How Can Single-Case Data Be Analyzed? Software Resources, Tutorial, and Reflections on Analysis

Rumen Manolov[1] and Mariola Moeyaert[2]

Abstract

The present article aims to present a series of software developments in the quantitative analysis of data obtained via single-case experimental designs (SCEDs), as well as the tutorial describing these developments. The tutorial focuses on software implementations based on freely available platforms such as R and aims to bring statistical advances closer to applied researchers and help them become autonomous agents in the data analysis stage of a study. The range of analyses dealt with in the tutorial is illustrated on a typical single-case dataset, relying heavily on graphical data representations. We illustrate how visual and quantitative analyses can be used jointly, giving complementary information and helping the researcher decide whether there is an intervention effect, how large it is, and whether it is practically significant. To help applied researchers in the use of the analyses, we have organized the data in the different ways required by the different analytical procedures and made these data available online. We also provide Internet links to all free software available, as well as all the main references to the analytical techniques. Finally, we suggest that appropriate and informative

[1]Universitat de Barcelona, Spain
[2]State University of New York, New York City, USA

Corresponding Author:
Rumen Manolov, Departament de Metodologia de les Ciències del Comportament, Facultat de Psicologia, Universitat de Barcelona, Passeig de la Vall d'Hebron, 171, 08035 Barcelona, Spain.
Email: rrumenov13@ub.edu

data analysis is likely to be a step forward in documenting and communicating results and also for increasing the scientific credibility of SCEDs.

The current article aims to bring analytical developments closer to applied researchers conducting single-case studies by presenting and illustrating free software available for carrying out the analysis—software that is accompanied by free 270-plus-page tutorial.[1] First, we offer a brief presentation of single-case experimental designs (SCEDs). Second, we stress the need to focus on data analysis, including both visual and statistical analyses, and we explain why the current article represents a step forward in this topic. Third, we provide an illustration, applying several analytical techniques to a real dataset, with the latter being selected as representative of SCED studies.

Sustained and Increased Attention to SCEDs

The three main characteristics of an SCED (focus on one entity, repeated measures across time, and experimental control; Kratochwill et al., 2010) can be used to define SCEDs in a broader research context, which might help understanding how these designs differ from closely related designs. In terms of the focus, unlike group-comparison design study dealing with average treatment effect estimates, SCED studies focus on a limited number of prese-lected individuals and subject-specific treatment effect estimates are obtained (Barlow, Nock, & Hersen, 2009; Kratochwill & Levin, 2014). Moreover, unlike the most common between-group studies, SCEDs make possible iden-tifying within-subject trends. In terms of experimental control, SCEDs do not always entail random assignment of measurement occasions to treatments. However, a SCED should not be confounded with a (qualitative) case study (Blampied, 2000) or observational case study research, as in these latter types of studies there is not a purposeful manipulation of an independent variable nor are there necessarily repeated measures.

From an applied perspective, the relevance and recognition of the use of SCEDs has been made evident from the number of papers dedicated recently to the topic in disciplines as varied such as speech-language pathology (Byiers, Reichle, & Symons, 2012), pediatric psychology (Cohen, Feinstein, Masuda, & Vowles, 2014), education (Plavnick & Ferreri, 2013), technology-based medical interventions (Dallery, Cassidy, & Raiff, 2013), sport psychology

(Gorczynski, 2013), rehabilitation (Graham, Karmarkar, & Ottenbacher, 2012), and group work (Macgowan & Wong, 2014). This is well aligned with the recognition of SCEDs as a means of obtaining evidence about interventions (Howick et al., 2011). Accordingly, the conclusion of Smith's (2012) review of 409 studies was that "recently published SCED research is largely in accordance with contemporary criteria for experimental quality" (p. 510).

From an academic perspective, the salience of SCEDs is also illustrated in the publication in the last decade of revised editions of classical and major reference books on SCED methodology and analysis, such as the ones by Barlow et al. (2009), Kazdin (2011), Gast and Ledford (2014), and Kratochwill and Levin (2014). Moreover, there have been an important number of journal special issues dedicated to SCEDs from a methodological and/or analytical point of view (e.g., *Journal of Behavioral Education, Journal of Applied Sport Psychology, Remedial and Special Education* in 2013, *Journal of School Psychology* in 2014, *Neuropsychological Rehabilitation*). Accordingly, the U.S. Institute of Education Sciences (2014) continues to show interest in funding research related to single-case methodology.

SCED Researchers' Data-Analytical Practices

The salience and wider acceptance of SCEDs as a valid methodology for obtaining scientific evidence has been translated in methodological advances such as standards (Kratochwill et al., 2010; Smith, 2012), methodological quality scales (Tate et al., 2013), and further recommendations (Horner et al., 2005; Ledford & Gast, 2014). Concrete recommendations about reporting SCED studies are also available (Tate et al., 2016). In terms of data analysis, it has been recommended that objective summary measures be used for documenting results, communication across researchers and meta-analysis (Busse, Kratochwill, & Elliott, 1995; Jenson, Clark, Kircher, & Kristjansson, 2007; Kromrey & Foster-Johnson, 1996), but this is not always the case, given the strong predominance of visual analysis (Kratochwill & Brody, 1978; Parker et al., 2005; Perdices & Tate, 2009). In that sense, it is possible that apart from methodological improvements, such as randomization (Kratochwill & Levin, 2010), the scientific credibility of SCEDs could be boosted by improving the analytical practice of applied researchers and practitioners in their everyday life work.

Several reviews have been performed on the way in which SCED data are analyzed. Parker and Brossart (2003) performed an informal review of SCED articles in counseling, clinical and school psychology journals in the period 1987-2002, reporting that visual analysis was used in absence of statistical analysis in more than 65% of articles. The review of Perdices and

Tate (2009) in the field of neuropsychological rehabilitation in the period 1991-2008 showed that 78% of the articles reported graphed data, 64% used some kind of statistical analysis, and 26% used visual analysis alone. Smith (2012) reviewed articles from peer-reviewed journals for the period 2000-2010, and reports that visual analysis alone is used in 21.6% of the studies using multiple-baseline designs, 17.1% of the studies using a reversal design, and 23.1% of alternating treatment designs, whereas the corresponding percentages for statistical analysis alone are 13.4, 12.9, and 7.7 and those for combined visual and statistical analysis are 6.4, 5.7, and 19.2. Considering the repeated calls to use visual and statistical analysis jointly (Fisch, 2001; Franklin, Gorman, Beasley, & Allison, 1996; Harrington & Velicer, 2015; Houle, 2009), apparently it is still necessary to stress that point, given the relatively infrequent combined use of these two types of analyses.

To complement visual analysis with quantitative analysis, the following steps could be followed: (a) create sound techniques and test their statistical properties with simulated data, and field test them with real behavioral data; (b) illustrate their application in a step-by-step fashion in the same way as some of the papers published in the abovementioned special issues (e.g., Heyvaert & Onghena, 2014; Shadish, Hedges, & Pustejovsky, 2014); (c) present the developments at conferences and workshops (e.g., Beretvas, Van den Noortgate, & Ferron, 2014; Manolov, Krasny-Pacini, Evans, & Chevignard, 2014; Shadish, Hedges, & Pustejovsky, 2013); (d) develop software and tutorials; and (e) write papers presenting the software and tutorials (e.g., Bulté & Onghena, 2009, 2012; Onghena & Van Damme, 1994). The current article corresponds to the fifth step, whereas a description of the tutorial is provided in the appendix. In the following, we present the breadth and limitations of the usefulness of the current article.

Usefulness of the Software, the Tutorial, and Current Article

We consider that there are two basic ways in which guidance can be provided to applied researchers conducting SCEDs and willing to analyze their data quantitatively. The first option is to review the evidence and discussions available on the performance and the characteristics of all/many analytical techniques, and recommend which techniques should be used when. This option would deal with the question "What is the use of a good analytical technique, if applied researchers are not aware of its existence and/or qualities?" We do mention here most of the recent developments and show how they can be applied, but we do not compare their quality or appropriateness for establishing recommendations for choosing among them in different

situations. This latter topic is dealt with in Manolov and Moeyaert (2016) and it is important for future research.

The second option is to make the application of all/many analytical techniques a feasible task by providing free software and a tutorial explaining how to use it. We have been creating such software and reviewing the software created by other authors, and we have also created such a tutorial. Therefore, in the current article we aim to illustrate the capabilities of a variety of analytical techniques with a typical SCED dataset. Moreover, the current article helps dealing with the question "What is the use of a myriad of analytical proposals if these cannot be applied easily and with no additional cost?" Thus, the purpose was to bring recent SCED analytical developments closer to applied researchers to bridge the gap between statistical advances and actual analytical practice by making applied researchers aware of the fact that there are multiple user-friendly software resources available. However, we do not claim that such information is sufficient for improving the analytical skills of researchers.

In summary, the need for software, tutorial, and an article presenting them is based on the following points: (a) the software presented is freely available; (b) the tutorial guides the application of the analytical techniques in a step-by-step fashion, relying heavily on screenshots, commenting on the results obtained, and referring the interested reader to the original literature presenting the techniques; (c) many different quantitative techniques are included in the software and tutorial and illustrated here, given that SCED data are complex, and that the necessary visual analysis deals with several aspects of the data such as assessing baseline stability, within-phase level and trend, changes in level, trend and variability, overlap, immediacy of effects, consistency of the patterns, and comparing observed and projected data patterns (Kratochwill et al., 2010; Lane & Gast, 2014), apart from taking into account whether the predictions make sense or not (Parker, Cryer, & Byrns, 2006); (d) visual analysis is also part of the software, tutorial, and current illustration, given that it can be useful for selecting an analytical technique and for evaluating whether the quantitative results obtained are intuitively meaningful (Parker et al., 2006); actually, there is a graphical representation of the data accompanying almost every quantitative procedure, to ensure that the assessment of intervention effectiveness takes into account both types of information as suggested repeatedly (Davis et al., 2013; Fisch, 2001; Smith, 2012); (e) the techniques included and illustrated show a wide variety of complexities, from simple graph rotation to building multilevel models; (f) one of the main complexities of the software are the different data structures required by the different pieces of R code/packages, but we here include an Excel file, as complementary online material, with separate worksheets

representing all the data structures required for all analytical techniques included in the tutorial (even if they are not illustrated here); (g) illustrating the different uses and types of information provided by the different techniques, we are also implicitly giving some indications regarding the crucial questions "Which techniques should be used when?" and "How should I analyze my data?"

Regarding previous papers on single-case data analysis (SCDA) software tools, most have a narrower focus than the current article. In chronological order, the following documents have been made public: Nagler, Rindskopf, and Shadish (2008) illustrated how UnGraph can be used for retrieving data from graphs, and they also show how multilevel models can be applied with SPSS and HLM software; Dixon et al. (2009) explained how to create graphs with Microsoft Excel; Bulté and Onghena (2013) presented the SCDA plugin for R; Parker, Vannest, and Davis (2014a) mentioned the WinPepi free software (http://www.brixtonhealth.com/pepi4windows.html) for computing Tau-*U* for designs beyond the basic AB (baseline phase followed by an intervention) and also for meta-analytical use; Levin, Ferron, and Gafurov (2014) and Levin, Evmenova, and Gafurov (2014) described the use of the Excel-based *ExPRT* for randomization tests; de Vries, Hartogs, and Morey (2015) presented R code for Bayesian analysis about estimating effect size and hypothesis testing; Maric, de Haan, Hogendoorn, Wolters, and Huizenga (2015) explained the use of SPSS for piecewise regression analysis, and Busse, McGill, and Kennedy (2015) mentioned a webpage (http://www.interventioncentral.org/teacher-resources/graph-maker-free-online) that graphs the data and allows computing trend, the percentage of nonoverlapping data and the standardized mean difference using the standard deviation of the baseline data in the denominator. The most general article on SCED analysis software is by Chen, Peng, and Chen (2015), published after the submission of the current article and dealing with several different types of software and analyses. Chen et al. (2015) provided less detail regarding use and interpretation of the output than the one available in the current article and in our tutorial. In that sense, Chen et al. (2015) focused more on whether the software tools function properly, whereas we deal with how to interpret their outcomes. In terms of interpretation, we rely heavily on graphical representations to aid the interpretation of the results of the techniques, given the importance of visual analysis as a way of validating quantitative results (Parker et al., 2006). In contrast, Chen et al. (2015) paid more attention to the formulaic expression of the analytical techniques. Finally, we present a tutorial created by us, which describes the use of many R scripts also developed by us and useful for implementing a variety of procedures (created or suggested by a variety of authors), apart from describing software developed by other researchers. In

summary, we consider that the Chen et al. (2015) article, together with the current text, and the tutorial we are presenting here offer sufficient information for applied researchers to know how to implement virtually any SCED analytical technique, once it has been chosen.

Method

Illustrative Data

To choose an appropriate dataset for illustrating the analytical techniques and the output of the software, we took into account the characteristics of SCED data as reported in recent reviews (Shadish & Sullivan, 2011; Solomon, 2014) and also the design requirements for making possible the demonstration of intervention effects (Kratochwill et al., 2010; Tate et al., 2013). As a result, we chose the data collected by Singh and colleagues (2007) on mindfulness training for controlling aggressive behavior in people diagnosed with several mental disorders such as depression, schizoaffective disorder, borderline personality, and antisocial personality. These data have the following characteristics: (a) a multiple-baseline design is used, which is the most common design structure (present in 54% of studies), including three cases which represent the median and modal number of cases per study (Shadish & Sullivan, 2011) and meets the design requirements from the What Works Clearinghouse Standards (Kratochwill et al., 2010); (b) the amounts of data points for each comparison between baseline and treatment condition are 16 (for Jason), 18 (for Michael), and 22 (for Tim), matching well the median number of 20 measurements found by Shadish and Sullivan (2011); the baselines have lengths of 3, 4, and 6, which matches well the finding that 54% of the baselines have less than 5 measurements, and meets current standards of a minimum of 3 (Kratochwill et al., 2010; Tate et al., 2013); (c) the number of outcomes measured per case is 2 (verbal and physical aggression), which corresponds well to the finding that most SCEDs (60%) include more than one outcome per case, and these outcomes represent the number of aggressive behaviors,[2] which is well aligned with the fact that 48% of the SCEDs use total counts as measures (Shadish & Sullivan, 2011); (d) some of the AB comparisons present baseline trend, but these trends are heterogeneous (some improving, some deteriorating), as found in the Solomon (2014) review. Thus, we can consider that the dataset analyzed here is typical for SCEDs, although the complexity of possible SCED data patterns is impossible to illustrate with a single dataset and impossible to represent in a single study. The reader interested in learning more about SCED data is referred to scholarly texts by Barlow et al. (2009), Gast and Ledford (2014), Johnston and

Pennypacker (2009), Kazdin (2011), Kennedy (2005), and Kratochwill and Levin (2014).

Data Analysis

We use the free software (mainly R packages and R code, but also web pages) that is described in the tutorial. Practically, all analyses performed require only entering the data in a specific way. We have included as supplementary material an Excel file containing the data organized in all necessary ways. The only technique requiring more than data input is the multilevel model. We provide the exact code used here as supplementary material in a text file.

Results

The results will be presented in the following order: First, we focus on the within-phase levels and comparisons of the levels in different conditions. We start inspecting visually the whole dataset, including two outcomes for each of the three participants, before we move to comment specific AB comparisons according to the data characteristics of interest and according to the aspects that visual aids and quantitative analysis deal with. Second, we take a look at within-phase trends and changes in trend across conditions, once again beginning the analysis with all the data, before moving to some comparisons between pairs of conditions (baseline and treatment data for one of the outcomes for one of the participants). Third, we present procedures that quantify changes in both level and slope. The reader will see that some of the analyses also entail a comparison between projected and actually obtained (intervention phase) data. Fourth, we comment on procedures for quantifying data overlap. All these analytical techniques (and several more) can be implemented with the free software commented in the tutorial we present here.

In the process of looking at the different aspects of the data quantifying any changes taking place, we will show that some analytical techniques (e.g., the *d*-statistic by Hedges and colleagues and multilevel models) combine results from separate AB comparisons in a more direct way than others (e.g., nonoverlap indices). For the latter, we will illustrate how such integrations can easily (although slightly more laboriously) be obtained from the separate quantifications of behavioral change.

Level

Figure 1 shows that for two outcomes of all three participants, there is a change in level consistent with the desired intervention effect. This effect is
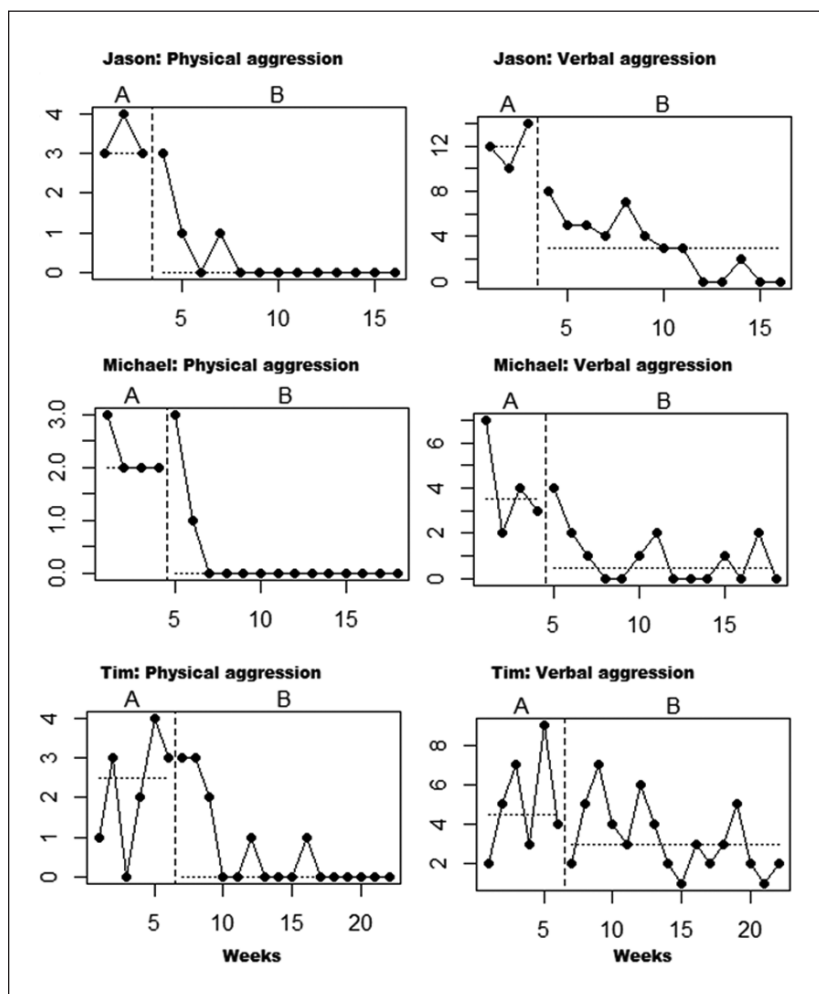
Figure 1. Graphical representation of the data obtained by Singh et al. (2007).
Note. Obtained using the SCDA plug-in for R (http://cran.r-project.org/web/packages/ RcmdrPlugin.SCDA/index.html) described in the Visual tools section of the tutorial. The dashed line represents the within-phase median. SCDA = single-case data analysis.

visually clearer for physical aggression as an outcome and for Jason as a participant.

The naked-eye visual inspection can be complemented with visual aids (Visual tools section of the tutorial), such as the standard deviation bands
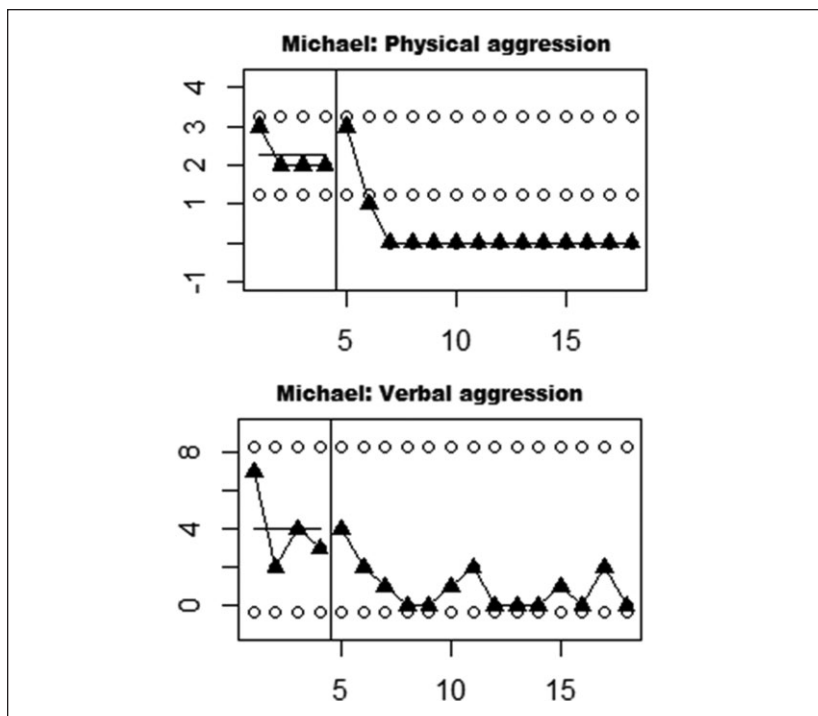
Figure 2. Graphical representation of the 2 standard deviation band superimposed on the data obtained by Singh et al. (2007) for Michael.
Note. Obtained using R code (https://dl.dropboxusercontent.com/s/elhy454ldf8pij6/SD_band.R) and described in the Visual tools section of the tutorial.

(Callahan & Barisa, 2005; Pfadt & Wheeler, 1995), which are more appropriate when the baseline does not show a clear trend. When looking into Michael's data (Figure 2), we see that the physical aggressions are lower than what can be expected by projecting the baseline mean and considering the baseline variability; for verbal aggression, this visual aid[3] suggests lack of effect. This is consistent with our general impression about effect according to the outcome stated above.

If a researcher is willing to quantify the difference in level, she or he can turn to the procedures included in the following sections of the tutorial: percentage indices not quantifying overlap, unstandardized indices and their standardized versions, or application of two-level multilevel models for analyzing data (see the appendix). One option is to use the mean baseline reduction
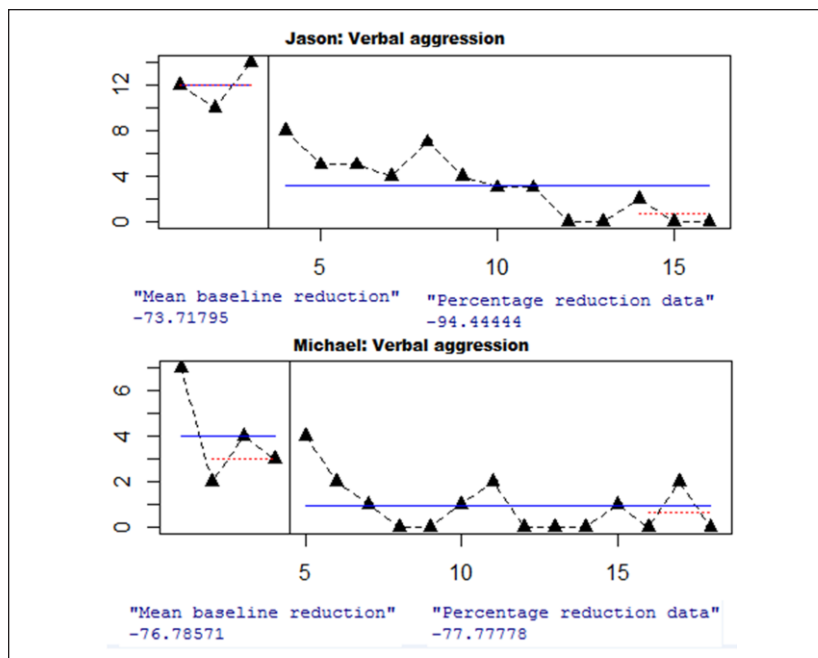
Figure 3. Graphical representation of the mean baseline reduction and percentage change index superimposed on the data obtained by Singh et al. (2007) for verbal aggressions by Jason and Michael.
Note. Obtained using R code (https://www.dropbox.com/s/wt1qu6g7j2ln764/MBLR.R?dl=0) and described in the Percentage indices not quantifying overlap section of the tutorial.

(Campbell, 2004) or the percentage change index (Hershberger, Wallace, Green, & Marquis, 1999; also called percentage reduction data, Wendt, 2009). Focusing on the verbal aggressions by Jason and Michael (Figure 3), we can see that the mean baseline reduction taking all the data into account is similar (73.7% vs. 76.8%), whereas if the quantification is based only on the last 3 points of each phase due to some substantive reason the effect is clearer for Jason (percentage reduction data 94.4% vs. 77.8%), confirming our visual impression and attaching an objective quantitative summary to it. Note that both indices convert the raw measures into percentages and thus make the results comparable (e.g., despite the fact that Michael showed initially less verbal aggressions). However, in some cases it may be justified to expect the effect to accumulate at the last (three) intervention data points, whereas the choice of the last baseline measurements may be due to the stability (if the

researcher waited for the baseline to stabilize before intervening). All values of the percentage change index (100% and 94.4% for Jason; 100% and 77.8% for Michael; 100% and 68.8% for Tim for physical and verbal aggressions, respectively) indicate that the reduction in target behavior is substantial and potentially clinically relevant. Moreover, this index confirms our visual impression that the change for physical aggression is greater.

Another option is to compute a standardized mean difference, as proposed by Hedges, Pustejovsky, and Shadish (2012, 2013). This index allows obtaining a single quantification (controlling for small sample size and for autocorrelation) for a multiple-baseline design or an AB^k design if there are at least three cases present. For physical aggression, we obtained $d = -2.25$ ($SE = 0.55$) and for verbal aggression $d = -1.44$ ($SE = 0.55$), once again reflecting the clearer effect for the former type of behavior. In comparison with the previous indices expressed in percentages, this one is expressed in standard deviations: The variability within- and between cases is taken into account. Actually, it should be noted that while the numerator of the d-statistic deals with differences in level, its denominator takes variability (another data feature which is object of visual analysis) into account. Moreover, as the d-statistic by Hedges et al. (2012, 2013) is comparable with the classical d-statistic for group designs, it could be argued that this index can be interpreted in terms of overlap, assuming normality of the distributions (Vacha-Haase & Thompson, 2004). It is also possible to combine meta-analytically the d-statistic values (Shadish, Hedges, & Pustejovsky, 2014). For merely illustrative purposes, we will here proceed, as if the physical and verbal aggression outcomes were independent[4] (a requisite for meta-analysis), although they are not as the data are obtained from the same participants. Figure 4 shows the forest plot from which we see that the weighted average $d = -1.85$ and its 95% confidence interval ranges from -2.64 to -1.05, indicating the relatively low precision of the point estimate and its statistical significance at the .05 level. The true heterogeneity observed in the effects is very small $I^2 = 7.18\%$, as it is smaller than the usual cutoff for small heterogeneity, 25%. The results obtained so far indicate that the effect of the intervention is large, at least in quantitative terms.

Immediate Change

When looking at and quantifying the level in each condition, it is possible to focus not only on average but also on the immediacy of the effect, which is one of the aspects assessed in visual analysis (Kratochwill et al., 2010). However, the data for verbal aggression for Jason seem to suggest an immediate change. Immediate change can be quantified via piecewise regression analysis (Center, Skiba, & Casey, 1985-1986), from the "Unstandardized
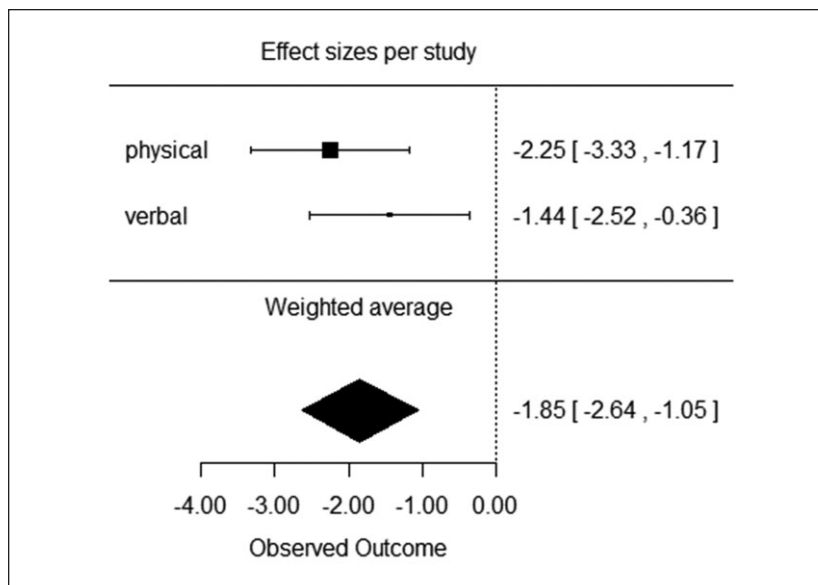
Figure 4. Graphical representation of a meta-analysis of the data obtained by Singh et al. (2007), combining the *d*-statistic values for physical and verbal aggressions across all three participants.
Note. The values of the *d*-statistic are obtained using the "scdhlm" package for R (http://blogs. edb.utexas.edu/pusto/software/), whereas the forest plot is obtained using R code (https:// www.dropbox.com/s/41gc9mrrt3jw93u/Across%20studies_d.R?dl=0) and described in the Integrating results of several studies section of the tutorial.

indices and their standardized versions" section of the tutorial, also quantifying change in slope, which is why it is also commented in the "Level and Trend" subsection later. Most of the datasets in Figure 1 show that the main reduction takes place after 1 or 2 weeks, suggesting that a change in slope might describe better the type of effect than a change in level. The graphical representation and quantification for Jason provided in Figure 9 show that the immediate effect of the intervention consists in a reduction of 7.15 behaviors. In contrast, for Michael, the decrease is estimated, according to piecewise regression, as half a behavior (−0.5). However, if we look at the data actually obtained and not at the regression lines fitted, the immediate effect of the intervention is actually a deterioration; an increase from two to three physically aggressive behaviors. Moreover, the regression trend fitted to the intervention phase data does not seem to represent them well. Thus, this quantification should also be interpreted with caution.

Trend

As the previously presented graphical representations and quantifications do not take trend into account, further analyses are necessary, as the consideration of trend might change our initial conclusions. If we turn our attention to trend (Figure 5), we can see that in all AB comparisons there seems to be a change in (ordinary least squares regression) trend with the introduction of the intervention. For Jason and Tim, flat and worsening trends give way to improving trends after the intervention (indicating intervention effectiveness), whereas for Michael the trend stops being as improving as it was before the intervention. The latter finding is related to the fact that counts lower than 0 are impossible. Note how looking at the trends makes the evaluation of the data an easier task, as complexity is reduced. However, at the same time, one should assess to what extent the trends match well the measurements obtained. Actually, we here illustrated the simplest linear model, although the type of data (with an achievable minimum of 0 may require a logistic model). Moreover, we used the ordinary least squares estimation, which is appropriate for continuous data (i.e., not counts)—more complex and potentially more appropriate options can be consulted in Shadish, Zuur, and Sullivan (2014).

Apparently, the only baseline trends that are cause for concern are the ones observed for Michael. One way of dealing with trends (included in the Visual tools section of the tutorial) is to physically rotate the graph (Figure 6) after a trisplit trend has been fitted, so that this trend is now perfectly horizontal (Parker, Vannest, & Davis, 2014b). Afterward, a quantification of choice can be computed on the rotated data (Parker et al., 2014b, suggest the nonoverlap of all pairs [NAP]; Parker & Vannest, 2009).

Comparing Observed and Projected Data Patterns

It is also possible to fit split-middle (i.e., bisplit) trend to the data and project it into the next phase to explore whether the projected and actual data are similar, taking baseline data variability into account (Manolov, Sierra, Solanas, & Botella, 2014). The illustration provided for Michael's verbal aggression (Figure 7) shows that no intervention phase measurements improve what could already be predicted from the baseline. If a quantification is desired, the percentage of data points exceeding the split-middle trend (Wolery, Busick, Reichow, & Barton, 2010) can be computed. Figure 7 shows graphically and numerically that few (28.6%) of the intervention phase data points improve the projected baseline trend. These results suggest that the effect on Michael's verbal aggressive behavior is not clear.
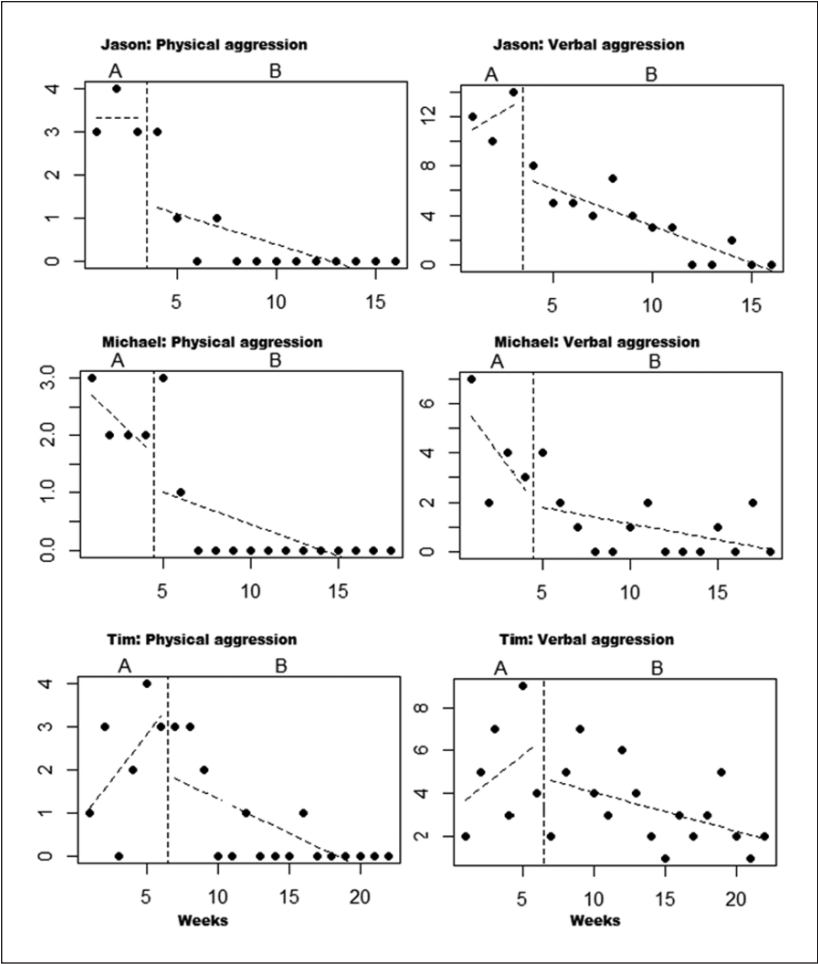
Figure 5. Graphical representation of the data obtained by Singh et al. (2007).
Note. Obtained using the SCDA plug-in for R (http://cran.r-project.org/web/packages/
RcmdrPlugin.SCDA/index.html) and described in the Visual tools section of the tutorial. The
dashed line represents the within-phase trend estimated by ordinary least squares regression.
SCDA = single-case data analysis.

Another quantification possible is using the mean phase difference (MPD; Manolov & Rochat, 2015; Manolov & Solanas, 2013b) from the "Unstandardized indices and their standardized versions" section of the tutorial. In this case, similar to the percentage of data points exceeding the split-middle trend, a trend
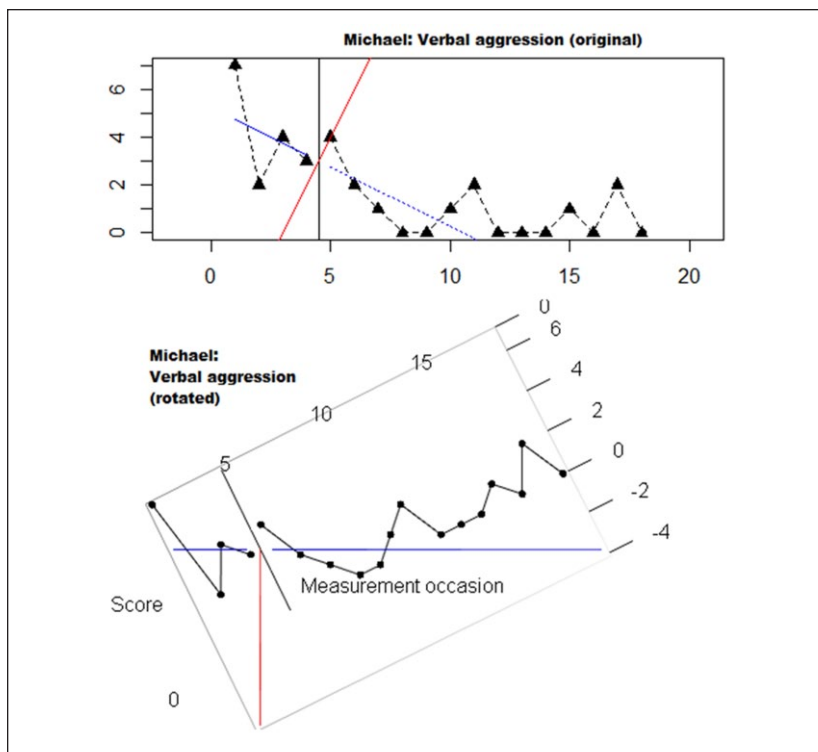
Figure 6. Graphical representation of the use of the graph rotation technique (Parker et al., 2014b) on the data obtained by Singh et al. (2007) for the verbal aggressions by Michael.
Note. The graphs are obtained using R code (https://www.dropbox.com/s/jxfoik5twkwc1q4/ GROT.R?dl=0) and described in the Visual tools section of the tutorial.

line is extended, but there are two differences: (a) in MPD trend is based on differencing not on the split-middle method, and (b) the difference between projected and actual intervention phase measurements is computed, instead of focusing on overlap. As can be seen from Figure 8, the projections for Michael (into impossible negative values, as was the case using split-middle trend) and for Tim differ from the actual measurements obtained.

Level and Trend

According to what we have seen from Figure 5, it might be interesting to quantify the changes in both level and slope, given that both types of effects
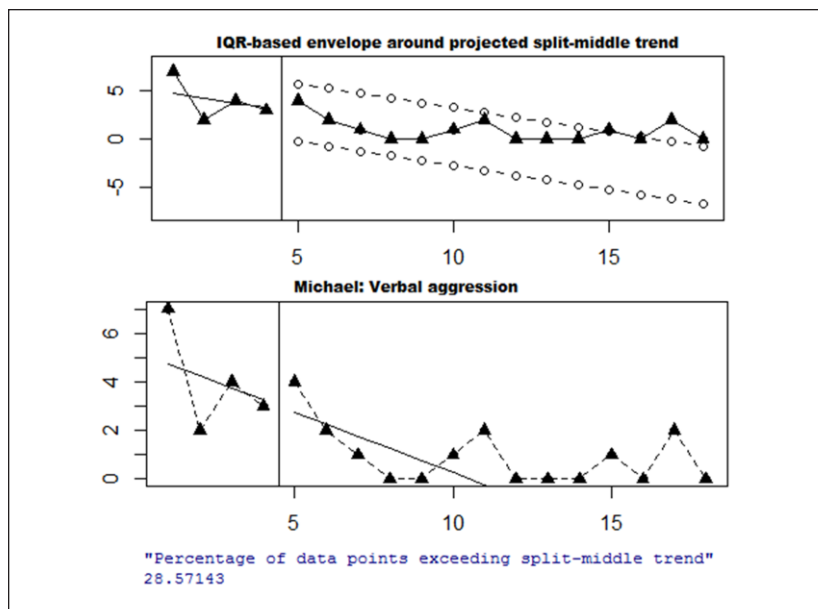
Figure 7. Graphical representation of the use of the split-middle trend and percentage of data points exceeding it, as well as a projection taking baseline data variability into account.

Note. The data are the verbal aggressions by Michael, as collected by Singh et al. (2007). The graphs are obtained using R code (https://www.dropbox.com/s/rlk3nwfoya7rm3h/ PEM-T.R?dl=0, described in the Nonoverlap indices section of the tutorial, and https:// dl.dropboxusercontent.com/s/5z9p5362bwlbj7d/ProjectTrend.R described in the Visual tools section of the tutorial). Note the different ways in which the *y*-axis is represented and its possible effect on visual inspection.

are present. However, these effects are not always in the same direction. Here, we will use tools from the tutorial sections "Unstandardized indices and their standardized versions" and "Application of two-level multilevel models for analyzing data" (see the appendix). Using the slope and level change procedure (Solanas, Manolov, & Onghena, 2010), we obtain the following quantifications of change in slope for physical aggression: −0.25 for Jason, 0.10 for Michael, and −0.60 for Tim, indicating that the improving change in slope is stronger for Tim than for Jason, and Michael's physically aggressive behavior is getting reduced at a slightly slower rate after the intervention than before. Looking the net level change (once slope change is controlled for), we obtain −0.86 for Jason, 0.37 for Michael, and −1.44 for Tim. We once again observe apparently a deterioration for Michael, but it is due to
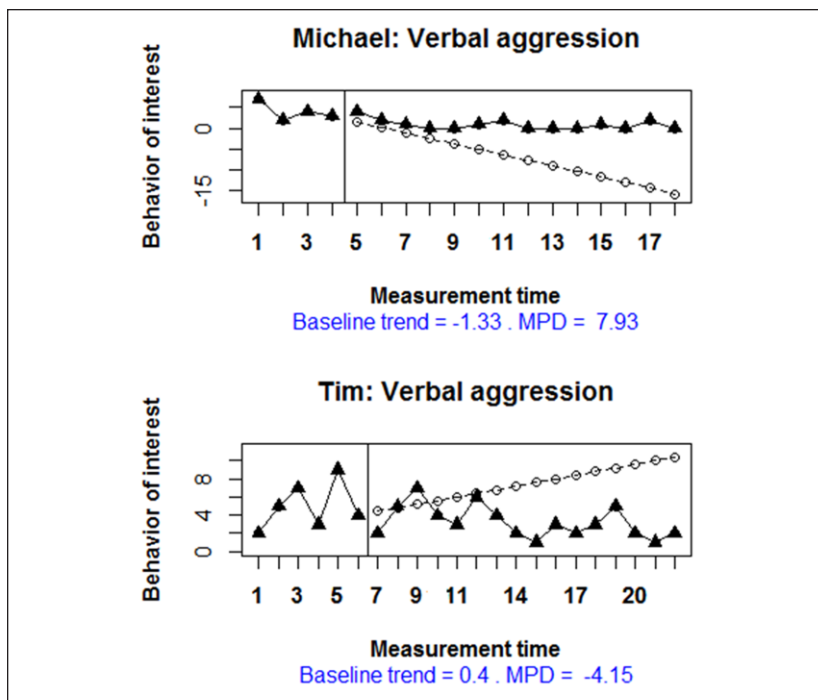
Figure 8. Graphical representation of the use of the mean phase difference procedure.
Note. The data are the verbal aggressions by Michael and Tim, as collected by Singh et al. (2007). The graphs are obtained using R code (https://www.dropbox.com/s/nky75oh40f1gbwh/MPD.R?dl=0, described in the Unstandardized indices and their standardized versions section of the tutorial).

the strong effect of the correction of baseline trend. Thus, in this case, the numerical result for the net level change does not seem to agree with the visual impression and we will stick to the latter, as it seems to represent better the data. The average of the slope change estimate for physical aggression, weighted by the number of measurements in each AB comparison, is −0.27 (see Figure 9). This value can be interpreted as an average progressive decrease of almost three physically aggressive behaviors per each 10 intervention phase measurements. The weighted average level change for physical aggression is −0.86, that is, less than one behavior average difference. When we look at the same quantifications for verbal aggressions, we see a weighted average slope change of −0.30 (similar to physical aggression) and
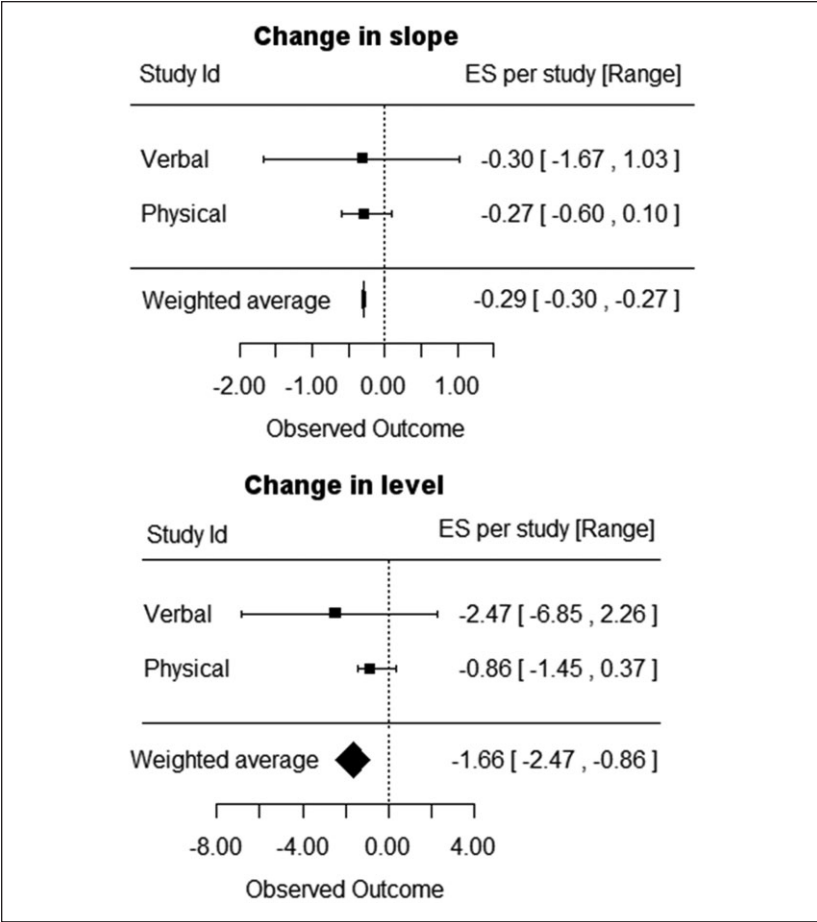
Figure 9. Graphical representation of the results obtained using the slope and level change procedure on the data collected by Singh et al. (2007).
Note. The average verbal and physical behaviors across the three participants are obtained first (using https://www.dropbox.com/s/74lr9j2keclrec0/Within-study_SLC_std.R, and described in the Unstandardized indices and their standardized versions section of the tutorial), before obtaining the global weighted average. This graph is obtained using R code (https://www.dropbox.com/s/wtboruzughbjg19/Across%20studies.R) and described in the Integrating results of several studies section of the tutorial.

a weighted average level change of −2.47 (indicating a much larger change than for physical aggression). The latter estimate for the average net level change diverges from our initial visual impression that the intervention effect

is clearer for physical aggression. This difference is due to the strong influence of the baseline trend estimated and controlled for in the case of Jason, leading to a level change of -6.85 for this participant (apart from the change in slope of -1.67). Thus, the numerical values in this case have to be interpreted with caution.

To continue exploring the changes in level and in slope, it is possible to use piecewise regression analysis (Center et al., 1985-1986). We will turn our attention to the verbally aggressive behaviors for Jason—we previously mentioned an immediate reduction of 7.15 behaviors. Moreover, the change in slope is negative, suggesting a reduction of 1.6 behaviors per measurement occasion, as compared with the baseline trend. In this case, the results are very similar to the ones provided by the slope and level change procedures, and in both cases it has to be considered whether a trend can be fitted reliably to only three baseline data points and whether it can be expected to continue in the same way (to very high values) throughout the intervention phase.

Next, we turn our attention to the verbal aggression measurements for Michael, as we said that the level change estimate of the slope and level change procedure seemed to disagree with the visual impression. From Figure 10, we observe that the change in slope (0.19) as estimated through piecewise regression is consistent with the slope and level change procedure and with the visual impression of small change.

Finally, another analytical technique that makes possible quantifying change in level and in slope[5] is a multilevel model (Moeyaert, Ferron, Beretvas, & Van Den Noortgate, 2014). Multilevel models (also referred to as hierarchical linear models and mixed linear models) allow modeling the dependencies in SCED studies (i.e., the measurements belonging to the same individuals are autocorrelated, see the reviews by Shadish & Sullivan, 2011; Solomon, 2014) and also when combining several SCED studies (i.e., the outcomes from the same study are not assumed to be independent). Moreover, these models permit modeling both fixed effects (e.g., the same baseline level for all individuals) and random effects (e.g., different effects of the intervention on time trend for the different individuals). In that sense, what are obtained are both average estimates across cases (e.g., an average level change for verbal aggression and for physical aggression) and estimates of the variation between cases in these changes. Moreover, individual-level (shrunken) estimates are provided using Empirical Bayes estimation. Other modeling capabilities include taking autocorrelation and heterogeneous variance into account. However, one of the potential limitations of the technique is that it requires many Level 2 units unless data series of least 20 measurements are available, to ensure the precision of the estimates of the average effects, especially of the variance components which should be interpreted
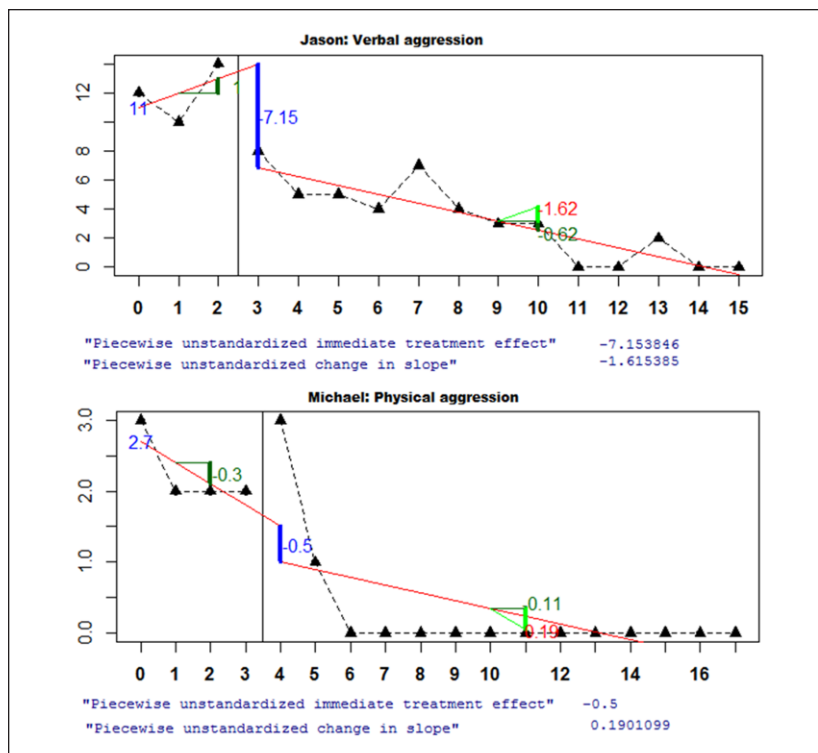
Figure 10. Graphical representation of the use of piecewise regression analysis on the data obtained by Singh et al. (2007) for the verbal aggressions by Jason and physical aggressions by Michael.

Note. The numerical results and the graphs are obtained using R code (https://www.dropbox.com/s/bt9lni2n2s0rv7l/Piecewise.R?dl=0), and described in the Unstandardized indices and their standardized versions section of the tutorial.

with caution in datasets such as the current one (Ferron, Bell, Hess, Rendina-Gobioff, & Hibbard, 2009). For more details about how models can be made increasingly more complex to model different types of effects (e.g., change in level and change in slope) and to account for autocorrelation and within-case or between-cases variability, see Moeyaert, Ferron, et al. (2014), whereas for more information about how to specify design matrices for a variety of SCEDs, Moeyaert, Ugille, Ferron, Beretvas, and Van den Noortgate (2014) could be consulted.

Given that two outcomes per participant we measured, we apply 2 two-level models (measurements nested into individuals): one for verbal

aggression and one for physical aggression. In the current example, with only three participants, it was not possible to model all data aspects that we wanted to model as random effects (i.e., allowing for variation between participants): baseline trend, change in level, change in trend, and autocorrelation. Thus, we did not model baseline trend. The graphical representation of the results (Figure 11) shows that for physical aggression, all participants start from a very similar level (2.19; *SE* = 0.33) and they also show similar changes in level (−0.56; *SE* = 0.43) and in slope (−0.16; *SE* = 0.04). First, considering that baseline trend was not modeled, the average estimate of the initial baseline value (i.e., the intercept) may not be a good representation of the actual measurements. Second, these values for the two types of effects are somewhat smaller than the ones obtained by the slope and level change procedure, but both are consistent with the general visual impression and offer a quantification of the amount of change. The amount of difference should be assessed, taking into account the fact the number of physically aggressive behaviors observed ranges from 0 to 4. Such information is potentially useful for stating whether the effect observed is rather small or large.

For verbal aggression, there is greater variability in the initial levels (an average of 6.97, *SE* = 2.50, with individual estimates equal to 11.81, 4.02, and 5.09), in the change in level (an average of −2.58, *SE* = 1.46, with individual estimates equal to −4.99, −2.15, and −0.62), and in slope (an average of −0.30, *SE* = 0.16, with individual estimates equal to −0.61, −0.13, and −0.18). The greater variation is consistent with the visual inspection and makes clear how much larger the effect is for Jason. When such amount of variability (which is also statistically significant) is observed, multilevel models allow for including moderator variables to account for it, but we will not complicate the analysis further here. Another interesting result is that the average effects (Table 1) are very similar to the ones provided by the slope and level change procedure (−2.58 vs. −2.47 for change in level; −0.30 according to both for change in slope).

Overlap

We have already shown a progression from simpler analysis (i.e., visual) to more complex ones (i.e., multilevel models). However, there is one other type of information usually taken into account by visual analysts: data overlap. From Figure 1, we see that there is very little overlap between the measurements pre- and postintervention for Jason and Michael, and thus, according to this criterion, the intervention effect is clearer for them than for Tim. In terms of quantification, there have been many nonoverlap indices proposed (see Parker, Vannest, & Davis, 2011, and the "Nonoverlap indices" section of the

Figure 11. Graphical representation of the multilevel (two-level) models (modeling random immediate effect and change in slope) applied separately on data obtained by Singh et al. (2007) for physical and verbal aggressions.
Note. The graphs are obtained adapting slightly the R code (https://www.dropbox.com/s/ slakgbeok8x19of/Two-level.R?dl=0) and described in the application of two-level multilevel models for analyzing data section of the tutorial. The slim lines are the predicted values for each participant. The thick line for the model for verbal aggression represents the average predicted values for all three participants.

tutorial), but we will focus here on two relatively recent and promising proposals (NAP by Parker & Vannest, 2009; and Tau-U by Parker, Vannest, Davis, & Sauber, 2011) instead of the classical Percentage of nonoverlapping data

Table 1. Results of the 2 Two-Level Analysis Performed Separately for Verbal and Physical Aggressions.

	Estimate	CI		p value
		Lower limit	Upper limit	
Two-level model: Verbal aggression				
Fixed coefficient				
Average baseline level	6.97	1.94	12	.008
Average effect on level	−2.58	−5.51	0.34	.082
Average effect on trend	−0.30	−0.63	0.03	.07
Variance component (*SD*)				
Baseline level	4.27	1.55	11.81	NA
Effect on level	2.29	0.65	8.07	NA
Effect on trend	0.27	0.09	0.83	NA
Residual variation	1.59	1.3	1.94	NA
Autocorrelation	−0.06	−0.37	0.25	NA
Two-level model: Physical aggression				
Fixed coefficient				
Average baseline level	2.19	NP	NP	<.001
Average effect on level	−0.56	NP	NP	.194
Average effect on trend	−0.16	NP	NP	<.001
Variance component (*SD*)				
Baseline level	0.0037	NP	NP	NA
Effect on level	0.0007	NP	NP	NA
Effect on trend	0.0002	NP	NP	NA
Residual variation	0.8969	NP	NP	NA
Autocorrelation	0.44	NP	NP	NA

Source. The results are obtained using the code available at https://www.dropbox.com/s/lbw59lsx6crh4m5/Multilevel%20code_Singh.R?dl=0.
Note. SAS proc mixed is a commercial alternative for carrying our multilevel analysis (Moeyaert, Ferron, et al., 2014). NA = result not available from the output of the R package "nlme"; NP = result not provided for this model by the R package "nlme"; CI = 95% confidence interval.

(Scruggs, Mastropieri, & Casto, 1987). The main difference between the two indices is the possibility to control for baseline trend. Figure 12 focuses on the verbal aggressive behaviors by Jason and Michael. For Jason, there is no data overlap (NAP ≡ A vs. B = −1.00), and there is also a trend that is deteriorating in general. However, due to the fact that the second baseline point is actually an improvement compared with the first one, Tau-U (A vs. B − trendA) controls for this improvement and yields a value of −.95. For Michael, the general
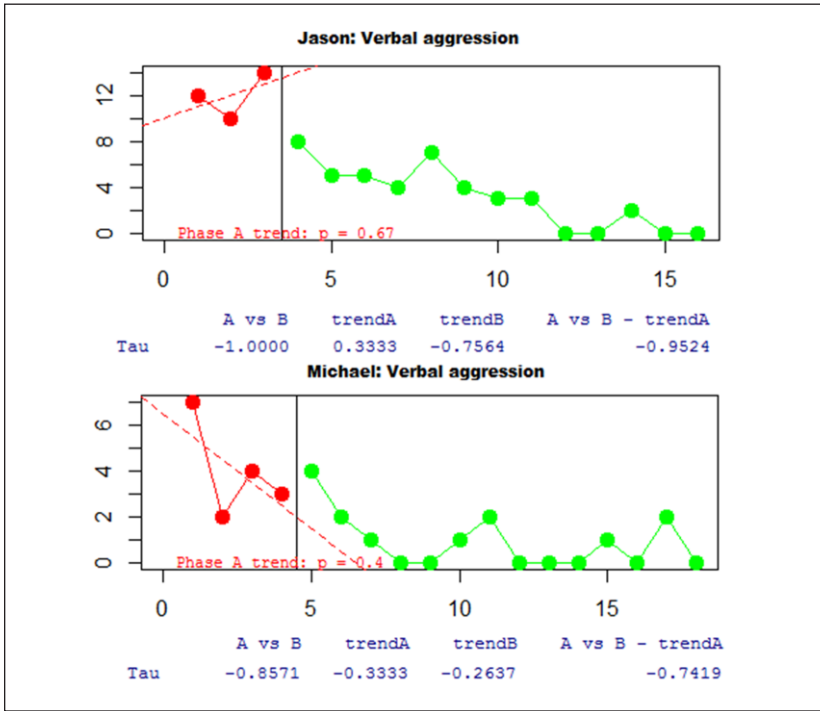
Figure 12. Graphical representation of the use of Tau-U and the nonoverlap of all pairs (A vs. B) on the data obtained by Singh et al. (2007) for the verbal aggressions by Jason and Michael.

Note. The numerical results and the graphs are obtained using R code (https://dl.dropboxusercontent.com/u/2842869/Tau_U.R) described in the Nonoverlap indices section of the tutorial.

baseline trend is improving, and thus the effect of the correction is greater (NAP = −.86 vs. Tau-U = −.74). To avoid overcorrections such as the one performed by Tau-U for Jason, we obtained the NAP values for all AB comparisons (−.95 and −1.00 for Jason; −.88 and −.86 for Michael; −.60 and −.43 for Tim), which leads to a median of −.87, an average of −.79, and an average weighted by series length equal to −.76, suggesting a moderate effect (in the range from |.66-.92| according to Parker & Vannest, 2009).

An index sometimes (e.g., Wehmeyer et al., 2006) used jointly with nonoverlap indices is the percentage of zero data (Wolery et al., 2010): a quantification that could be useful for the current data, given that the aim is to eliminate the aggressive behavior. According to this index, the majority of

measurement occasions after the first 0 remain at 0 (90.9% and 80% for Jason; 100% and 63.6% for Michael; 84.6% for Tim's physically aggressive behavior), except for Tim's verbal aggression (28.6%). Taking this result into account, together with the fact that there is only one 0 value in all six baseline phases, the effect of the intervention seems to be a practically relevant one.

Variance

Variance is important when analyzing SCED data as stable baselines are traditionally considered necessary for further comparisons (Kazdin, 1978). Moreover, data variability reduces the degree to which estimates of average level and trend are meaningful and representative of the data. Therefore, data variability is considered when constructing standard deviation bands around average levels (Figure 2) and when projecting trend (Figure 7): see the "Visual tools" section of the tutorial. Despite the importance of variance, it is not usually the object of the intervention or the focus of the data analysis (but see Winkens, Ponds, Pouwels-van den Nieuwenhof, Eilander, & van Heugten, 2014, for whom the increased variability of behavior after the intervention was an important reason for deeming it unsuccessful). Nevertheless, the information about data variability is incorporated into the standardized mean difference indices (Glass et al., 1981; Hedges et al., 2012, 2013), and the information about the variability around the baseline trend lines is suggested as part of the weighting strategy for the MPD (Manolov & Rochat, 2015). Finally, multilevel models allow modeling heterogeneous unexplained data variability in the baseline and the intervention phases (see Table 4 in Moeyaert, Ferron, et al., 2014).

Consistency of the Data Across Phases

The degree to which the data pattern is similar across replications, that is, for all three participants and for both types of behaviors can be assessed visually. The data for Michael and Jason (as per Figure 1) are similar, with a rapid and progressive reduction of aggressive behavior and elimination for physical aggression. Nevertheless, the data for Michael include an improving trend in the baseline, which makes causal attributions more difficult. For Tim, the data are more variable, but the intervention effect is also clearer for physical aggression. Tim is the only person for whom there is a deteriorating trend before the intervention.

The degree to which the effect is similar across cases has been incorporated as an output of the new version of the MPD (Manolov & Rochat, 2015), and it was also suggested to be used as part of the weight assigned to the
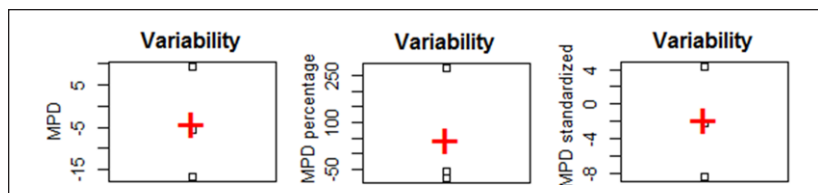
Figure 13. Graphical representation of the use of the mean phase difference procedure.
Note. The data are the verbal aggressions by all three participants, as collected by Singh et al. (2007). The graphs are obtained using R code (https://www.dropbox.com/s/ll25c9hbprro5gz/Within-study_MPD_percent.R and https://www.dropbox.com/s/g3btwdogh30biiv/Within-study_MPD_std.R), described in the Unstandardized indices and their standardized versions section of the tutorial.

average intervention effect for the whole study. Variability is represented via a strip chart, for the raw version of MPD expressed in number of behaviors, and for the percentage change and standardized versions (see Figure 13 for an example with the data for verbal aggression).

Moreover, multilevel models offer the possibility to model[6] the intervention effect as fixed (i.e., the same for all AB comparisons in the multiple-baseline design) or as random (i.e., varying across cases). The information for verbal aggression from Table 1 suggests that the amount of variation in the intervention effect on level and on slope is statistically significant at the .05 level, as the 95% confidence intervals do not include 0 as a plausible value for the variance. This information concurs with the visual impression from Figure 11, in which both the immediate effect and the slopes differ across individuals.

General Assessment of the Intervention Effect

Both visual analysis and the quantifications indicate that there is, in general, an effect of the intervention reducing the target behaviors (an average of two–three behaviors; almost 2 standard deviations) or even eliminating them. Nevertheless, there are differences across the participants (differences that are shown to be statistically significant). In view of the general results and the graphical information, we encourage researchers not to interpret any quantification in an isolated way, for instance, in relation to cases such as Michael's data, where estimating and projecting trends may obscure the fact that the problematic behavior is progressively reduced to elimination during the intervention phases, whereas the relatively short baseline does not allow predicting such an outcome with sufficient certainty. Thus, a sensitivity

analysis is called for to explore whether similar conclusions are obtained using different modeling options (see Moeyaert, Ferron, et al., 2014, for a detailed example).

A Remark on Formulaic Representations

Regression and multilevel analyses require specifying a model to be used as a representation of the data. For fitting and comparing ordinary least squares regression trend lines, such as the ones represented in Figure 5, the R code described in the tutorial can be used by only entering the data. The same is the case for piecewise regression model represented in Figure 9, for which the formulae are presented in Center et al. (1985-1986). Multilevel modeling is somewhat more complex, as the model is building according to the data features that the researcher considers relevant. Van den Noortgate and Onghena (2003, 2008) and Moeyaert, Ferron, et al. (2014) provide indications and formulaic representations for multilevel models. Moreover, general indications about data modeling (e.g., how to model change in level and in trend in different design structures) are available in Huitema and McKean (2000) and Moeyaert, Ugille, et al. (2014). We preferred to offer verbal instead of formulaic expressions in the current article.

A Remark on Meta-Analysis

To perform a meta-analysis of SCED studies, there are several options. First, one could use a three-level model taking into account the nested structure of the data: measurements within cases within studies (Moeyaert, Ugille, et al., 2014). Second, one could use classical meta-analytical techniques with the values of the d-statistic obtained from each study (Shadish, Hedges, et al., 2014), using inverse variance as a weight and obtaining confidence intervals, heterogeneity tests, and assessment of potential publication bias. Third, using other quantifications initially proposed for comparing a pair of conditions (e.g., nonoverlap indices, slope and level change, percentage change index) it is possible to obtain the average or the mean of the effects observed for each comparison. In the latter case, among the possibilities for using weights, the number of measurements seems to be a parsimonious solution (Kratochwill et al., 2010; Manolov, Guilera, & Sierra, 2014; Shadish, Rindskopf, & Hedges, 2008). Once an average effect is observed to represent the whole study[7] (or selected due to a substantive criterion), the effects for several studies can again be combined in the same way. Fourth, although it is not exactly meta-analysis, the results of different studies can be integrated combining probabilities (Darlington & Hayes, 2000; Rosenthal, 1978), such as the ones

obtained via randomization tests (Edgington & Onghena, 2007). All four options can be performed using free code and are illustrated in the tutorial—see the section "Integrating results of several studies".

Discussion

In this article, the aim was to bring analytic developments within the field of SCEDs closer to the applied SCED researcher. Another possible step that would potentially improve analytical practice would be statisticians and methodologists to collaborate with applied researchers sharing their knowledge. Such collaborations can also bring analytical developments to real-world practice, but we consider that the efforts (such as the current article) to make practitioners autonomous are justified, as the latter are the people who have most intimate knowledge of the client, the context (and the data in general) and should be the main actors in data analysis, instead of being detached from it.

Implications and Recommendations for Applied Researchers

The current article is intended to inform applied researchers about an easy-to-use set of analytical tools, so that they can obtain as much information as possible beyond what visual analysis can offer, always keeping in mind their objectives and the characteristics of data. We consider that informing applied researchers about tools for visual and quantitative analyses is justified, especially because the free availability of the techniques means that the software implementations come with no additional economic cost. Moreover, the current illustration alongside the ones included in the tutorial can guide researchers in their data-analytical process.

Papers such as the current one, together with journal special issues on the topic, are intended to increase the awareness of applied researchers about the existence of different analytical techniques, their usefulness, and the way in which they can actually be applied. In that sense, it would be important to know in which research areas it is most necessary to provide such information. To deal with this topic, we first review the research areas in which SCED studies appear most frequently, and then compare these areas with areas in which special issues about SCED methodology and analysis are available. This comparison can lead to a tentative idea about the domains in which statistical analysis is lacking visibility.

On one hand, Shadish and Sullivan (2011) identified 113 studies in the fields of psychology and education from 2008. The most frequently represented journals in this sample of studies deal with behavior modification/

intervention, developmental disorders (including autism spectrum disorders), and education. The review performed by Smith (2012) identified 409 articles in peer-reviewed journals in the period 2000-2010 reporting an SCED study, with the most frequent research areas being behavioral modification/interventions, developmental disorders, and school psychology. (In both reviews, special education articles are generally included in journals on developmental disorders or school psychology and education.) Moreover, area-specific reviews show that SCEDs are common in special education (Hammond & Gast, 2010), school-based intervention (Solomon, 2014), and neuropsychological rehabilitation (Perdices & Tate, 2009).

On the other hand, the areas in which special issues on SCED methodology and data have been offered include (special) education (*Journal of Behavioral Education* in 2012, Volume 21, Issue 3; *Remedial and Special Education* in 2013, Volume 34, Issue 4; *Journal of School Psychology* in 2014, Volume 52, Issue 2), rehabilitation (*Neuropsychological Rehabilitation* in 2014, Volume 24, Issues 3-4), developmental disorders (*Developmental Neurorehabilitation* with first papers appearing online in 2016), and interventions related to communication impairments (*Evidence-Based Communication Assessment and Intervention* in 2008, Volume 2, Issue 3; see also a forum on a target article in aphasiology in 2015, Volume 5, Issue 5). Special issues with more emphasis on methodology than on data analysis are available in sport psychology (*Journal of Applied Sport Psychology* in 2013, Volume 25, Issue 1), counseling (*Journal of Counseling and Development* in 2015, Volume 93, Issue 4; with an analytical focus on nonoverlap indices), and behavior modification (*Journal of Contextual Behavioral Science* in 2014, Volume 3; with an analytical focus on randomization tests).

If we compare the areas in which SCED studies are most frequently published with the areas in which special issues on SCED analysis are available, apparently the field of behavior modification/intervention has been object of less systematic efforts to cover a broad range of possibilities for statistical analysis and present them jointly in a synthesized for. In that sense, the current summary is timely for a journal focused on behavior modification. Nevertheless, identifying areas needing more emphasis on data analysis requires a proper review and, moreover, the references of the current article show that *Behavior Modification* has published articles related to SCED statistical analysis, with the current article presenting a broad overview.

In summary, although it is difficult to know whether special issues are a result of an increased awareness of the importance of quantitative analyses or they are created due to the detection of the continuous omission of any quantitative techniques, efforts such as the special issues, as well as articles

offering a broad overview such as the current one (see also Manolov & Moeyaert, 2016), along with conference presentations and workshops are deemed to be the way to promote the use of appropriate analytical techniques.

As our aim is not only to increase awareness and provide a broad overview in a research area which would apparently benefit from such a synthesis but also to show how analytical techniques can actually be applied, we consider that with information available here, in Chen et al. (2015) and in Manolov and Moeyaert (2016) applied researchers are capable to meet the criteria proposed by Tate et al. (2013) for scoring maximum the data analysis item in the Methodological Quality scale: structured visual analysis as detailed by Kratochwill and colleagues (2010) or Lane and Gast (2014), or visual analysis together with quasi-statistical quantifications (e.g., non-overlap or other percentage-based indices) or statistical techniques accompanied by the justification of their choice. Improved analysis makes more likely the publication of the results of an SCED study, although the methodological rigor and the interest of the results are also necessary for achieving this aim.

We consider that any potential improvement in the way in which data are analyzed or the way in which results are documented and communicated across researchers is useful establishing the evidence basis of interventions. This in turn can contribute to increasing the scientific credibility of SCEDs. Moreover, the tutorial presented here might prompt methodologists and statisticians to improve the usability of their software or to enhance the way in which the techniques they have authored are presented in that document. The result of this interest of basic researchers could be twofold. On the one hand, software implementations, illustrations, and tutorials can lead to methodological papers having impact on the real world (not only academic impact factor), when their proposals are being used and useful for making decisions about the (degree of) effectiveness of an intervention. On the other hand, large-scale collaborations may take place, leading to the development of a single software including all major analytical proposals for SCEDs and providing quantifications as well as a graphical representation of their meaning. Such a software would help applied researchers use their time more efficiently as, for instance, they would not need to organize their data in different ways according to the analysis to be used.

Limitations and Future Research

Several potential limitations need to be made explicit. First, in the current illustration of the possibilities for SCED data analysis, we included several

but not all possible analytical techniques. A detailed illustration and discussion of each of the techniques would have taken a lot of space—the tutorial that we are presenting here has more than 270 pages. In the appendix, we provide a list of analytical techniques for which free software is available, including references, URLs where to find them, and a brief specification of the type of comparison performed with the software. One of the procedures we did not include in the current illustration, but which is included in the tutorial (as there is free R software for it), is randomization test. Randomization tests were not included due to two reasons. First, we did not want to illustrate such an analysis in the absence of random assignment in the design, as it is necessary for ensuring the validity of the analysis (Edgington, 1980) and for the adequate performance of the test (Ferron, Foster-Johnson, & Kromrey, 2003). Thus, an illustration would have been potentially misleading. Second, the SCDA plug-in for R (Bulté & Onghena, 2012) requires that all AB comparisons have the same length, which is not the case here. Nevertheless, this omission should not be understood as an inadequacy of randomization tests for all types of SCED data (see Heyvaert & Onghena, 2014; Kratochwill & Levin, 2010, regarding the importance of randomization in the design; Levin, Ferron, & Kratochwill, 2012, for a discussion on randomization tests).

Second, part of the R code developed is not available via the CRAN website (http://cran.r-project.org), given that it is more difficult to develop and maintain the packages when R versions change. We have also not used any website (other than Dropbox URLs and ResearchGate and Academia personal web pages), as the current work is developed without external funding and we do not wish to bother the users with unnecessary advertisements (by companies which we do not endorse) from free websites.

Third, no new technique was proposed. However, we consider that given the myriad of possibilities, applied researchers first need to get acquainted with existing alternatives and how they can actually be used, before conducting further basic methodological research. Such methodological research should also be connected to (and made useful for) actual professional practice.

Fourth, the presentation of statistical techniques according to the criteria used in visual analysis was chosen due to the common use of the latter. Nevertheless, it should be noted that the quantifications may focus mainly on a single data feature but may be affected by other data features as well. For instance, a greater difference in level will refer to a greater nonoverlap, according to the amount of variability in the data. Also, the importance of a difference in means is subjected to how well these means represent the data, which is related to the amount of variability. Finally, some techniques may

yield different results according to decisions made by the researcher—in multilevel analysis, the point in which the comparison in level is made is relevant (e.g., at the beginning of the intervention phase or at the end); when using Tau-U, the decision to control or not for monotonic baseline trend and to quantify or not for intervention monotonic trend has effect on the quantification of overlap. Therefore, although there is no technique taking into account absolutely all data features on which visual analysts focus, several data features are relevant for the quantifications obtained. This further illustrates the need for joint use of visual and statistical analysis.

Finally, although dealing with the question "Which techniques should be used when?" was not our aim here, the indications provided are restricted to data that have characteristics similar to the ones collected by Singh et al. (2007). Accordingly, the instances in which we detected that a procedure was not as helpful as we would have desired (e.g., a multilevel model including all interesting data aspects) or that some procedures (e.g., slope and level change procedure, piecewise regression) may entail interpretative challenges when integrating the information with visual analysis are also restricted to the current dataset. Therefore, more detailed and comprehensive discussion is necessary (e.g., Manolov & Moeyaert, 2016) on that topic, and this is one of the possible lines of future research.

Another potentially interesting task for the future is to combine all the pieces of software into the same package, just as all explanations are available in the same tutorial. The tutorial itself is continuously updated and can certainly be improved, especially if several experts in different analytical techniques collaborate on this task. A study of the acceptability of several analytical techniques and also the acceptability and perceived usability of the software implementations and the tutorial is also potentially useful (e.g., consulting the tens of users that have already downloaded the tutorial from ResearchGate and Academia).

Appendix

The online tutorial we have created includes the following points about each of the analytical techniques: (a) name of the technique, (b) authors/proponents of the technique and suggested readings, (c) free[8] software that can be used and its author, (d) how the software can be obtained, (e) how the analysis can be run to obtain the results, and (f) how to interpret the results. The reader is guided along these steps using screenshots from the software. The implementations that the tutorial describes refer to 30 procedures, and we have created R code for 15 of them (not restricted to proposals made by us but extensible to

procedures for which software implementations were not available). In most cases, this R code provides as output both numerical results and graphical representations, as illustrated in the main text of the article. The software generally requires either that (a) the user locates an already prepared data file (e.g., a data matrix in Excel or .txt file) or (b) the measurements are introduced separated by commas. The results are obtained with a couple of clicks in the SCDA package (http://cran.r-project.org/web/packages/RcmdrPlugin.SCDA/index.html), or by copying and pasting the corresponding R code. The tutorial includes the details about downloading and using the packages and code. It is not necessary to learn programming languages but only to follow the examples with the datasets available online on how to use the techniques and afterward repeat the analyses with own data. Interpretation is also given, although applied researchers are still advised to read primary papers presenting the techniques for more detailed information.

The current version of the tutorial (https://www.dropbox.com/s/bocsaj8yvnrlyvq/Tutorial.pdf?dl=0; also available from the www.researchgate.net and www.academia.edu profiles of the first author) covers the procedures listed below. We also provide here the main references and a URL to the freely available software for each procedure. All the R code we created (but not the one created by other authors) is available in the same pdf file of the tutorial (see the Dropbox link presented before and from the www.researchgate.net and www.academia.edu profiles of the first author).

Tools for Visual Analysis

- Visual analysis with the SCDA package

Comparison: whole data pattern (e.g., multiple-baseline design, a reversal ABAB design in which initial baseline is followed by intervention introduction, withdrawal of the intervention, and re-introduction of the intervention, alternating treatments design, apart from AB)
References: Bulté and Onghena (2012), Gast and Spriggs (2010), Kratochwill et al. (2010)
URL: http://cran.r-project.org/web/packages/RcmdrPlugin.SCDA/index.html

- Using standard deviation bands as visual aids

Comparison: a pair of phases (e.g., AB)
References: Callahan and Barisa (2005), Pfadt and Wheeler (1995)
URL: https://dl.dropboxusercontent.com/s/elhy454ldf8pij6/SD_band.R

- Estimating and projecting baseline trend

Comparison: a pair of phases (e.g., AB)
Reference: Manolov, Sierra, Solanas, and Botella (2014)
URL: https://dl.dropboxusercontent.com/s/5z9p5362bwlbj7d/ProjectTrend.R

- Graph rotation for controlling for baseline trend

Comparison: a pair of phases (e.g., AB)
Reference: Parker, Vannest, and Davis (2014b)
URL: https://www.dropbox.com/s/jxfoik5twkwc1q4/GROT.R?dl=0

Nonoverlap Indices

- Percentage of nonoverlapping data

Comparison: whole data pattern (e.g., multiple-baseline design, ABAB, alternating treatments design, apart from AB)
References: Scruggs, Mastropieri, and Casto (1987); Scruggs and Mastropieri (2013)
URL: http://cran.r-project.org/web/packages/RcmdrPlugin.SCDA/index.html

- Percentage of data points exceeding the median

Comparison: whole data pattern (e.g., multiple-baseline design, ABAB, alternating treatments design, apart from AB)
Reference: Ma (2006); Parker and Hagan-Burke (2007)
URL: http://cran.r-project.org/web/packages/RcmdrPlugin.SCDA/index.html

- Pairwise data overlap

Comparison: a pair of phases (e.g., AB)
Reference: Wolery, Busick, Reichow, and Barton (2010)
URL: https://www.dropbox.com/s/jd8a6vl0nv4v7dt/PDO2.R?dl=0

- Nonoverlap of all pairs

Comparison: a pair of phases (e.g., AB)
References: Parker and Vannest (2009); Brossart, Vannest, Davis, and Patience (2014)
URL: http://www.singlecaseresearch.org and https://dl.dropboxusercontent.com/u/2842869/Tau_U.R.

- Improvement rate difference

Comparison: a pair of phases (e.g., AB)
Reference: Parker, Vannest, and Brown (2009)
URL: http://www.singlecaseresearch.org

- Tau-*U*

Comparison: a pair of phases (e.g., AB)
References: Parker, Vannest, Davis, and Sauber (2011); Brossart et al. (2014)
URL: http://www.singlecaseresearch.org and https://dl.dropboxusercontent. com/u/2842869/Tau_U.R and https://s3.amazonaws.com/ktarlow.com/stats/R/ Tau.R

- Percentage of data points exceeding median trend

Comparison: a pair of phases (e.g., AB)
Reference: Wolery et al. (2010)
URL: https://www.dropbox.com/s/rlk3nwfoya7rm3h/PEM-T.R?dl=0

- Percentage of nonoverlapping corrected data

Comparison: a pair of phases (e.g., AB)
Reference: Manolov and Solanas (2009)
URL: https://dl.dropboxusercontent.com/s/8revawnfrnrttkz/PNCD.R

Percentage Indices Not Quantifying Overlap

- Percentage of zero data

Comparison: a pair of phases (e.g., AB)
Reference: Wolery et al. (2010)
URL: https://www.dropbox.com/s/k57dj32gyit934g/PZD.R?dl=0

- Percentage reduction data (percentage change index) and mean base-line reduction

Comparison: a pair of phases (e.g., AB)
References: Hershberger, Wallace, Green, and Marquis (1999); Wendt (2009); Campbell (2004)
URL: https://www.dropbox.com/s/wt1qu6g7j2ln764/MBLR.R?dl=0.

Unstandardized Indices and Their Standardized Versions

- Ordinary least squares regression analysis

Comparison: a pair of phases (e.g., AB)
References: Huitema and McKean (2000); Gorsuch (1983)
URL: https://www.dropbox.com/s/v0see3bto1henod/OLS.R?dl=0.

- Piecewise regression analysis

Comparison: a pair of phases (e.g., AB)
References: Center, Skiba, and Casey (1985-1986); Van den Noortgate and Onghena (2008)
URL: https://www.dropbox.com/s/bt9lni2n2s0rv7l/Piecewise.R?dl=0.

- Generalized least squares regression analysis

Comparison: a pair of phases (e.g., AB)
References: Swaminathan, Rogers, Horner, Sugai, and Smolkowski (2014); Swaminathan, Rogers, and Horner (2014)
URL: https://www.dropbox.com/s/dni9qq5pqi3pc23/GLS.R?dl=0.

- Classical mean difference indices

Comparison: whole data pattern (e.g., multiple-baseline design, ABAB, alternating treatments design, apart from AB)
References: Beretvas and Chung (2008)
URL: http://cran.r-project.org/web/packages/RcmdrPlugin.SCDA/index.html

- SCED-specific mean difference indices

Comparison: multiple-baseline design with at least three cases; AB^k design across at least three cases
References: Hedges, Pustejovsky, and Shadish (2012, 2013); Shadish et al. (2014)
URL: http://blogs.edb.utexas.edu/pusto/software/ and https://github.com/jepusto/scdhlm

- Design-comparable effect size

Comparison: whole data pattern (e.g., multiple-baseline design, ABAB, alternating treatments design, apart from AB)

References: Pustejovsky, Hedges, and Shadish (2014)
URL: http://blogs.edb.utexas.edu/pusto/software/ and https://github.com/jepusto/scdhlm

- Mean phase difference (MPD)

Comparison: a pair of phases (e.g., AB)
Reference: Manolov and Solanas (2013b)
URL: https://www.dropbox.com/s/nky75oh40f1gbwh/MPD.R?dl=0

- MPD—Percentage and standardized versions

Comparison: whole data pattern (e.g., multiple-baseline design, ABAB, alternating treatments design, apart from AB)
Reference: Manolov and Rochat (2015)
URL: https://www.dropbox.com/s/ll25c9hbprro5gz/Within-study_MPD_percent.R and https://www.dropbox.com/s/g3btwdogh30biiv/Within-study_MPD_std.R

- Slope and level change

Comparison: a pair of phases (e.g., AB)
Reference: Solanas, Manolov, and Onghena (2010)
URL: https://www.dropbox.com/s/ltlyowy2ds5h3oi/SLC.R?dl=0

- Slope and level change—Percentage and standardized versions

Comparison: whole data pattern (e.g., multiple-baseline design, ABAB, alternating treatments design, apart from AB)
References: Manolov and Rochat (2015)
URL: https://www.dropbox.com/s/o0ukt01bf6h3trs/Within-study_SLC_percent.R and https://www.dropbox.com/s/74lr9j2keclrec0/Within-study_SLC_std.R.

Tools for Implementing Randomization and Using Randomization Tests

- Randomization tests with the SCDA package

Comparison: whole data pattern (e.g., multiple-baseline design, ABAB, alternating treatments design, apart from AB)

References: Edgington and Onghena (2007); Heyvaert and Onghena (2014)
URL: http://cran.r-project.org/web/packages/RcmdrPlugin.SCDA/index.html

- Randomization tests with *ExPRT*

Comparison: whole data pattern (e.g., multiple-baseline design, ABAB, alternating treatments design, apart from AB)
References: Levin, Ferron, and Kratochwill (2012); Levin, Lall, and Kratochwill (2011)
URL: http://code.google.com/p/exprt/

Carrying Out Simulation Modeling Analysis

Comparison: a pair of phases (e.g., AB)
References: Borckardt et al. (2008), Borckardt and Nash (2014)
URL: http://clinicalresearcher.org/software.htm

Implementing the Maximal Reference Approach

Comparison: a pair of phases (e.g., AB)
References: Manolov and Solanas (2012, 2013a), Manolov, Jamieson, Evans, and Sierra (2015)
URL: https://www.dropbox.com/s/56tqnhj4mng2wrq/Probabilities.R?dl=0

Application of Two-Level Multilevel Models for Analyzing Data

Comparison: whole data pattern (e.g., multiple-baseline design, ABAB, alternating treatments design, apart from AB)
References: Moeyaert, Ferron, et al. (2014); Van den Noortgate and Onghena (2003).
URL: https://www.dropbox.com/s/slakgbeok8x19of/Two-level.R?dl=0; code used in the current illustration https://www.dropbox.com/s/lbw59lsx6crh4m5/Multilevel%20code_Singh.R?dl=0. Moreover, the following papers offer code in SAS for two-level analysis: Baek and Ferron (2013) and Ferron, Moeyaert, Van den Noortgate, and Beretvas (2014).

Integrating Results of Several Studies

- Meta-analysis using the SCED-specific standardized mean difference

Comparison: a set of studies

References: Shadish, Hedges, et al. (2014); see also their R code in the article referenced

URL: https://www.dropbox.com/s/41gc9mrrt3jw93u/Across%20studies_d.R?dl=0

- Meta-analysis using the MPD and SLC (also applicable to other indices)

Comparison: a set of studies
References: Manolov and Rochat (2015)
URL: https://www.dropbox.com/s/wtboruzughbjg19/Across%20studies.R

- Application of three-level multilevel models for meta-analyzing data

Comparison: a set of studies
References: Moeyaert, Ferron, et al. (2014); Van den Noortgate and Onghena (2003).
URL: https://www.dropbox.com/s/qamahg4o1cu3sf8/Three-level.R?dl=0; the code used in the current illustration is available at the following address: https://www.dropbox.com/s/lbw59lsx6crh4m5/Multilevel%20code_Singh.R?dl=0. Moreover, the following paper offers code in SAS for three-level analysis: Moeyaert, Ugille, Ferron, Beretvas, and Van den Noortgate (2013)

- Integrating results combining probabilities

Comparison: a set of studies
References: Edgington (1972); Jones and Fiske (1953)
URL: http://cran.r-project.org/web/packages/RcmdrPlugin.SCDA/index.html

Note on the Data Used in the Current Article

Regarding the example used in the current article, the URL for the Excel file including data, organized in different ways according to the analytical technique, and used in this article is https://www.dropbox.com/s/atkly341lo1b503/Singh%20data.xlsx?dl=0. This file is also available as supplementary online material to the article.

Note on the Availability and Updating of R Tools

Regarding the R software resources described in the tutorial, some of them have been formalized as R packages (i.e., the SCDA plug-in for the

R-Commander package, the *nlme* and *scdhlm* packages), whereas others consist of code that has to be copied, in some cases slightly modified to specify the data to be analyzed and pasted in the R console (i.e., all remaining tools not mentioned in the previous parentheses). Moreover, only the SCDA plug-in and the *nlme* package are available in the CRAN repository (https://cran.r-project. org/web/packages/), and thus only these can be installed directly using the R menus. Given that R is continuously evolving, to remain in the CRAN repository, it is necessary for their maintainer to ensure that these packages still function properly for more recent versions of R. Regarding the R software tools that are not included in the repository, if they depend on other packages (i.e., the R code for Tau-U depends on the Kendall package, the R code for generalized least squares regression depends on the *lmtest* package, the R code for graph rotation depends on the *rgl* package, and the R code for meta-analysis depends on the *metafor* package), the creator of those tools needs to check whether these other packages are still available for more recent versions of R, although their importance for data analysis in general suggests that this is almost certainly so. In case, the R code does not depend on any other packages (i.e., all remaining tools not mentioned in the previous parentheses), it is expected to function properly regardless of the R version. The tutorial presented here described how the packages and codes existing at the time of its creation work (at the end of year 2015). The discussion provided in this article, regarding the use of the single-case analytical techniques, is still applicable (albeit representing our own view on the topic) regardless of the version or the existence of the software tools.

Declaration of Conflicting Interests

Funding

Notes

1. All the necessary links are provided in the appendix.
2. The ordinate axes of all figures represent counts of the corresponding type of aggressive behavior.
3. We do not suggest using the standard deviations band as a formal statistical tool, given that the data are not likely to be normally distributed, as assumed in the procedure.

4. We only perform this analysis to show how *d*-statistic values can be meta-analyzed using classical meta-analytical techniques. For the type of data collected by Singh et al. (2007) —two outcomes per participant—it is possible to carry out multivariate analysis or to use a multilevel model (Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013).

5. Actually, multilevel models can be used to model only change in level or only change in slope; they are flexible enough to be adapted to the data aspects that the researcher considers relevant to be modeled.

6. Although the design used by Singh et al. (2007) is a multiple baseline, it is relevant to mention that with piecewise regression it is possible to compare whether effects are similar in the different comparisons (A_1B_1 and A_2B_2) involved in a reversal design using design matrices 5 and 8 from Moeyaert, Ugille, Ferron, Beretvas, and Van den Noortgate (2014).

7. It is also possible to select at random one of the outcomes reported in a study or to perform this selection on a substantive basis (i.e., because it is the outcome of interest for the meta-analysis).

8. Commercial software is available for multilevel analysis, including options in SAS and SPSS as well as the specific programs HLM and MLwiN.

References

Baek, E., & Ferron, J. M. (2013). Multilevel models for multiple-baseline data: Modeling across participant variation in autocorrelation and residual variance. *Behavior Research Methods*, *45*, 65-74.

Barlow, D. H., Nock, M. K., & Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior change* (3rd ed.). Boston, MA: Pearson.

Beretvas, S. N., & Chung, H. (2008). A review of meta-analyses of single-subject experimental designs: Methodological issues and practice. *Evidence-Based Communication Assessment and Intervention*, *2*, 129-141.

Beretvas, S. N., Van den Noortgate, W., & Ferron, J. M. (2014, April). *Using multilevel modeling to meta-analyze single-case experimental design studies' results*. Workshop presented at the AERA 2014 Annual Meeting, Philadelphia, PA.

Blampied, N. M. (2000). Single-case research designs: A neglected alternative. *American Psychologist*, *55*, 960.

Borckardt, J. J., & Nash, M. R. (2014). Simulation modelling analysis for small sets of single-subject data collected over time. *Neuropsychological Rehabilitation*, *24*, 492-506.

Borckardt, J. J., Nash, M. R., Murphy, M. D., Moore, M., Shaw, D., & O'Neil, P. (2008). Clinical practice as natural laboratory for psychotherapy research: A guide to case-based time-series analysis. *American Psychologist*, *63*, 77-95.

Brossart, D. F., Vannest, K., Davis, J., & Patience, M. (2014). Incorporating nonoverlap indices with visual analysis for quantifying intervention effectiveness in single-case experimental designs. *Neuropsychological Rehabilitation*, *24*, 464-491.

Bulté, I., & Onghena, P. (2009). Randomization tests for multiple-baseline designs: An extension of the SCRT-R package. *Behavior Research Methods*, *41*, 477-485.

Bulté, I., & Onghena, P. (2012). When the truth hits you between the eyes: A software tool for the visual analysis of single-case experimental data. *Methodology*, *8*, 104-114.

Bulté, I., & Onghena, P. (2013). The single-case data analysis package: Analysing single-case experiments with R software. *Journal of Modern Applied Statistical Methods*, *12*, 450-478.

Busse, R. T., Kratochwill, T. R., & Elliott, S. N. (1995). Meta-analysis for single-case consultation outcomes: Applications to research and practice. *Journal of School Psychology*, *33*, 269-285.

Busse, R. T., McGill, R. J., & Kennedy, K. S. (2015). Methods for assessing single-case school-based intervention outcomes. *Contemporary School Psychology*, *19*, 136-144.

Byiers, B. J., Reichle, J., & Symons, F. J. (2012). Single-subject experimental design for evidence-based practice. *American Journal of Speech-Language Pathology*, *21*, 397-414.

Callahan, C. D., & Barisa, M. T. (2005). Statistical process control and rehabilitation outcome: The single-subject design reconsidered. *Rehabilitation Psychology*, *50*, 24-33.

Campbell, J. M. (2004). Statistical comparison of four effect sizes for single-subject designs. *Behavior Modification*, *28*, 234-246.

Center, B. A., Skiba, R. J., & Casey, A. (1985-1986). A methodology for the quantitative synthesis of intra-subject design research. *The Journal of Special Education*, *19*, 387-400.

Chen, L.-T., Peng, C.-Y. J., & Chen, M.-E. (2015). Computing tools for implementing standards for single-case designs. *Behavior Modification*, *39*, 835-869.

Cohen, L. L., Feinstein, A., Masuda, A., & Vowles, K. E. (2014). Single-case research design in pediatric psychology: Considerations regarding data analysis. *Journal of Pediatric Psychology*, *39*, 124-137.

Dallery, J., Cassidy, R. N., & Raiff, B. R. (2013). Single-case experimental designs to evaluate novel technology-based health interventions. *Journal of Medical Internet Research*, *15*, 268-284.

Darlington, R. B., & Hayes, A. F. (2000). Combining independent p values: Extensions of the Stouffer and binomial methods. *Psychological Methods*, *5*, 496-515.

Davis, D. H., Gagné, P., Fredrick, L. D., Alberto, P. A., Waugh, R. E., & Haardörfer, R. (2013). Augmenting visual analysis in single-case research with hierarchical linear modeling. *Behavior Modification*, *37*, 62-89.

de Vries, R. M., Hartogs, B. M., & Morey, R. D. (2015). A tutorial on computing Bayes factors for single-subject designs. *Behavior Therapy*, *46*, 809-823.

Dixon, M. R., Jackson, J. W., Small, S. L., Horner-King, M. J., Lik, N. M. K., Garcia, Y., & Rosales, R. (2009). Creating single-subject design graphs in Microsoft Excel™ 2007. *Journal of Applied Behavior Analysis*, *42*, 277-293.

Edgington, E. S. (1972). An additive method for combining probability values from independent experiments. *Journal of Psychology*, *80*, 351-363.

Edgington, E. S. (1980). Validity of randomization tests for one-subject experiments. *Journal of Educational Statistics*, *5*, 235-251.

Edgington, E. S., & Onghena, P. (2007). *Randomization tests* (4th ed.). London, England: Chapman & Hall/CRC.

Ferron, J. M., Bell, B. A., Hess, M. R., Rendina-Gobioff, G., & Hibbard, S. T. (2009). Making treatment effect inferences from multiple-baseline data: The utility of multilevel modeling approaches. *Behavior Research Methods*, *41*, 372-384.

Ferron, J. M., Foster-Johnson, L., & Kromrey, J. D. (2003). The functioning of single-case randomization tests with and without random assignment. *The Journal of Experimental Education*, *71*, 267-288.

Ferron, J. M., Moeyaert, M., Van den Noortgate, W., & Beretvas, S. N. (2014). Estimating causal effects from multiple-baseline studies: Implications for design and analysis. *Psychological Methods*, *19*, 493-510.

Fisch, G. S. (2001). Evaluating data from behavioral analysis: Visual inspection or statistical models? *Behavioural Processes*, *54*, 137-154.

Franklin, R. D., Gorman, B. S., Beasley, T. M., & Allison, D. B. (1996). Graphical display and visual analysis. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 119-158). Mahwah, NJ: Lawrence Erlbaum.

Gast, D. L., & Ledford, J. R. (2014). *Single subject research methodology: Applications in special education and behavioral sciences* (2nd ed.). London, England: Routledge.

Gast, D. L., & Spriggs, A. D. (2010). Visual analysis of graphic data. In D. L. Gast (Ed.), *Single subject research methodology in behavioral sciences* (pp. 199-233). London, England: Routledge.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Gorczynski, P. (2013). The use of single-case experimental research to examine physical activity, exercise, and physical fitness interventions: A review. *Journal of Applied Sport Psychology*, *25*, 148-156.

Gorsuch, R. L. (1983). Three methods for analyzing limited time-series (N of 1) data. *Behavioral Assessment*, *5*, 141-154.

Graham, J. E., Karmarkar, A. M., & Ottenbacher, K. J. (2012). Small sample research designs for evidence-based rehabilitation: Issues and methods. *Archives of Physical Medicine and Rehabilitation*, *93*, S111-S116.

Hammond, D., & Gast, D. L. (2010). Descriptive analysis of single subject research designs: 1983-2007. *Education and Training in Autism and Developmental Disabilities*, *45*, 187-202.

Harrington, M., & Velicer, W. F. (2015). Comparing visual and statistical analysis in single-case studies using published studies. *Multivariate Behavioral Research*, *50*, 162-183.

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods*, *3*, 224-239.

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods*, *4*, 324-341.

Hershberger, S. L., Wallace, D. D., Green, S. B., & Marquis, J. G. (1999). Meta-analysis of single-case data. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 107-132). London, England: SAGE.

Heyvaert, M., & Onghena, P. (2014). Analysis of single-case data: Randomisation tests for measures of effect size. *Neuropsychological Rehabilitation*, *24*, 507-527.

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*, 165-179.

Houle, T. T. (2009). Statistical analyses for single-case experimental designs. In D. H. Barlow, M. K. Nock, & M. Hersen (Eds.), *Single case experimental designs: Strategies for studying behavior change* (3rd ed., pp. 271-305). Boston, MA: Pearson.

Howick, J., Chalmers, I., Glasziou, P., Greenhalgh, T., Heneghan, C., Liberati, A., . . . Thornton, H. (2011). *The 2011 Oxford CEBM evidence table* (introductory document). Oxford Centre for Evidence-Based Medicine. Retrieved from http://www.cebm.net/index.aspx?o=5653

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

Institute of Education Sciences. (2014). *Request for applications*. *Statistical and research methodology for education*. Retrieved August 5, 2016 from ies.ed.gov/funding/pdf/2015_84305D.pdf

Jenson, W. R., Clark, E., Kircher, J. C., & Kristjansson, S. D. (2007). Statistical reform: Evidence-based practice, meta-analyses, and single subject designs. *Psychology in the Schools*, *44*, 483-493.

Johnston, J. M., & Pennypacker, H. S., Jr. (2009). *Strategies and tactics of behavioral research* (3rd ed.). New York, NY: Routledge.

Jones, L. V., & Fiske, D. W. (1953). Models for testing the significance of combined results. *Psychological Bulletin*, *50*, 375-382.

Kazdin, A. E. (1978). Methodological and interpretive problems of single-case experimental designs. *Journal of Consulting and Clinical Psychology*, *46*, 629-642.

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). Oxford, UK: Oxford University Press.

Kennedy, C. H. (2005). *Single-case designs for educational research*. Boston, MA: Pearson.

Kratochwill, T. R., & Brody, G. H. (1978). Single subject designs: A perspective on the controversy over employing statistical inference and implications for research and training in behavior modification. *Behavior Modification*, *2*, 291-307.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_scd.pdf

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods*, *15*, 124-144.

Kratochwill, T. R., & Levin, J. R. (Eds.). (2014). *Single-case research designs and analysis: New directions for psychology and education* (2nd ed.). New York, NY: Routledge.

Kromrey, J. D., & Foster-Johnson, L. (1996). Determining the efficacy of intervention: The use of effect sizes for data analysis in single-subject research. *Journal of Experimental Education*, *65*, 73-93.

Lane, J. D., & Gast, D. L. (2014). Viual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation*, *24*, 445-463.

Ledford, J. R., & Gast, D. L. (2014). Measuring procedural fidelity in behavioural research. *Neuropsychological Rehabilitation*, *24*, 332-348.

Levin, J. R., Evmenova, A. S., & Gafurov, B. S. (2014). The single-case data-analysis ExPRT (Excel7 Package of Randomization Tests). In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case intervention research: Methodological and statistical advances* (pp. 185-219). Washington, DC: American Psychological Association.

Levin, J. R., Ferron, J. M., & Gafurov, B. S. (2014). Improved randomization tests for a class of single-case intervention designs. *Journal of Modern Applied Statistical Methods*, *13*, 2-52.

Levin, J. R., Ferron, J. M., & Kratochwill, T. R. (2012). Nonparametric statistical tests for single-case systematic and randomized ABAB . . . AB and alternating treatment intervention designs: New developments, new directions. *Journal of School Psychology*, *50*, 599-624.

Levin, J. R., Lall, V. F., & Kratochwill, T. R. (2011). Extensions of a versatile randomization test for assessing single-case intervention effects. *Journal of School Psychology*, *49*, 55-79.

Ma, H. H. (2006). An alternative method for quantitative synthesis of single-subject research: Percentage of data points exceeding the median. *Behavior Modification*, *30*, 598-617.

Macgowan, M. J., & Wong, S. E. (2014). Single-case designs in group work: Past applications, future directions. *Group Dynamics: Theory, Research, and Practice*, *18*, 138-158.

Manolov, R., Guilera, G., & Sierra, V. (2014). Weighting strategies in the meta-analysis of single-case studies. *Behavior Research Methods*, *46*, 1152-1166.

Manolov, R., Jamieson, M., Evans, J. J., & Sierra, V. (2015). Probability and visual aids for assessing intervention effectiveness in single-case designs: A field test. *Behavior Modification*, *39*, 691-720.

Manolov, R., Krasny-Pacini, A., Evans, J. J., & Chevignard, M. (2014, May). *Evidence-based rehabilitation using single-case experimental designs: From protocol design to statistical interpretation of intervention effect*. Workshop presented at 19th European Congress of Physical and Rehabilitation Medicine, Marseille, France.

Manolov, R., & Moeyaert, M. (2016, May 16). Recommendations for choosing single-case data analytical techniques. *Behavior Therapy*. Advance online publication. doi:10.1016/j.beth.2016.04.008

Manolov, R., & Rochat, L. (2015). Further developments in summarizing and meta-analyzing single-case data: An illustration with neurobehavioural interventions in acquired brain injury. *Neuropsychological Rehabilitation*, *25*, 637-662.

Manolov, R., Sierra, V., Solanas, A., & Botella, J. (2014). Assessing functional relations in single-case designs: Quantitative proposals in the context of the evidence-based movement. *Behavior Modification*, *38*, 878-913.

Manolov, R., & Solanas, A. (2009). Percentage of nonoverlapping corrected data. *Behavior Research Methods*, *41*, 1262-1271.

Manolov, R., & Solanas, A. (2012). Assigning and combining probabilities in single-case studies. *Psychological Methods*, *17*, 495-509.

Manolov, R., & Solanas, A. (2013a). Assigning and combining probabilities in single-case studies: A second study. *Behavior Research Methods*, *45*, 1024-1035.

Manolov, R., & Solanas, A. (2013b). A comparison of mean phase difference and generalized least squares for analyzing single-case data. *Journal of School Psychology*, *51*, 201-215.

Maric, M., de Haan, E., Hogendoorn, S. M., Wolters, L. H., & Huizenga, H. M. (2015). Evaluating statistical and clinical significance of intervention effects in single-case experimental designs: An SPSS method to analyze univariate data. *Behavior Therapy*, *46*, 230-241.

Moeyaert, M., Ferron, J., Beretvas, S., & Van Den Noortgate, W. (2014). From a single-level analysis to a multilevel analysis of since-case experimental designs. *Journal of School Psychology*, *52*, 191-211.

Moeyaert, M., Ugille, M., Ferron, J., Beretvas, S. N., & Van den Noortgate, W. (2013). The three-level synthesis of standardized single-subject experimental data: A Monte Carlo simulation study. *Multivariate Behavioral Research*, *48*, 719-748.

Moeyaert, M., Ugille, M., Ferron, J., Beretvas, S. N., & Van den Noortgate, W. (2014). The influence of the design matrix on treatment effect estimates in the quantitative analyses of single-case experimental designs research. *Behavior Modification*, *38*, 665-704.

Nagler, E., Rindskopf, D., & Shadish, W. (2008). *Analyzing data from small N designs using multilevel models: A procedural handbook*. Unpublished manuscript. Retrieved on August 5, 2016 from http://www.ipr.northwestern.edu/workshops/past-workshops/quasi-experimental-design-and-analysis-in-education/recent-articles/docs/QE2010-SCD%20Manual.pdf

Onghena, P., & Van Damme, G. (1994). SCRT 1.1: Single-case randomization tests. *Behavior Research Methods, Instruments, & Computers*, *26*, 369.

Parker, R. I., & Brossart, D. F. (2003). Evaluating single-case research data: A comparison of seven statistical methods. *Behavior Therapy*, *34*, 189-211.

Parker, R. I., Brossart, D. F., Vannest, K. J., Long, J. R., Garcia De-Alba, R., Baugh, F. G., & Sullivan, J. R. (2005). Effect sizes in single case research: How large is large? *School Psychology Review*, *34*, 116-132.

Parker, R. I., Cryer, J., & Byrns, G. (2006). Controlling baseline trend in single-case research. *School Psychology Quarterly*, *21*, 418-443.

Parker, R. I., & Hagan-Burke, S. (2007). Median-based overlap analysis for single case data: A second study. *Behavior Modification*, *31*, 919-936.

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

Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single-case research. *Exceptional Children*, *75*, 135-150.

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, *35*, 303-322

Parker, R. I., Vannest, K. J., & Davis, J. L. (2014a). Non-overlap analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case intervention research: Methodological and statistical advances* (pp. 127-151). Washington, DC: American Psychological Association.

Parker, R. I., Vannest, K. J., & Davis, J. L. (2014b). A simple method to control positive baseline trend within data nonoverlap. *Journal of Special Education*, *48*, 79-91.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy*, *42*, 284-299.

Perdices, M., & Tate, R. L. (2009). Single-subject designs as a tool for evidence-based clinical practice: Are they unrecognised and undervalued? *Neuropsychological Rehabilitation*, *19*, 904-927.

Pfadt, A., & Wheeler, D. J. (1995). Using statistical process control to make data-based clinical decisions. *Journal of Applied Behavior Analysis*, *28*, 349-370.

Plavnick, J. B., & Ferreri, S. J. (2013). Single-case experimental designs in educational research: A methodology for causal analyses in teaching and learning. *Educational Psychology Review*, *25*, 549-569.

Pustejovsky, J. E., Hedges, L. V., & Shadish, W. R. (2014). Design-comparable effect sizes in multiple baseline designs: A general modeling framework. *Journal of Educational and Behavioral Statistics*, *39*, 368-393.

Rosenthal, R. (1978). Combining results of independent studies. *Psychological Bulletin*, *85*, 185-193.

Scruggs, T. E., & Mastropieri, M. A. (2013). PND at 25: Past, present, and future trends in summarizing single-subject research. *Remedial and Special Education*, *34*, 9-19.

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education*, *8*, 24-33.

Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2013, March). *An SPSS macro for a d-statistic for single-case designs*. Workshop presented at the 2013 Spring conference of the Society for Research on Educational Effectiveness, Washington, DC.

Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. *Journal of School Psychology*, *52*, 123-147.

Shadish, W. R., Rindskopf, D. M., & Hedges, L. V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention*, *2*, 188-196.

Shadish, W. R., & Sullivan, K. J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods*, *43*, 971-980.

Shadish, W. R., Zuur, A. F., & Sullivan, K. J. (2014). Using generalized additive (mixed) models to analyze single case designs. *Journal of School Psychology*, *52*, 149-178.

Singh, N. N., Lancioni, G. E., Winton, A. S., Adkins, A. D., Wahler, R. G., Sabaawi, M., & Singh, J. (2007). Individuals with mental illness can control their aggressive behavior through mindfulness training. *Behavior Modification*, *31*, 313-328.

Smith, J. D. (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods*, *17*, 510-550.

Solanas, A., Manolov, R., & Onghena, P. (2010). Estimating slope and level change in N = 1 designs. *Behavior Modification*, *34*, 195-218.

Solomon, B. G. (2014). Violations of assumptions in school-based single-case data: Implications for the selection and interpretation of effect sizes. *Behavior Modification*, *38*, 477-496.

Swaminathan, H., Rogers, H. J., & Horner, R. H. (2014). An effect size measure and Bayesian analysis of single-case designs. *Journal of School Psychology*, *52*, 213-230.

Swaminathan, H., Rogers, H. J., Horner, R., Sugai, G., & Smolkowski, K. (2014). Regression models for the analysis of single case designs. *Neuropsychological Rehabilitation*, *24*, 554-571.

Tate, R. L., Perdices, M., Rosenkoetter, U., Shadish, W., Vohra, S., Barlow, D. H., . . . Wilson, B. (2016). The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016 statement. *Journal of Clinical Epidemiology*, *73*, 142-152.

Tate, R. L., Perdices, M., Rosenkoetter, U., Wakima, D., Godbee, K., Togher, L., & McDonald, S. (2013). Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: The 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. *Neuropsychological Rehabilitation*, *23*, 619-638.

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*, *51*, 473-481.

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, *45*, 576-594.

Van den Noortgate, W., & Onghena, P. (2003). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments, & Computers*, *35*, 1-10.

Van den Noortgate, W., & Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental design studies. *Evidence-Based Communication Assessment and Intervention*, *2*, 142-151.

Wehmeyer, M. L., Palmer, S. B., Smith, S. J., Parent, W., Davies, D. K., & Stock, S. (2006). Technology use by people with intellectual and developmental disabilities to support employment activities: A single-subject design meta-analysis. *Journal of Vocational Rehabilitation*, *24*, 81-86.

Wendt, O. (2009, May 18). *Calculating effect sizes for single-subject experimental designs: An overview and comparison*. Paper presented at The Ninth Annual Campbell Collaboration Colloquium, Oslo, Norway. Retrieved from http://www.campbellcollaboration.org/artman2/uploads/1/Wendt_calculating_effect_sizes.pdf

Winkens, I., Ponds, R., Pouwels-van den Nieuwenhof, C., Eilander, H., & van Heugten, C. (2014). Using single-case experimental design methodology to evaluate the effects of the ABC method for nursing staff on verbal aggressive behaviour after acquired brain injury. *Neuropsychological Rehabilitation*, *24*, 349-364.

Wolery, M., Busick, M., Reichow, B., & Barton, E. E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *Journal of Special Education*, *44*, 18-29.

Author Biographies

Rumen Manolov, PhD, is lecturer at the Faculty of Psychology at the University of Barcelona in Spain. His investigation is focused on single-case designs data analysis: proposing, testing, and discussing analytical techniques, as well as developing and promoting R code for their implementation.

Mariola Moeyaert, PhD, is Assistant professor Statistics and Research design at the Department of Educational Psychology and Methodology at the State University of New York. Her Major research interests and publications are in the field of multilevel analysis, meta-analysis and single-case experimental data.