inexpensive StatsDirect (2002), designed for medical research. Also suitable is the web-based calculator from the open-source *R* statistics (Hornik 2012) program. Both of these options require a few hand calculations on the statistical program output (Schwarz 2007), as detailed by Parker et al. 2011a, b. More user-friendly is the dedicated web-based calculator offered by our SCR research group at http://www.singlecaseresearch.org.

The TauU ES for each selected phase contrast are saved, along with their respective standard errors (SE). The final step is to combine those ES and their SE for one omnibus ES and SE, which is accomplished by a meta-analysis or multi-tier analysis statistical module. Contrary to its name, an omnibus ES will not always make use of all possible phases or phase contrasts, as seen in our example, and more than one omnibus ES may need to be calculated from a single design's data in order to answer two or more research questions. The statistical procedure for hand calculation was described in detail above. Here, we rely on real data, and menu-based calculations, which readers can replicate.

Analysis Results

Manufactured data to roughly represent the Fig. 1 example are the following: Adam, (3, 2, 2, 3, 2/2, 3, 2, 4, 5, 6, 5, 7/6, 5, 4, 5, 3); Bob, (2, 3, 2, 3, 2, 3, 3/2, 3, 3, 3, 5, 7, 6, 8/7, 7, 6, 7, 6, 5, 4, 5); Carol, (1, 1, 2, 2, 3, 2, 2, 3, 3/3, 3, 3, 5, 4, 6, 6, 7, 7/4, 5, 5, 6, 5, 6); and Dan, (2, 3, 1, 2, 3, 1, 1, 3, 4, 2, 3, 4, 2, 3, 4, 3, 3, 3, 4, 3, 4, 4, 4, 5, 4). Using the <singlecaseresearch.org> TauU calculator, we input phase A and B data for Adam, Bob, and Carol, in turn. We delete the first three phase B data points

(acquisition period) for all clients. The web-based calculation gives us the following TauU ES (with SE): Adam, 1.0 (.365); Bob, .886 (.352); Carol, .593 (.314). For Carol's analysis only, we click the "control baseline" option. Without this option, her TauU would be much higher (i.e., 1.0). The three ES can now be combined arithmetically, by formulas presented above. Or we can use the multi-tier analysis program, WinPepi (Abramson 2012). Within WinPepi, we select the menu: "Compare2"≫"i. Any of the above for multi-strata"≫"Other, with SE/CI"≫select "Standard Error". Then enter each ES and its SE as one stratum and click "next stratum." Repeat until all three strata have been entered then click "All strata" to obtain the omnibus TauU ES with its SE. The 90 % confidence intervals around the omnibus TauU are also provided.

For our example data, overall TauU = .80, with SE = .20, which is highly significant (<.000). So across all phase contrasts included, 80 % of treatment scores exceeded all baseline scores. This design-wide 80 % nonoverlap between phases is believed to represent the treatment effect, because it was obtained from a design with three AB phase contrasts, each of which was validated by a baseline in a separate series, concurrent with the intervention onset.

Issues in Bottom-Up Analysis

Next are discussed some issues or unresolved questions in conducting a bottom-up analysis. In fact, these same questions also apply to top-down analyses, but for bottom-up, they are more sharply defined.

Which Phase Contrasts in a Reversal Design?

A reversal design such as $A_1B_1A_2B_2$ contains three adjacent phase contrasts: A_1 versus B_1, B_1 versus A_2, and A_2 versus B_2. All three contrasts must contribute to the predicted response pattern (weak, strong, weak, strong) in order for the design to possess conclusion validity. But should all three contrasts be included in the calculation of the omnibus effect size for the ABAB design? The question is whether B_1 versus A_2 (the reversal) measures the same sort of intervention effect as does A_1 versus B_1. When behavior does not fully reverse (regress to baseline levels), B_1 versus A_2 can be small, even when treatment onset (A_1 vs B_1, A_2 vs B_2) causes a large change. This question has not been well discussed in the literature, but at present it seems best to omit the reversal contrast (B_1 vs A_2) from effect size calculation.

Figure 2 shows an ABABAB reversal design, which demonstrates substantial maintenance of improvement after treatment is withdrawn. The A versus B contrasts consistently show greater behavior change than do the B versus A contrasts, although both are consistent. A gradual improvement trend is noted through all phases due to maintenance.

Effect sizes are intended to communicate positive change due to treatment implementation. They are not intended to represent deterioration in behavior when the treatment is removed—that is a conclusion validity concern. In this example,

Fig. 2 Example of ABABAB reversal with increasing maintenance of improvement
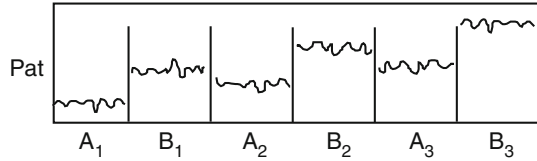
including B versus A contrasts in the calculation of an omnibus ES would seriously underestimate the improvement effect. This issue deserves discussion.

How to Control Confounding Baseline Trend?

Pre-existing improvement trend in baseline clearly confounds the inference that the treatment caused the behavior change. Statistical control of baseline trend is possible, reflected in the extended celeration line model of White and Haring (1980), and the regression model of Allison and Gorman (1993), among other models. These two models have been most popular, both controlling by trend line slope. The baseline slope is subtracted (semipartialled) from the entire data series, and then the main analysis is conducted on the remaining or residual scores. However, trend line slope control faces three major criticisms, (a) the ES reduction undesirably depends on the length of phase B; a long enough B phase will over-control trend, reducing the ES to zero or to negative (deteriorating) values, (b) it is not unusual for phase A trend control to project impossible future performance, beyond the limits of the score scale, casting doubt on the whole procedure, and (c) trend line slope control is conducted without regard for the reliability of the trend line. A lengthy phase A trend with many values tightly clustered around the trend line (low variability and high reliability) will be treated the same as a very short, very bouncy and hence very unreliable baseline trend line.

Figure 3 graphically depicts three weaknesses of traditional (Allison et al. 1996; White and Haring 1980) control of positive linear baseline trend. The top graph contains original data, with a potentially confounding improvement trend in baseline. The baseline trend line is extended through phase B. The bottom graph shows the result of removing phase A trend from the entire data set. Although this is a fabricated (drawn) line graph, the change between the top and bottom graphs is accurate, using a graph rotation (GROT; Parker et al. in press) method which exactly replicates the Allison et al. (1996) trend control method.

This example shows three undesirable features, very short baseline phase, highly variable baseline trend, and long treatment phase. Controlling this short, unreliable phase A trend would excessively impact the long treatment phase, forcing the strange conclusion that controlling for baseline trend, the client severely deteriorated in phase B. This extreme example would also lead us to conclude that by the middle of phase B, the client's behavior would be so severe as to fall off the bottom of the score scale. Such extreme examples are not rare and call for both cautious application of trend control and alternative methods of doing so.

Alternative methods of controlling baseline trend exist, including (a) control not of trend, but of the *proportion of variance* (R^2) accounted for by trend (yielding

Fig. 3 Three weaknesses of traditional control of positive baseline trend
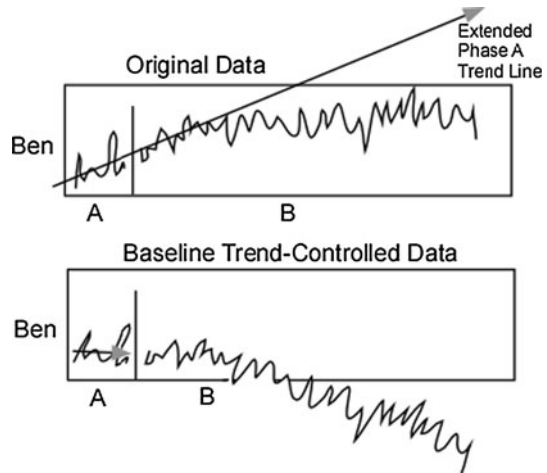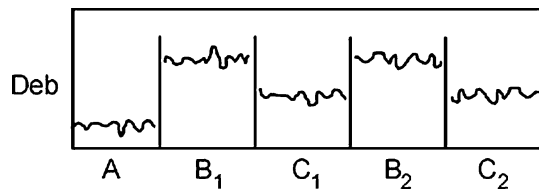

Fig. 4 Example of alternating treatment design

more moderate results), (b) control of trend by *amount of variance* accounted for by phase A trend (also more conservative), and (c) control not of linear trend, but of nonlinear monotonic trend (much more conservative). TauU employs two of those options. First, it controls not linear, but monotonic trend. Second, it subtracts a finite amount of baseline trend, rather than projecting it out into the future as a vector. This issue of how to best control baseline trend was raised over a decade ago (Scruggs and Mastropieri 1994, 1998, 2001), but has not benefitted from recent debate in the published literature.

Adjacent Contrasts Only?

The cautious approach to selecting pairs of phases for contrast analysis is to restrict analysis to adjacent phases. But cases exist where design logic and data patterns suggest a more liberal approach. Figure 4 portrays an alternating treatment design in which, for ethical or pragmatic reasons, there was no attempt to return to baseline. The question is, which contrasts are warranted? At minimum, A versus B_1 is acceptable, but that would include only 40 % of available data. The broader A versus $(B_1 + B_2)$ contrast would be desirable because the additional B_2 data reduce its measurement error. The best argument for this broader contrast seems that B_1 and B_2 response patterns are so similar.

Another contrast we would like to be able to make is A versus C_1 or even A versus $(C_1 + C_2)$, but note that there is no A phase contiguous with C. There seems

to be justification for contrasting nonadjacent pairs (assuming, of course that the contrast follows the design logic) in cases like the Fig. 4 example, where responses in C have demonstrated independence from B. In this example, independence is noted by a repeated pronounced drop in behavior on switching from B to C. A counter-argument to accepting the A versus C_1 contrast and ES is that in this design, the effects in C are always contaminated by the effects of B. Adherents to the adjacent contrasts only rule would likely find the above design flawed and recommend replacement with ABCACB series that permits clean AB and AC adjacent contrasts. An even more critical conservative point of view would require two separate, shorter data series, ABC and ACB, to permit adjacent contrasts only.

The conservative adjacent contrasts only rule has some traction (Horner et al. 2005), and its caution is appealing. However, other researchers have recommended using all phases of a design to compute ES (Maggin et al. 2011). One should not discard data without reason, but it is difficult to see how all phases can always have a legitimate role in bottom-up analysis of many designs. To conclude, whether to permit contrasts other than for adjacent phases is an unresolved issue, and one that has received little attention in the literature.

Conclusion

This paper described a bottom-up analytic strategy for SCR design data, describing its major benefit as a very close linkage to, and governance by visual analysis. Additional benefits accrued from use of a new, powerful TauU (nonoverlap+trend) ES. The bottom-up strategy is distinct from a top-down strategy in which an overall or omnibus analytic model is fit to the entire design. Whereas top-down appears more elegant, it entails a marked risk, which is to ignore the idiosyncrasies or uniqueness of a design and its data patterns. It is true that any template can be modified, but to do so in HLM, for example, requires statistical skills beyond those of most interventionists. This raises the broadest concern with the top-down analytic strategy because the behavior analyst is not able to maintain decision-making control, may not even be able to confirm the legitimacy of a model fit, and may not even be able to interpret the results. Sophisticated, omnibus analytic models appear insufficiently governed by thoughtful visual analysis by the informed interventionist. Visual and statistical analysis becomes disconnected, with risk to the validity of results and conclusions.

In group research, standard designs are typically followed, and the resulting data can be analyzed almost automatically, or by template. Knowing only the design and procedures, one can ascribe a level of conclusion validity. But SCR presents three major challenges not present in group research. First, interventionists employ not just a few main design types, but rather myriad variations. Behavior analysts are encouraged to blend design types for greater efficiency and to better meet local exigencies (Kazdin 2010). Second, which phases to include in an analysis cannot be known until client response patterns can be visually scrutinized. Response patterns such as delayed or lagged responding, unplanned maintenance of learning, unexpected growth trends in data, and drops or spikes in response levels in

apparent reaction to external influences all make a difference in which phases should be contrasted, and whether trend should be included. The third challenge relates to conclusion validity. Any SCR design has an "on the shelf" level of conclusion validity, based on the number of important phase contrasts, number of behavioral reversals, and number of key contrasts with concurrent baseline control, but the actual validity of the design may be considerably weaker, as it also depends on data patterns and consistency. Only visual analysis can tell the final strength of the design. The visual analyst looks first at response patterns within and between phases. Next, the analyst looks for consistency in responding across important contrasts. Lack of consistency in responding can invalidate a design. The need for continued visual analysis to establish level of conclusion validity is unique to SCR.

The debate on visual versus statistical analysis of 10 years ago has matured to a search for how visual and statistical analysis can be mutually complementary. Our contention is that visual analysis must govern statistical analysis and that the complex analytic process best remains in the hands of the behavior analyst. Moreover, it is helpful to invoke a simple, accessible, and intuitive ES index such as TauU. Visual analysis can inform us about which data to analyze, how to analyze it, and even what conclusion validity the omnibus ES will have. The bottom-up analytic model, which, informed by visual analysis, proceeds additively, contrast-by-contrast, seems most likely to keep the knowledgeable behavior analyst in control.

References

Abramson, J. H. (2012). Programs for epidemiologists-Windows version (WinPepi) [Computer software]. Retrieved from http://www.brixtonhealth.com/pepi4windows.html.

Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behavior, Research, and Therapy, 31*, 621–631.

Allison, D. B., Silverstein, J. M., & Gorman, B. S. (1996). Power, sample size estimation, and early stopping rules. In R. D. Frandlin, D. B. Allison, & B. S. Gorman (Eds.), *Design and analysis of single case research* (pp. 335–371). Mahwah, NJ: Lawrence Erlbaum Associates.

Borenstein, M., Hedges, L. V., & Rothstein, H. (2007). *Meta-analysis: Fixed effect vs random effects*. Retrieved February 20, 2012, from http://www.Meta-Analysis.com.

Faith, M. S., Allison, D. B., & Gorman, B. S. (1996). Meta-analysis of single-case research. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 245–277). Hillsdale, NJ: Lawrence Erlbaum.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis. CITY?*. Harcourt: Academic Press.

Horner, R. J., 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*, 165–179.

Hornik, K. (2012). *The R FAQ*. Retrieved February 14, 2012 from http://cran.r-project.org/doc/FAQ/R-FAQ.html#Citing-this-document. ISBN 3-900051-08-9.

Huitema, B. E., & McKean, J. W. (2000). Design specification issues in time series intervention models. *Educational and Psychological Measurement, 60*, 38–58.

Kazdin, A. E. (2010). *Single case research designs: Methods for clinical and applied settings* (2nd ed.). NY: Oxford University Press.

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods, 15*, 122–144.

Maggin, D. M., Swaminathan, H., Rogers, H. J., O'Keeffe, B. V., Sugai, G., & Horner, R. H. (2011). A generalized least squares regression approach for computing effect sizes in single-case research: application examples. *Journal of School Psychology, 49*, 301–321.

Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping data (PAND). *The Journal of Special Education, 40*, 194–204.

Parker, R. I., & Vannest, K. J. (2009). An improved effect size for single case research: NonOverlap of All Pairs (NAP). *Behavior Therapy, 40*, 357–367.

Parker, R., Vannest, K. J., & Davis, J. L. (2011a). Nine non-overlap techniques for single case research. *Behavior Modification, 35*, 303–322.

Parker, R., Vannest, K. J., & Davis, J. L. (in press). Simple graph rotation to control baseline trend within data non-overlap. *Journal of Special Education*.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011b). Combining nonoverlap and trend for single case research: Tau-*U*. *Behavior Therapy, 42*, 287–299.

Sanchez-Meca, J., & Marin-Martinez, F. (2010). Weighting by inverse variance or by sample size in meta, analysis: A simulation study. *Educational and Psychological Measurement, 70*(1), 56–73.

Schwarz, C. J. (2007). *Sampling, regression, experimental design and analysis for environmental scientists, biologists, and resource managers*. Burnaby: Simon Fraser University.

Scruggs, T. E., & Mastropieri, M. A. (1994). The utility of the PND statistic: A reply to Allison and Gorman. *Behaviour Research and Therapy, 32*, 879–883.

Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single-subject research: Issues and applications. *Behavior Modification, 22*, 221–242.

Scruggs, T. E., & Mastropieri, M. A. (2001). How to summarize single-participant research: Ideas and applications. *Exceptionality, 9*, 227–244.

StatsDirect Ltd. (2002). *StatsDirect statistical software*. England. Retrieved February 14, 2012 from http://www.statsdirect.com.

van den Noorgate, W., & Onghena, P. (2007). The aggregation of single-case results using hierarchical linear models. *The Behavior Analyst, 8*, 196–208.

van den Noortgate, W., & Onghena, P. (2008). A multilevel meta-analysis of single subject experimental design studies. *Evidence-Based Communication Assessment and Intervention, 2*, 142–158.

White, O. R., & Haring, N. G. (1980). *Exceptional teaching* (2nd ed.). Columbus, OH: Merrill.