

CONSISTENCY IN ALTERNATION DESIGNS

Assessing Consistency in Single-Case Alternation Designs

Rumen Manolov¹, René Tanious², Tamal Kumar De², & Patrick Onghena²

¹Department of Social Psychology and Quantitative Psychology,

Faculty of Psychology, University of Barcelona

²Faculty of Psychology and Educational Sciences, Methodology of Educational Sciences

Research Group, KU Leuven – University of Leuven, Leuven, Belgium

***Corresponding Author:**

Rumen Manolov*

Department de Psicologia Social i Psicologia Quantitativa, Universitat de Barcelona

Passeig de la Vall d'Hebron 171, 08035, Barcelona, Spain

E-mail: rrumenov13@ub.edu

René Tanious

Faculty of Psychology and Educational Sciences, KU Leuven

Tiensestraat 102, B-3000 Leuven, Belgium.

Phone: +32 16 32 82 19

E-mail: rene.tanious@kuleuven.be

Tamal Kumar De

Faculty of Psychology and Educational Sciences, KU Leuven

Tiensestraat 102, B-3000 Leuven, Belgium.

Phone: +32 16 37 92 03

E-mail: tamalkumar.de@kuleuven.be

Patrick Onghena

Faculty of Psychology and Educational Sciences, KU Leuven

Tiensestraat 102, B-3000 Leuven, Belgium

Phone: +32 16 32 59 54

E-mail: patrick.onghena@kuleuven.be

CONSISTENCY IN ALTERNATION DESIGNS

Abstract

Consistency is one of the crucial single-case data aspects that are expected to be assessed visually, when evaluating the presence of an intervention effect. Complementarily to visual inspection, there have been recent proposals for quantifying the consistency of data patterns in similar phases and the consistency of effects for reversal, multiple-baseline, and changing criterion designs. The current text continues this line of research by focusing on alternation designs using block randomization. Specifically, three types of consistency are discussed: consistency of superiority of one condition over another, consistency of the average level across blocks, and consistency in the magnitude of the effect across blocks. The focus is put especially on the latter type of consistency, which is quantified on the basis of partitioning the variance, as attributed to the intervention, to the blocking factor or remaining as residual (including the interaction between the intervention and the blocks). Several illustrations with real and fictitious data are provided in order to make clear the meaning of the quantification proposed. Moreover, specific graphical representations are recommend for complementing the numerical assessment of consistency. A freely available user-friendly webpage is developed for implementing the proposal.

Keywords: Single-case design, Alternating treatments designs, Randomized blocks, Consistency, Analysis of variance

CONSISTENCY IN ALTERNATION DESIGNS

Assessing Consistency in Single-Case Alternation Designs

Single-case experimental designs (SCEDs) are useful for providing the evidence basis for interventions, especially when several threats to validity are taken into account (Horner et al., 2005; Petursdottir & Carr, 2018; Reichow, Barton, & Maggin, 2018). Several methodology quality appraisal scales for SCEDs exist guiding researchers in the conduct of their studies (Ganz & Ayres, 2018; Maggin, Briesch, Chafouleas, Ferguson, & Clark, 2014; Zimmerman et al., 2018). Moreover, apart from ensuring that as many desirable elements (e.g., manipulation of the independent variable, reliable measurement of the dependent variable, procedural fidelity, and randomization) as possible are included in a specific study, it is also important to report the procedures in replicable detail (Ganz & Ayres, 2018), for example using the Single-Case Reporting In BEhavioral interventions (SCRIBE) guidelines (Tate et al., 2016). Detailed reporting is especially relevant, as it makes replication possible. Replication is especially relevant for studies following an idiographic approach (Kennedy, 2005; Sidman, 1960). Specifically, within-study replication is necessary for documenting the reliability of an experimental effect, whereas across studies replication is required for assessing the generality of the effect (Maggin et al., 2014).

Regarding the demonstration of effects, several authors coincide on the importance of visual analysis (Kratochwill et al., 2010; Lane, Shepley, & Spriggs, 2019; Ledford, Barton, Severini, & Zimmerman, 2019; Maggin, Cook, & Cook, 2018). This concurs with the SCED tradition (e.g., Fahmie & Hanley, 2008; Miller, 1985; Lane & Gast, 2014; Parker, Cryer, & Byrns, 2006; Parsonson & Baer, 1978) and with the fact that all methodological quality appraisal scales require visual analysis but not all require statistical analysis for evaluating effects (Heyvaert, Wendt, Van den Noortgate, & Onghena, 2015). In the presentation of how visual analysis is to be

CONSISTENCY IN ALTERNATION DESIGNS

performed (Kratochwill et al., 2010; Lane et al., 2019; Ledford et al., 2019; Maggin et al., 2018), the focus is put on six data features. Three of these data features (level, trend, variability) can be assessed within each phase and they can also be used when comparing adjacent phases, whereas two features (overlap and immediacy of effect) refer necessarily to a comparison across adjacent phases. The sixth aspect, consistency, has a twofold conceptualization. On the one hand, it is possible to assess the consistency of data patterns between phases implementing the same experimental condition(s). On the other hand, the consistency of effects can be evaluated: whether the amount and type of change in level, change in slope, change in variability, overlap and immediacy is similar across replications of the basic effect when comparing two conditions. Thus, in order to be able to state that there is evidence for an intervention effect or for a functional relation, the effects observed in the replications should be consistent (Maggin et al., 2018). Actually, some authors consider consistency to be the “most important” requirement (Ledford, 2018, p. 82), given that “consistency and replication are essential characteristics for a functional relation determination – large differences in level are not” (Lane, Ledford, & Gast, 2017, p. 7102300010p6).

Alternation Designs

Terms and characteristics. In the present text, we focus on the assessment of consistency of effects and we present new proposals for a specific type of SCEDs – alternation designs.

Alternation designs are SCEDS that are characterized by the rapid alternation of the treatment levels, in contrast to phase designs that are characterized by a larger number of consecutive measurement occasions under the same treatment level (Onghena & Edgington, 2005).

Systematic reviews by Smith (2012) and Shadish and Sullivan (2011) indicated that alternation designs are commonly used, accounting for six and eight percent respectively in their samples of

CONSISTENCY IN ALTERNATION DESIGNS

published research. According to the review by Hammond and Gast (2010), alternation designs are even more frequent in journals publishing SCED research on special education, representing approximately 16%.

There are different kinds of alternation designs and the focus of the current text is on randomized block designs (RBDs). In RBDs, usually two conditions (called “A” and “B” in the following) are being compared and the sequence of measurement occasions is divided into blocks of two measurement occasions. In each block, the A and the B conditions take place, in a random order. This randomly determined sequence is equivalent to the N-of-1 trials used in the health sciences (Nikles & Mitchell, 2015), where the several random-order AB blocks are called multiple crossovers. It is possible (and in the health sciences common) to replicate the series across several participants, each with its own randomly determined sequence. An RBD is different from other alternation designs, such as completely randomized designs, in which any sequence is possible (e.g., AAABBABBBA), without considering blocks and without ensuring rapid alternation. An RBD is also different from an alternating treatments design with restricted randomization (also called restricted alternating treatments design [ATD], Onghena & Edgington, 1994) in that certain sequences are not possible under an RBD, but are possible under the latter kind of design. For instance, restricting the maximum number of consecutive administrations of the same condition to two, a sequence such as AABBAABBAABB is possible for an alternating treatments design with restricted randomization, but it cannot be obtained following an RBD randomization scheme because the same treatment can only be administered once within each block. The current focus on RBDs is related to the quantifications proposed for assessing consistency: these quantifications are based on the existence of blocks and the random assignment taking place within the blocks.

CONSISTENCY IN ALTERNATION DESIGNS

The important distinction between restricted ATDs and RBDs is proposed in several methodological articles (e.g., Edgington, 1996; Manolov, 2019; Onghena & Edgington, 2005), but in applied SCED literature RBDs probably are not always denoted as such. For instance, an RBD can be referred to as an ATD with “blocked pairs random assignment procedure” (Lloyd, Finley, & Weaver, 2018, p. 215) or an ATD in which the order of conditions was “block randomized” (Warren, Cagliani, Whiteside, & Ayres, 2019, p. 9). Moreover, Wolery et al. (2018) mention two options when referring to how the alternation sequence is determined in an ATD.— The first option is “random alternation with no condition repeating until all have been conducted” and the second option is “random alternation with no more than two consecutive sessions in a single condition” (p. 304). The first option refers to an RBD and the second to a restricted ATD. Similarly, when referring to alternation designs, Ledford (2018) highlights the convenience of block randomization. Therefore, the quantifications proposed in the current text are also applicable to alternation designs with block randomization.

In an adapted ATD (referred to as AATD), in contrast to ATDs, at least two independent behaviors or outcome variables (Byiers, Reichle, & Symmons, 2012) are treated. These behaviors treated are nonreversible and the main aim is to explore which of two effective interventions is more efficient, i.e., enables faster learning (Shepley, Ault, Ortiz, Vogler, & McGee, 2019; Wolery et al., 2018). In an AATD it is critical to have the same number of sessions per condition and the authors “typically randomly select one condition and then automatically conduct the other condition for the next session” (Wolery et al., 2018, p. 315). This suggested way of determining the alternation sequence is consistent with randomized blocks, and, block randomization has actually been used in applied research using an AATD (e.g., Coleman, Cherry, Moore, Park, & Cihak, 2015; Klingbeil, January, & Ardoin, 2019; see also).

CONSISTENCY IN ALTERNATION DESIGNS

An equivalent way of proceeding is followed when both interventions take place during the same day, in two separate sessions, and the order of the sessions is randomly determined (e.g., Cihak, Alberto, Taber-Doughtly, & Gama, 2006) or when in each session both kinds of instruction are present and the order of the instructions is determined at random prior to the beginning of the session (e.g., Savaiano, Compton, Hatton, & Lloyd, 2016). Therefore, the quantifications proposed in the current text are also applicable to AATDs for which the order of the interventions is randomly determined within each block (which can represent a different day or session).

Building a case for using block randomization. The use of randomization within blocks has been recommended when working with alternation designs (Ledford, 2018). Using randomization within blocks ensures meaningful comparisons between measurements belonging to different conditions. As the comparisons are performed within blocks, randomization within blocks minimizes threats to the internal validity of the study. For example, a patient may consistently perform better in condition B than in condition A, but this difference may be an artefact if the order of treatment administration for the two interventions is not randomized. This also makes it easier to apply visual analysis for assessing the degree of differentiation between two data paths, e.g., when comparing adjacent data points (Wolery et al., 2018). By the same logic, randomization facilitates the use and interpretation of quantifications proposed for alternation designs such as the adaptation of the Percentage of nonoverlapping data (Wolery et al., 2014) and the average difference between successive observations (ADISO in Manolov & Onghena, 2018). Moreover, using a sequence that is consistent with an RBD avoids situations with two initial or final administrations of the same condition, which are possible for restricted ATD. Furthermore, in an RBD sequence comparing between data paths using the visual

CONSISTENCY IN ALTERNATION DESIGNS

structured criterion (Lanovaz et al., 2019) or the average difference obtained using actual and linearly interpolated values (ALIV in Manolov & Onghena, 2018) would entail a smaller loss of data.

In the following sections, we first review the proposals made for assessing consistency in other SCEDs, different from alternation designs. Second, we discuss the types of consistency that can be assessed for alternation designs with block randomization, making a proposal for the quantification of consistency of effects. Third, we illustrate the quantifications of consistency for alternation designs with block randomization using fictitious and real data.

Assessing Consistency in SCEDs

We consider that further research is required on how to assess consistency, given that most analytical proposals have focused on overlap (e.g., see Parker, Vannest, & Davis, 2011, for a review), level (e.g., Olive & Smith, 2005; Shadish, Hedges, & Pustejovsky, 2014), trend (in combination with level; Solanas, Manolov, & Onghena, 2010; Swaminathan, Rogers, Horner, Sugai, & Smolkowski, 2014), and immediacy (Center, Skiba, & Casey, 1985-1986; Michiels & Onghena, 2019; Natesan & Hedges, 2017). In contrast, the assessment of consistency has been restricted to “an overall gestalt analysis” (Geist & Hitchcock, 2014, p. 304) or to somewhat tautological recommendations such as “the extent to which there is consistency in the data patterns from phases with the same conditions” (Kratochwill et al., p. 19).

Regarding some specific proposals for addressing consistency, Maggin, Briesch, and Chafouleas (2013) suggest that the ratio of effects to no-effects within a study, should be at least 3:1, in order to constitute evidence for an intervention effect. For instance, in a multiple-baseline design across four participants, this would mean the need to demonstrate an effect for at least

CONSISTENCY IN ALTERNATION DESIGNS

three of these participants. The question still remains how an effect is objectively demonstrated in each of the AB-comparisons. A recent protocol on visual analysis attempts to make the visual assessment more systematic (Wolfe, Barton, & Meadan, 2019), but does not address this specific question of how an effect is objectively defined.

Consistency in Phase Designs

One formal treatment of consistency in SCEDs focuses on ABAB designs (Tanious, De, Michiels, Van den Noortgate, & Onghena, 2019a) and on multiple-baseline and changing criterion designs (Tanious, Manolov, & Onghena, 2019). There is one quantification of the consistency of data patterns for measurements taken in the same conditions, performing a point-by-point comparison using the Manhattan distance. This quantification is called CONDAP and it is applicable even if the two phases differ in the number of data points and regardless of the measurement units of the target variable. For CONDAP, there are interpretative benchmarks available helping applied researchers evaluate whether the consistency is very high, high, medium, low, or very low (Tanious, De, Michiels, Van den Noortgate, & Onghena 2019b). A second quantification has been proposed for the consistency of effects (changes in level, trend, variability, overlap, immediacy) when comparing across adjacent conditions (Tanious, De, et al., 2019a). This quantification is called CONEFF.

Consistency in Alternation Designs

Assessing consistency is important for alternation designs. For instance, when describing the visual analysis of ATD data, Wolery, Gast, and Ledford (2018) state that the aim is to assess the degree of differentiation between data paths and “differentiation is defined as a *consistent* difference in level between adjacent data points from different conditions” (p. 330, emphasis

CONSISTENCY IN ALTERNATION DESIGNS

added). To the best of our knowledge, no quantifications have been proposed or discussed specifically for assessing consistency in alternation designs. Nevertheless, some potentially applicable options, derived from the existing literature, are discussed next.

Wolery, Gast, and Ledford (2014) describe an adaptation of the Percentage of nonoverlapping data, which for an alternating treatments design would be computed by comparing the first measurement in one condition to the first measurement in the other condition, and so forth. If there are five measurements per condition (and a sequence of ten measurement occasions), there would be five comparisons. The final quantification is the percentage of comparisons for which one condition is superior to the other. Such a quantification could be conceptualized as a quantification of consistency of superiority (the closer the percentage to 100%, the more consistently that one condition is better than the other).

Similarly, Lanovaz, Cardinal, and Francis (2019) propose a comparison between data paths (i.e., the lines connecting the measurements for each condition). If there is a sequence of ten measurement occasions (with five measurements per condition, e.g., ABBAABBAAB), there would be eight comparisons, excluding the first and the last measurement occasion for which there is only one data path (e.g., the first A measurement and the last B measurement). Just as for the Percentage of nonoverlapping data described previously, the proportion of comparisons for which the one condition is superior (ordinally) to the other is tallied. Such a comparison could also be understood as leading to an assessment of consistency of superiority.

A visual approach (see Mengersen, McGree, & Schmid, 2015) to assessing superiority of one condition entails using a modified Brinley plot (Blampied, 2017). In this graphical representation the measurements from condition A are plotted against the measurements from condition B, corresponding to the same. A diagonal line is drawn representing no treatment effect. In case all

CONSISTENCY IN ALTERNATION DESIGNS

measurements from condition B values are superior to the measurements from condition A (e.g., above the diagonal line), once again the *presence* of an effect can be considered consistent, but not its *magnitude* (Mengersen et al., 2015).

All three options focus on the effects of the intervention, understood as the difference between conditions. These options for assessing the consistency of effects can be considered ordinal in that they do not evaluate whether the amount of difference between conditions is consistent in the different comparisons performed throughout the alternation sequence. That is, the previously mentioned analytical tools cannot be used to assess the consistency of the magnitude of effect.

Types of Consistency in an Alternation Design with Block Randomization

In the current section we discuss the different kinds of consistency that can be assessed in an alternation design and discussing possible quantifications. A more in-depth look into the interpretation and meaningfulness of these quantifications is presented in the next section, via illustrations.

Consistency in Similar Phases

Alternation designs do not entail comparing conditions across phases. Therefore, we consider that an assessment of the consistency of the data patterns in similar phases (e.g., Kratochwill et al., Ledford et al., 2019) would not make sense in this context. As a quantification of the (lack of) consistency of measurements in each condition, the standard deviation for all measurements belonging to the same condition could be computed. Nevertheless, such a quantification would not reflect any data pattern, as a pattern cannot be established when there is a single measurement per condition in each block.

Consistency of Superiority

CONSISTENCY IN ALTERNATION DESIGNS

Previously in the text, we reviewed several graphical and quantitative options that could be understood as assessing the superiority of one condition over the other. In contrast, in the following two sections, we present two quantifications of consistency which go beyond the ordinal information that can be obtained from assessing the consistency in superiority.

Consistency of the Average Level across Blocks

A single-case RBD is mathematically analogous to an RBD from group designs, in which there is a single participant in each cell, defined by the levels of the blocking variable and the treatment variable. For a group-design RBD, suppose that we are comparing two treatments, A and B, and that the blocks are matched pairs of participants (e.g., according to their age). Within each pair, it is randomly determined who receives treatment A and who receives treatment B. An analysis of variance for data collected in such an RBD consists in the independent partitioning of the variance explained by the treatment factor and the variance explained by the blocking factor (Kirk, 2013). The same can be done for an RBD as an SCED, although the blocks do not consist of participants, but consist of measurement occasions. The variability across blocks is the degree to which the average value for each block is different from the overall/grand mean (i.e., the mean of all measurements, regardless of the condition that they were obtained in). In an SCED, the average per block mixes a measurement in condition A and a measurement in condition B. Therefore, it does not inform about consistency of measurements in similar conditions or about consistency of effects across blocks. Moreover, the temporal order of the blocks is not taken into account, which further limits its usefulness.

According to the variance partitioning, the greater the variability attributed to the blocking factor, the greater the difference of the average level across blocks, and the lower the consistency of the average level across blocks. The variability attributed to blocking can be summarized as an

CONSISTENCY IN ALTERNATION DESIGNS

eta-squared, $\eta_{blocking}^2 = \frac{SS_{blocking}}{SS_{total}}$, where SS denotes sum of squares. Thus, it is possible to define the percentage consistency of the average level across blocks, as the complementary quantity, $\left(1 - \frac{SS_{blocking}}{SS_{total}}\right) \times 100$. In any case, it should be noted that the consistency of the average level across blocks is an attempt to interpret the variability attributed to blocking variable. However, we are not suggesting that the consistency of the average level across blocks is always meaningful or that it should be the main quantification of consistency for an alternation design. (Our main proposal for assessing consistency is described next.) Actually, for an AATD, the consistency of the average level across blocks is not desired, because an improvement is expected in both conditions (i.e., a trend), which would entail that the level is lower in the beginning of the alternating sequence and higher in the end.

Consistency of Effects across Blocks

In contrast to the consistency of the average level, the consistency of effects focuses on whether the *difference* between conditions is the same across blocks, regardless of the average level for the block. In that sense, the consistency of effect across blocks helps distinguishing between trends with parallel slopes and trends with different slopes, as will be illustrated later. The consistency of the average level across blocks compares the average in each block with the grand mean. The consistency of effects across blocks compares whether the difference in each block is similar to the mean difference between conditions. Thus, conceptually the latter is of greater interest for applied researchers.

Numerically, when partitioning the variance, apart from the effect of the intervention and the effect of blocking, there is likely to be residual variability (i.e., the variability that is left unexplained by the intervention and the blocking variables). Such residual variability would

CONSISTENCY IN ALTERNATION DESIGNS

represent both the error variability (random fluctuations, systematic variation due to factors not included in the model) and the interaction between intervention and blocking (Hays, 1994; Kirk, 2013). This is the case because the interaction cannot be separated from the error when there is only one measurement per cell (i.e., combination of a level of the blocking variable and a level of the treatment variable). Actually, if interaction were present, this would entail that the magnitude of the effect of the intervention depends on the blocking variable. Thus, an interaction would represent a lack of consistency of effects across blocks. That is, the degree of lack of consistency of effect across blocks is represented by the extent to which the difference between a measurement from condition A and a measurement for condition B differs across blocks.

Complementarily, the consistency of effects across blocks (abbreviated, CEAB), expressed as a percentage, would be: $CEAB = \left(1 - \frac{SS_{residual}}{SS_{total}}\right) \times 100$. In this expression, the sum of squares are computed exactly as in a two-way ANOVA, in which one of the factors is the intervention (condition A vs. condition B) and the other factor is the blocking variable (which has as many values as there are blocks). Specifically, the total sum of squares (SS_{total}) is, as usual, the sum of the squared differences between each measurement and the overall mean (computed as the average of all measurements, regardless of the condition and block that they belong to). The residual sum of squares ($SS_{residual}$) is the variability left unexplained from SS_{total} after removing (a) the variability attributed to the intervention (i.e., the sum of squared differences between the mean in each condition and the overall mean, multiplied by the number of blocks) and (b) the variability attributed to the blocking variable (i.e., the sum of squared difference between the mean level in each block and the overall mean, multiplied by number of conditions).

CONSISTENCY IN ALTERNATION DESIGNS

Illustrations of the Types of Consistency for Alternation Designs with Block

Randomization

The illustrations provided here are intended to help gaining a better understanding of the previously presented quantifications in different situations. Moreover, we are proposing the use of several graphical representations for making easier the assessment of consistency in alternation designs. The first six examples provided include fictitious data, whereas the last three use real data.

Figure 1 represents fictitious data in which the effect is consistent across blocks (there is always a 2-point difference between the A condition and the B condition), but the average level for some blocks is higher. These two types of consistency are visible from the upper right panel of Figure 1. In this panel, the horizontal red line represents the grand mean and the thick green line represent the average difference between the A-condition and the B-condition (the greater the slope, the greater the difference). Each A-measurement is connected to its corresponding B-measurement from the same block with a dashed line. The order of the A and the B measurements within the block is not represented on the graph; neither is the order of the blocks within the whole alternation sequence. The consistency of effects across blocks is represented in the dashed lines being parallel (versus crossing for lack of consistency of effect). The degree to which the average level of the blocks is not consistent is represented by the vertical distance between the middle points of the dashed lines and the middle point of the thick green line.

The data in Figure 2 (upper left panel) show measurements which follow a similar decreasing trend in both conditions. As for the data in Figure 1, the grey dashed lines from the upper right panel are parallel and the consistency of effects across blocks is complete (i.e., $CEAB = 100\%$). The greater separation between the dashed lines from the upper right panel illustrate the

CONSISTENCY IN ALTERNATION DESIGNS

fact that there is a greater difference of the average level across blocks ($\eta_{blocking}^2 = 0.20$ for data in Figure 1 and $\eta_{blocking}^2 = 0.74$ for the data in Figure 2). This is an example illustrating how the presence of a similar trend in both conditions is represented as a lack of consistency of the average level across blocks.

CONSISTENCY IN ALTERNATION DESIGNS

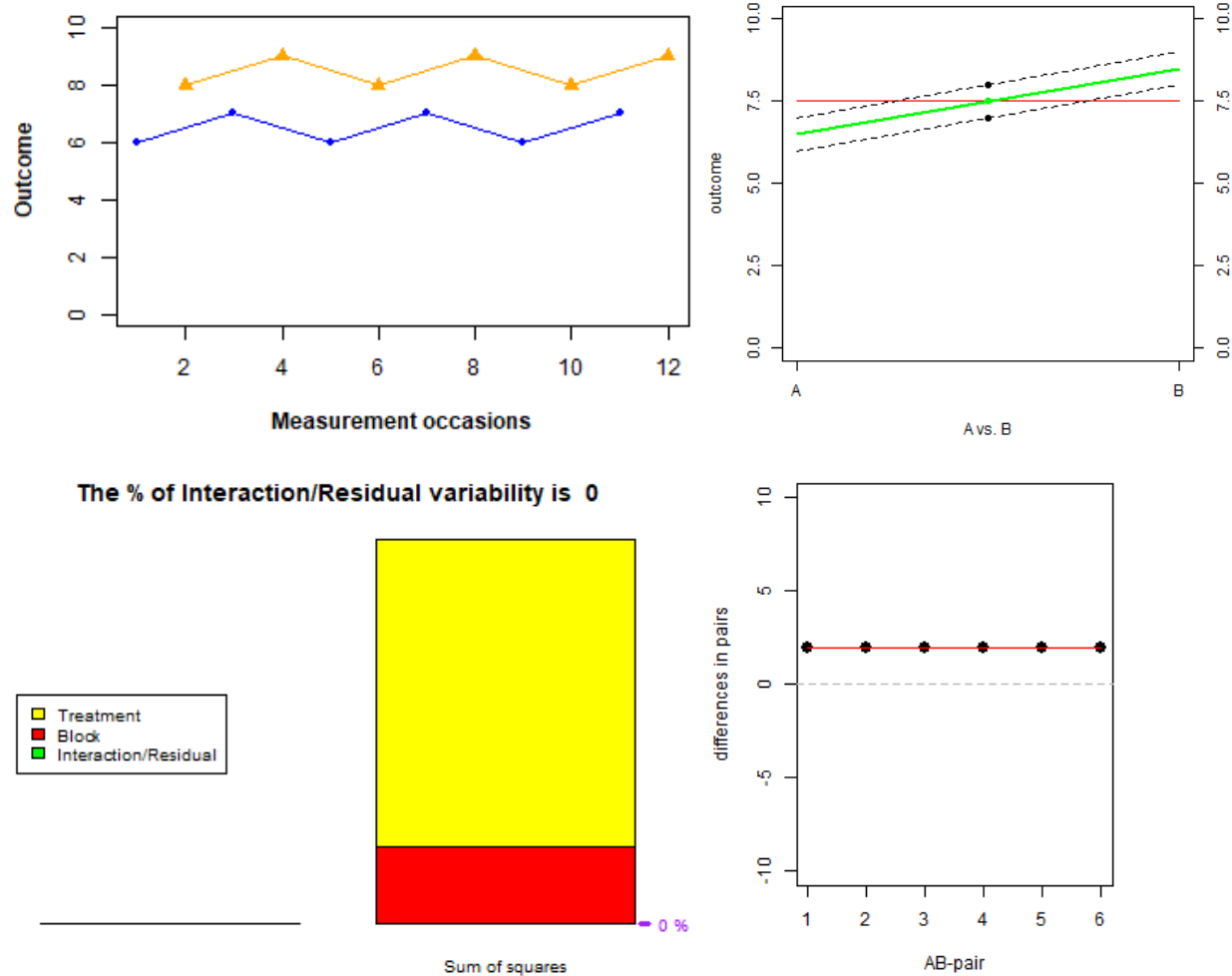


Figure 1. Fictitious example 1: CEAB=100%. The upper left panel is a time series plot with the A condition in blue and the B condition in orange. The upper right panel represents the A values on the left Y-axis, connected with a dashed line to the corresponding B-values on the right Y-axis. Only two dashed lines are visible due to overlapping across blocks. The horizontal red line represents the grand mean of the outcomes, whereas the thick green line connects the mean of the A-values to the mean of the B-values. The lower left panel represents the proportions of variability attributed to treatment (yellow area), blocking effect (red area) and interaction/residual (green area): the percentage to the left is the residual/interaction variability in relation to block plus residual variability, whereas the percentage to the right is the residual/interaction variability in relation to the total variability (including the treatment effect). The lower right panel represents the difference between the A and B measurement in each block, as compared to the average mean difference (horizontal red line).

CONSISTENCY IN ALTERNATION DESIGNS

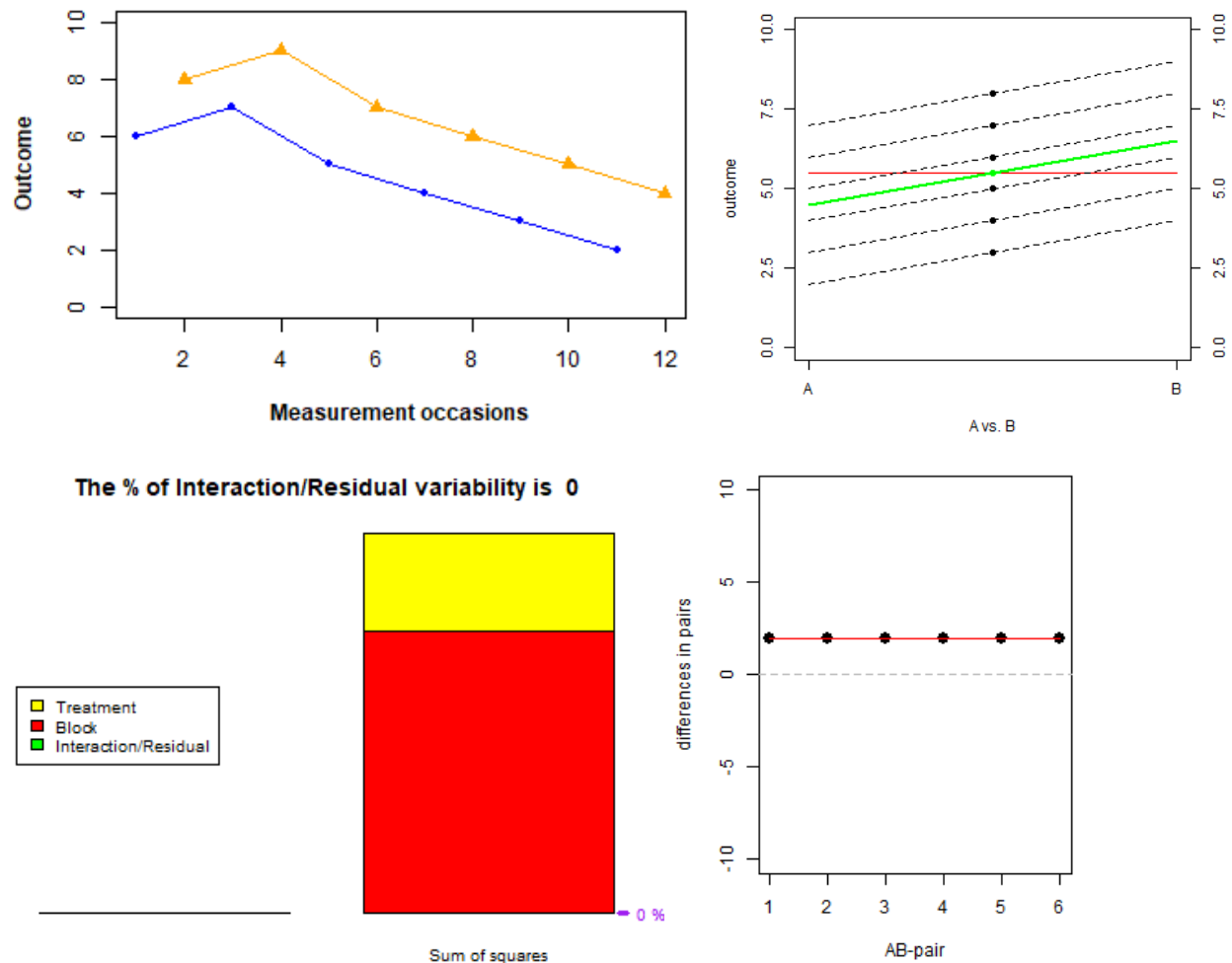


Figure 2. Fictitious example 2: CEAB=100%. The upper left panel is a time series plot with the A condition in blue and the B condition in orange. The upper right panel represents the A values on the left Y-axis, connected with a dashed line to the corresponding B-values on the right Y-axis. The horizontal red line represents the grand mean of the outcomes, whereas the thick green line connects the mean of the A-values to the mean of the B-values. The lower left panel represents the proportions of variability attributed to treatment (yellow area), blocking effect (red area) and interaction/residual (green area): the percentage to the left is the residual/interaction variability in relation to block plus residual variability, whereas the percentage to the right is the residual/interaction variability in relation to the total variability (including the treatment effect). The lower right panel represents the difference between the A and B measurement in each block, as compared to the average mean difference (horizontal red line).

CONSISTENCY IN ALTERNATION DESIGNS

The time series pattern of Figure 3 (upper left panel) is U-shaped in both conditions. There is not a complete consistency of effects across blocks, because two of the dashed lines cross ($CEAB = 93.65\%$). In terms of the consistency of the average level across blocks, it is equal to the one for Figure 2 ($\eta_{blocking}^2 = 0.74$), because the U-shaped pattern just like the decreasing trend also introduces lower consistency of the average level.

The Figure 4 data are in sharp contrast with Figure 3, in terms of data pattern (more stable here) and in terms of consistency. For the Figure 4 data, the dashed lines in the upper right panel cross to a greater extent, indicating lower consistency of effects across blocks ($CEAB = 74.81\%$). These lines are closer together, reflecting that there is greater consistency of the average level across blocks ($\eta_{blocking}^2 = 0.13$).

CONSISTENCY IN ALTERNATION DESIGNS

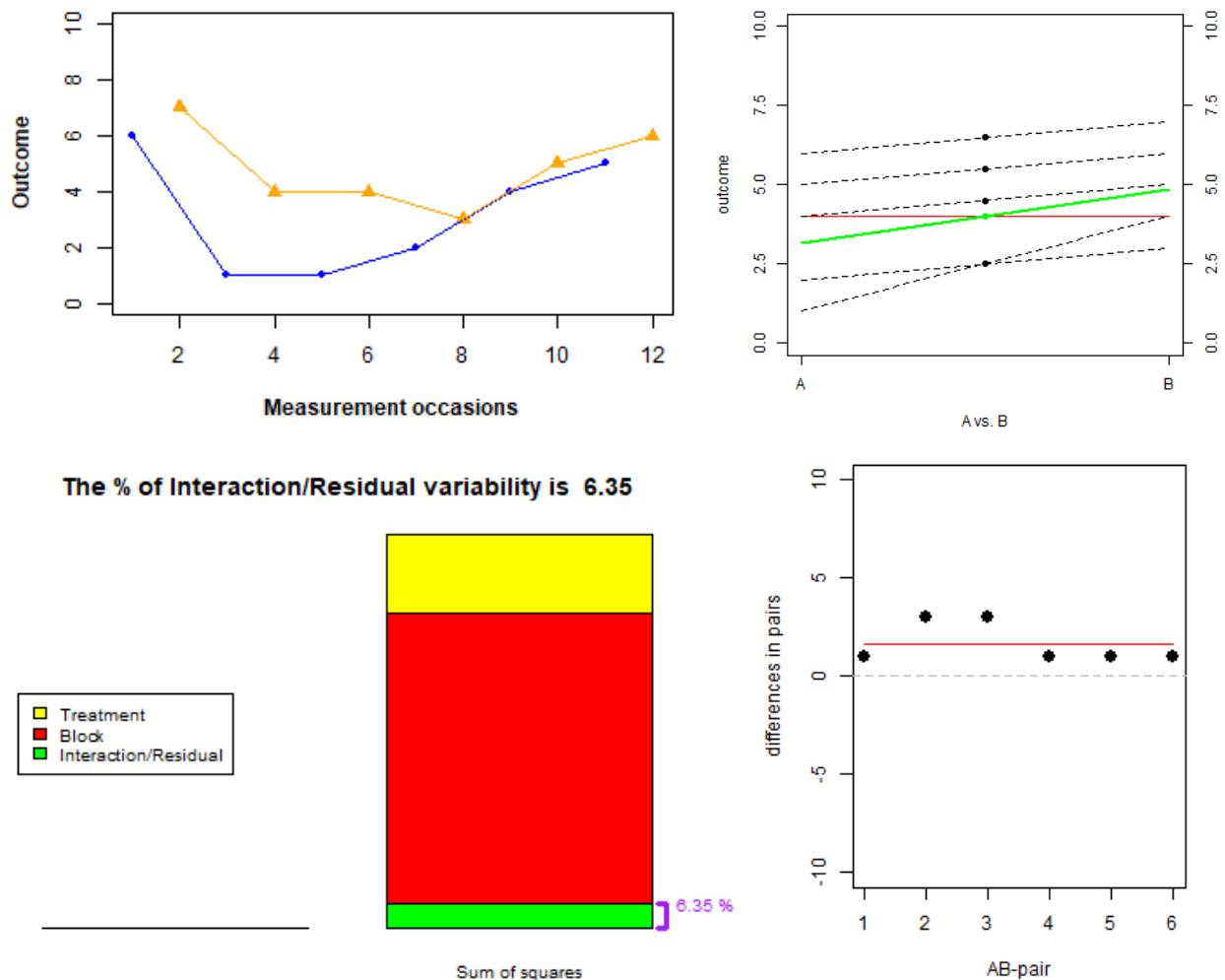


Figure 3. Fictitious example 3: CEAB=93.65%. The upper left panel is a time series plot with the A condition in blue and the B condition in orange. The upper right panel represents the A values on the left Y-axis, connected with a dashed line to the corresponding B-values on the right Y-axis. The horizontal red line represents the grand mean of the outcomes, whereas the thick green line connects the mean of the A-values to the mean of the B-values. The lower left panel represents the proportions of variability attributed to treatment (yellow area), blocking effect (red area) and interaction/residual (green area): the percentage to the left is the residual/interaction variability in relation to block plus residual variability, whereas the percentage to the right is the residual/interaction variability in relation to the total variability (including the treatment effect). The lower right panel represents the difference between the A and B measurement in each block, as compared to the average mean difference (horizontal red line).

CONSISTENCY IN ALTERNATION DESIGNS

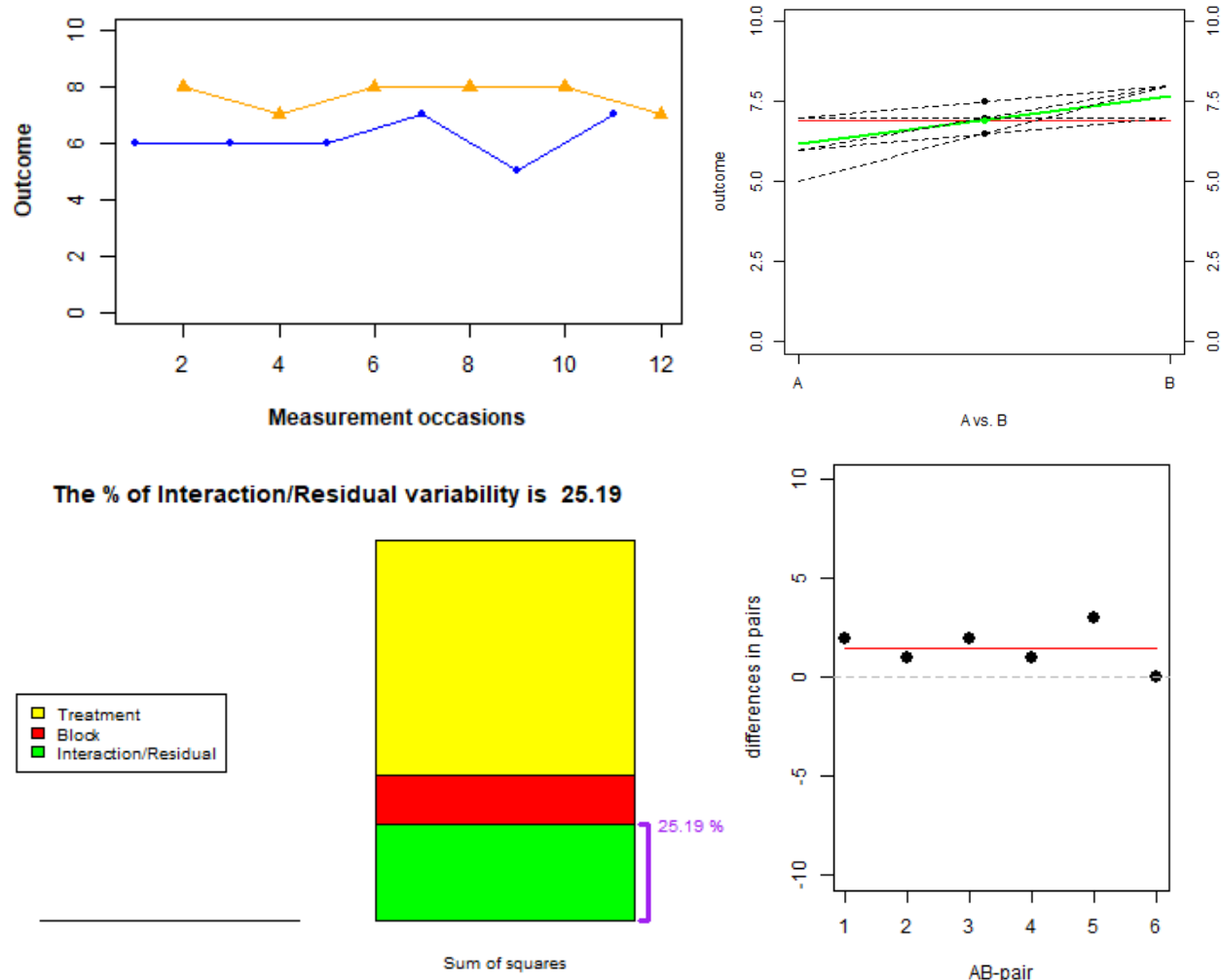


Figure 4. Fictitious example 4: CEAB=74.81%. The upper left panel is a time series plot with the A condition in blue and the B condition in orange. The upper right panel represents the A values on the left Y-axis, connected with a dashed line to the corresponding B-values on the right Y-axis. The horizontal red line represents the grand mean of the outcomes, whereas the thick green line connects the mean of the A-values to the mean of the B-values. The lower left panel represents the proportions of variability attributed to treatment (yellow area), blocking effect (red area) and interaction/residual (green area): the percentage to the left is the residual/interaction variability in relation to block plus residual variability, whereas the percentage to the right is the residual/interaction variability in relation to the total variability (including the treatment effect). The lower right panel represents the difference between the A and B measurement in each block, as compared to the average mean difference (horizontal red line).

CONSISTENCY IN ALTERNATION DESIGNS

Figure 5 shows data with considerable variability. In terms of the consistency of effects across blocks, three of the dashed lines (upper right panel) are parallel, indicating that the effect for three blocks is the same. However, the effect for the other two blocks is markedly different: large and positive for one block and negative for the other (see lower right panel). This lower consistency of effects is represented by $CEAB = 60\%$. In terms of consistency of the average level across blocks ($\eta_{blocking}^2 = 0.38$), it is lower than for the Figure 4 data, but higher than for the Figure 3 data, because three of the block averages coincide and are very close to the overall mean (see the upper right panel of Figure 5).

CONSISTENCY IN ALTERNATION DESIGNS

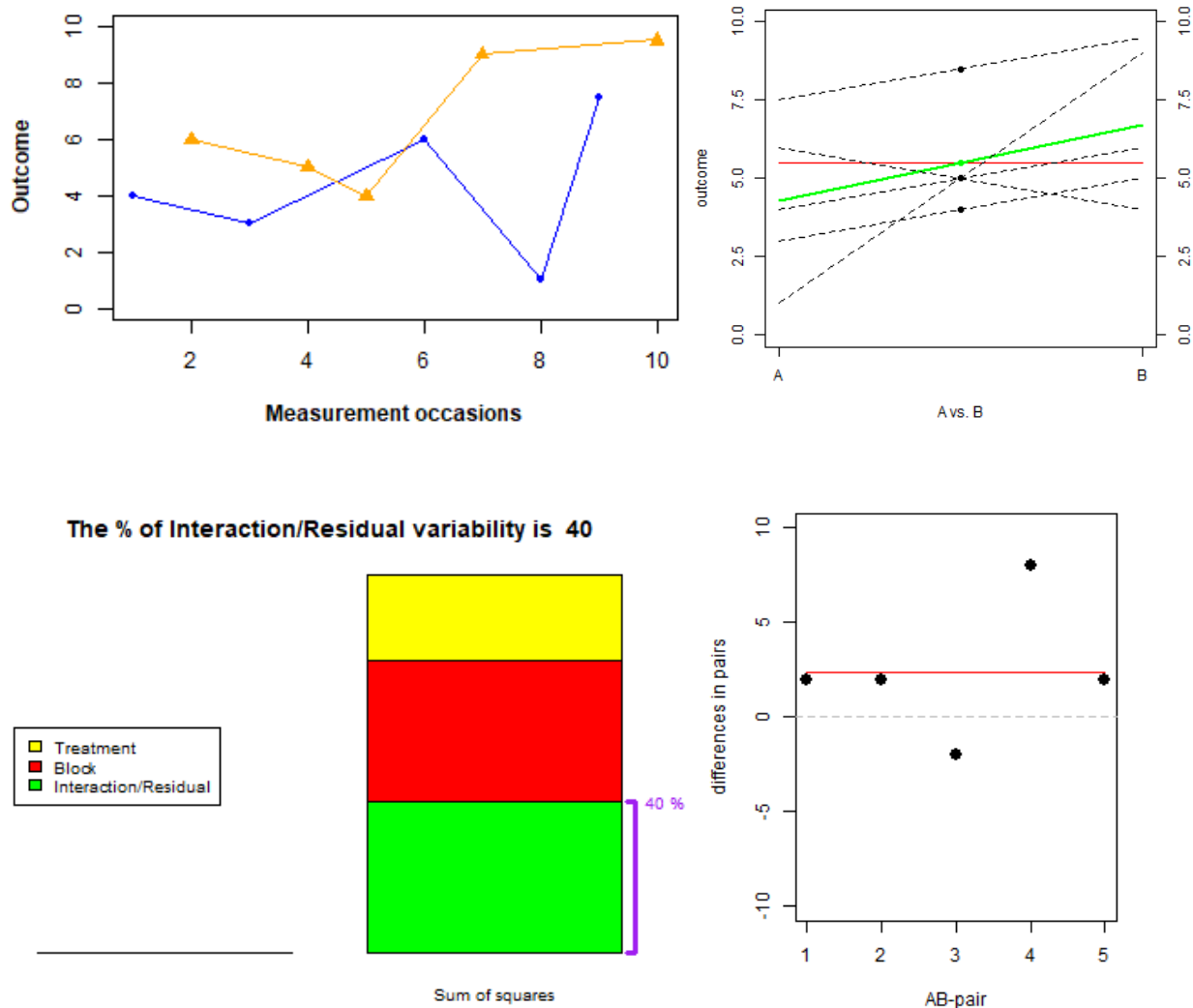


Figure 5. Fictitious example 5: CEAB=60%. The upper left panel is a time series plot with the A condition in blue and the B condition in orange. The upper right panel represents the A values on the left Y-axis, connected with a dashed line to the corresponding B-values on the right Y-axis. The horizontal red line represents the grand mean of the outcomes, whereas the thick green line connects the mean of the A-values to the mean of the B-values. The lower left panel represents the proportions of variability attributed to treatment (yellow area), blocking effect (red area) and interaction/residual (green area): the percentage to the left is the residual/interaction variability in relation to block plus residual variability, whereas the percentage to the right is the residual/interaction variability in relation to the total variability (including the treatment effect). The lower right panel represents the difference between the A and B measurement in each block, as compared to the average mean difference (horizontal red line).

CONSISTENCY IN ALTERNATION DESIGNS

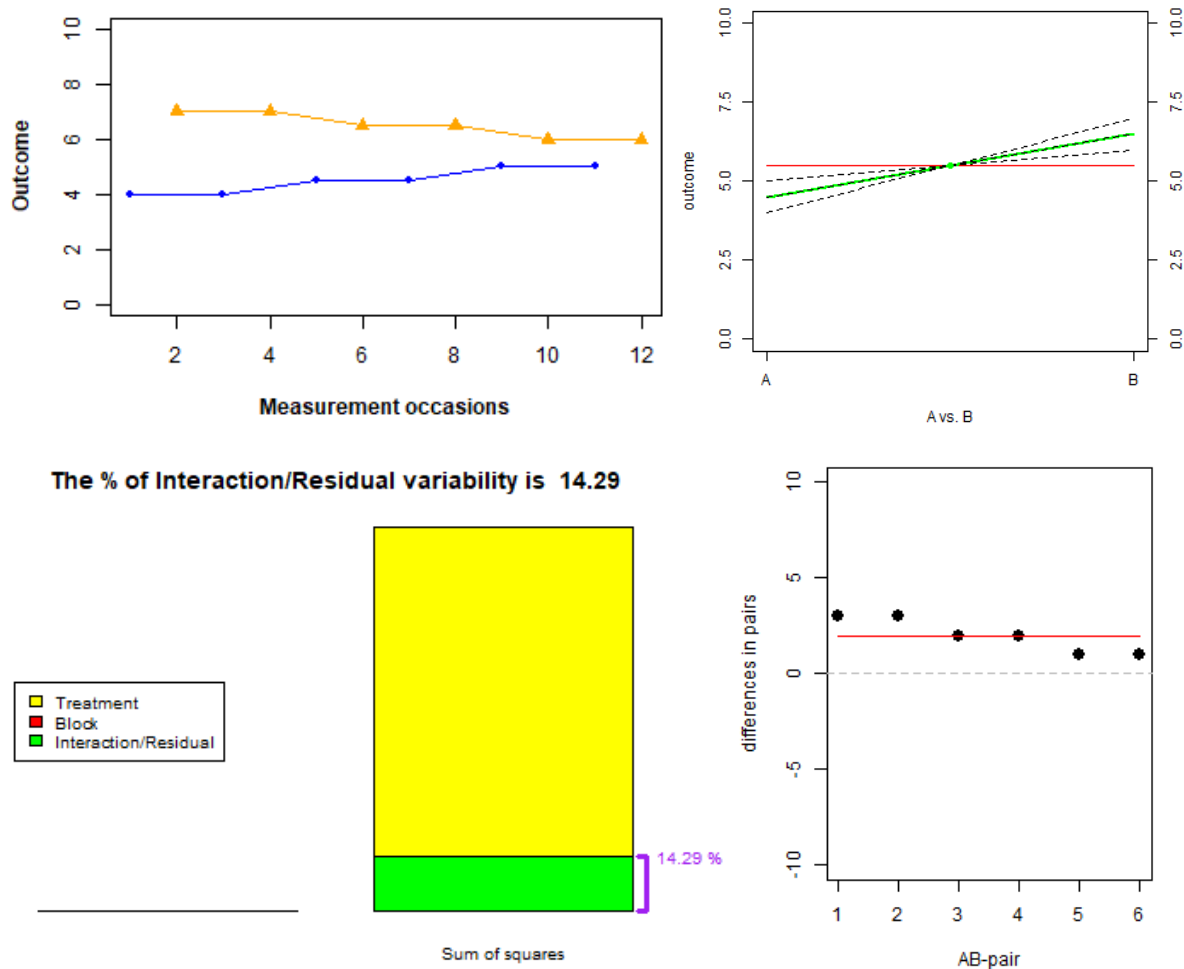


Figure 6. Fictitious example 6: CEAB=85.71%. The upper left panel is a time series plot with the A condition in blue and the B condition in orange. The upper right panel represents the A values on the left Y-axis, connected with a dashed line to the corresponding B-values on the right Y-axis. Only two dashed lines are visible due to overlapping across blocks. The horizontal red line represents the grand mean of the outcomes, whereas the thick green line connects the mean of the A-values to the mean of the B-values. The lower left panel represents the proportions of variability attributed to treatment (yellow area), blocking effect (red area) and interaction/residual (green area): the percentage to the left is the residual/interaction variability in relation to block plus residual variability, whereas the percentage to the right is the residual/interaction variability in relation to the total variability (including the treatment effect). The lower right panel represents the difference between the A and B measurement in each block, as compared to the average mean difference (horizontal red line).

CONSISTENCY IN ALTERNATION DESIGNS

Figure 6 represents data in which there are opposite trends in the two conditions (see upper left panel) and the same average level across blocks (i.e., the dashed lines share the same middle point on the upper right panel). This is relevant for the interpretation of the consistency quantifications. On the one hand, all the dashed lines sharing the same middle point implies that there is no effect of blocking ($\eta_{blocking}^2 = 0$) and it could be stated that there is perfect consistency of the average level across blocks. Nonetheless, note that this does not mean that all the A-measurements or the B-measurements are the same across blocks, or that the intervention effect is the same across blocks; just that the average of each block is equal to the grand mean. In terms of consistency of effects across blocks, there is consistency in the superiority of one condition over the other, but not perfect consistency in the magnitude of effect ($CEAB = 85.71\%$), because some of the dashed lines cross (see the upper right panel of Figure 6). The consistency is high, because the A-B differences for two of the blocks are exactly equal to the average A-B difference and there are two other values of the A-B difference that are repeated twice (see the lower right panel of Figure 6). The fact that certain values of the A-B differences are present more than once, with the A and B values also coinciding is represented by the fact that there are some of the dashed lines of the upper left panel are overlapping.

CONSISTENCY IN ALTERNATION DESIGNS

Figure 7 represents a data pattern in which there is no intervention effect. Actually, half of the variability is attributed to blocks and half to the interaction between blocks and intervention.

$CEAB = 50\%$ would be misleading, in case it is interpreted in isolation, but it has to be evaluated only in relation to the fact that there is actually no intervention effect (i.e., no variability explained by the intervention). Thus, it does not make sense to evaluate the consistency of an inexistent effect.

CONSISTENCY IN ALTERNATION DESIGNS

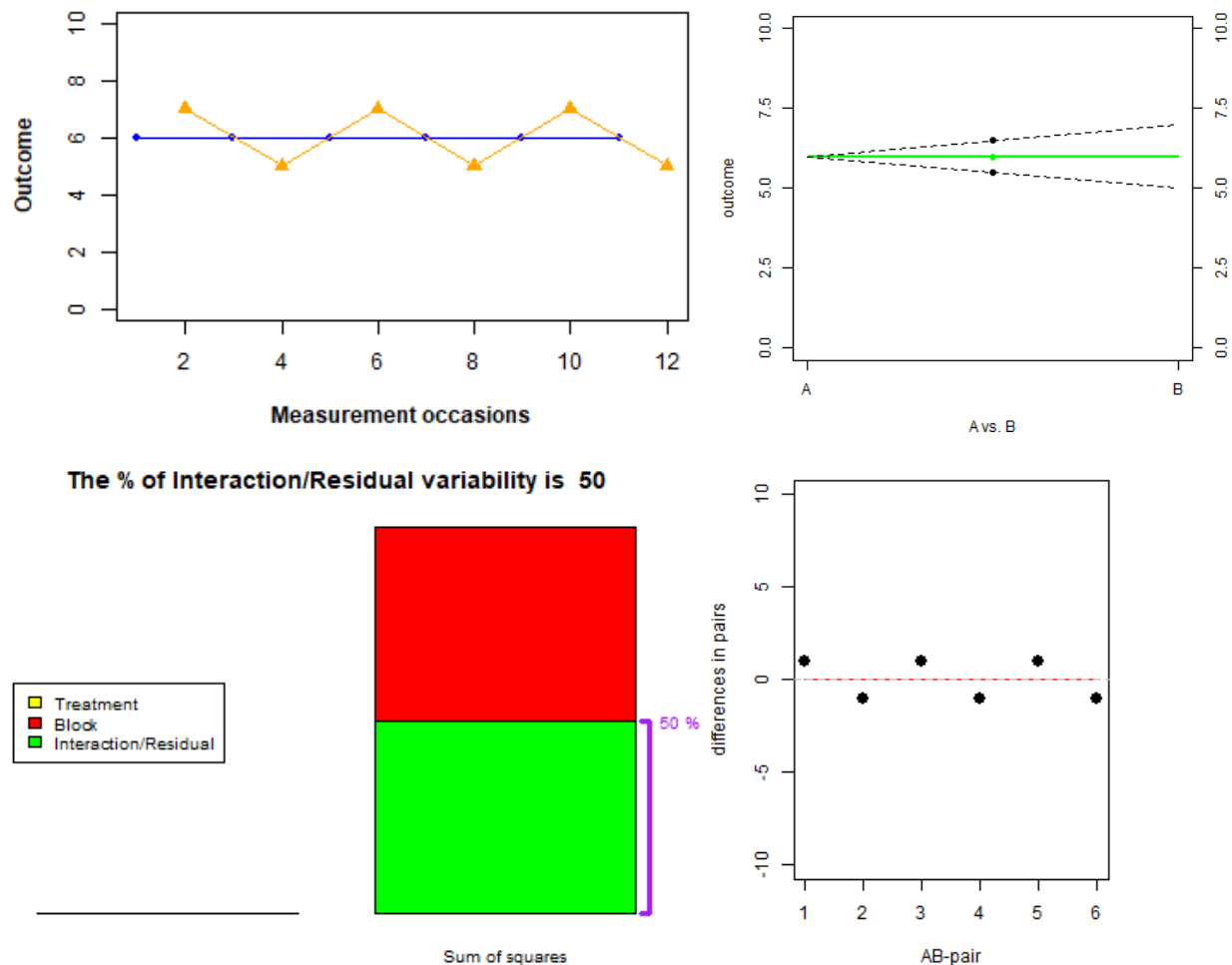


Figure 7. Fictitious example 7: CEAB=50%. The upper left panel is a time series plot with the A condition in blue and the B condition in orange. The upper right panel represents the A values on the left Y-axis, connected with a dashed line to the corresponding B-values on the right Y-axis. Only two dashed lines are visible due to overlapping across blocks. The horizontal red line represents the grand mean of the outcomes, whereas the thick green line connects the mean of the A-values to the mean of the B-values. The lower left panel represents the proportions of variability attributed to treatment (yellow area), blocking effect (red area) and interaction/residual (green area): the percentage to the left is the residual/interaction variability in relation to block plus residual variability, whereas the percentage to the right is the residual/interaction variability in relation to the total variability (including the treatment effect). The lower right panel represents the difference between the A and B measurement in each block, as compared to the average mean difference (horizontal red line).

CONSISTENCY IN ALTERNATION DESIGNS

The first example with real data focuses on the measurements of duration of hand-flapping obtained by Lloyd et al. (2018) from a participant called Martin, diagnosed with autism spectrum disorder and attention deficit hyperactivity disorder and presenting stereotypy (see Figure 8). For these data, across blocks, there is a mixture of five smaller A-B differences (similar among themselves) and five larger A-B differences (also similar among themselves, but different from the smaller A-B differences); this is visible from the lower right panel of Figure 8. As represented on the upper right panel, there are also several parallel dashed, that are crossing with several other parallel dashed lines. Numerically, this is summarized as $CEAB = 75.64\%$.

CONSISTENCY IN ALTERNATION DESIGNS

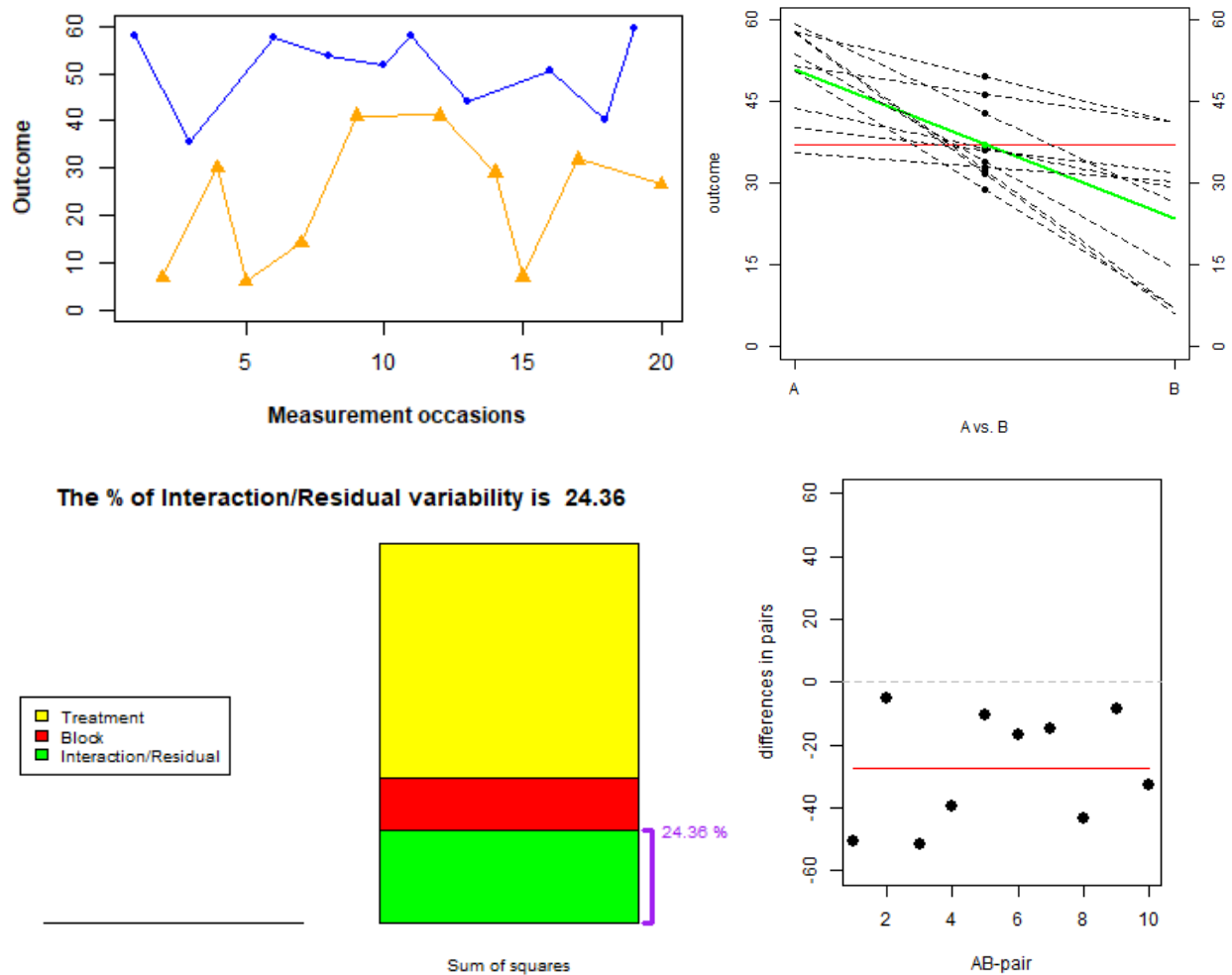


Figure 8. Lloyd et al. (2018) data: duration of hand-flapping – CEAB=75.64%. The upper left panel is a time series plot. The upper right panel represents the A values on the left Y-axis, connected with a dashed line to the corresponding B-values on the right Y-axis. The horizontal red line represents the grand mean of the outcomes, whereas the thick green line connects the mean of the A-values to the mean of the B-values. The lower left panel represents the proportions of variability attributed to treatment (yellow area), blocking effect (red area) and interaction/residual (green area): the percentage to the left is the residual/interaction variability in relation to block plus residual variability, whereas the percentage to the right is the residual/interaction variability in relation to the total variability (including the treatment effect). The lower right panel represents the difference between the A and B measurement in each block, as compared to the average mean difference (horizontal red line). The lower right panel represents the difference between the A and B measurement in each block, as compared to the average mean difference (horizontal red line).

CONSISTENCY IN ALTERNATION DESIGNS

The measurements of latency to hand-flapping by Martin, as obtained by Lloyd et al. (2018), are represented in Figure 9 (left panel), in order to include data with an apparent outlier. (It is not clear whether methods for detecting outliers such as the boxplot rule, [Tukey, 1977] or the rule based on the median of absolute deviations [Leys, Ley, Klein, Bernard, & Licata, 2013] are reasonable when there a few (e.g., five or six) measurements in a condition.). An outlier introduces lack of consistency. For instance, the outlier is related to certain line-crossing (upper right panel), but this is not the only reason for the lower consistency ($CEAB = 58.25\%$). Other contributions to the lack of consistency are visible in the lower right panel of Figure 9: there is one difference with a negative sign and two differences that are very close to zero, with the majority of differences being positive and close to 10 (seconds), plus a very large positive difference. Moreover, an outlier makes the average of one of the blocks farther away from the overall mean, leading to lower consistency of the average level across blocks ($\eta_{blocking}^2 = 0.32$ vs. $\eta_{blocking}^2 = 0.14$ for the Figure 8 data).

CONSISTENCY IN ALTERNATION DESIGNS

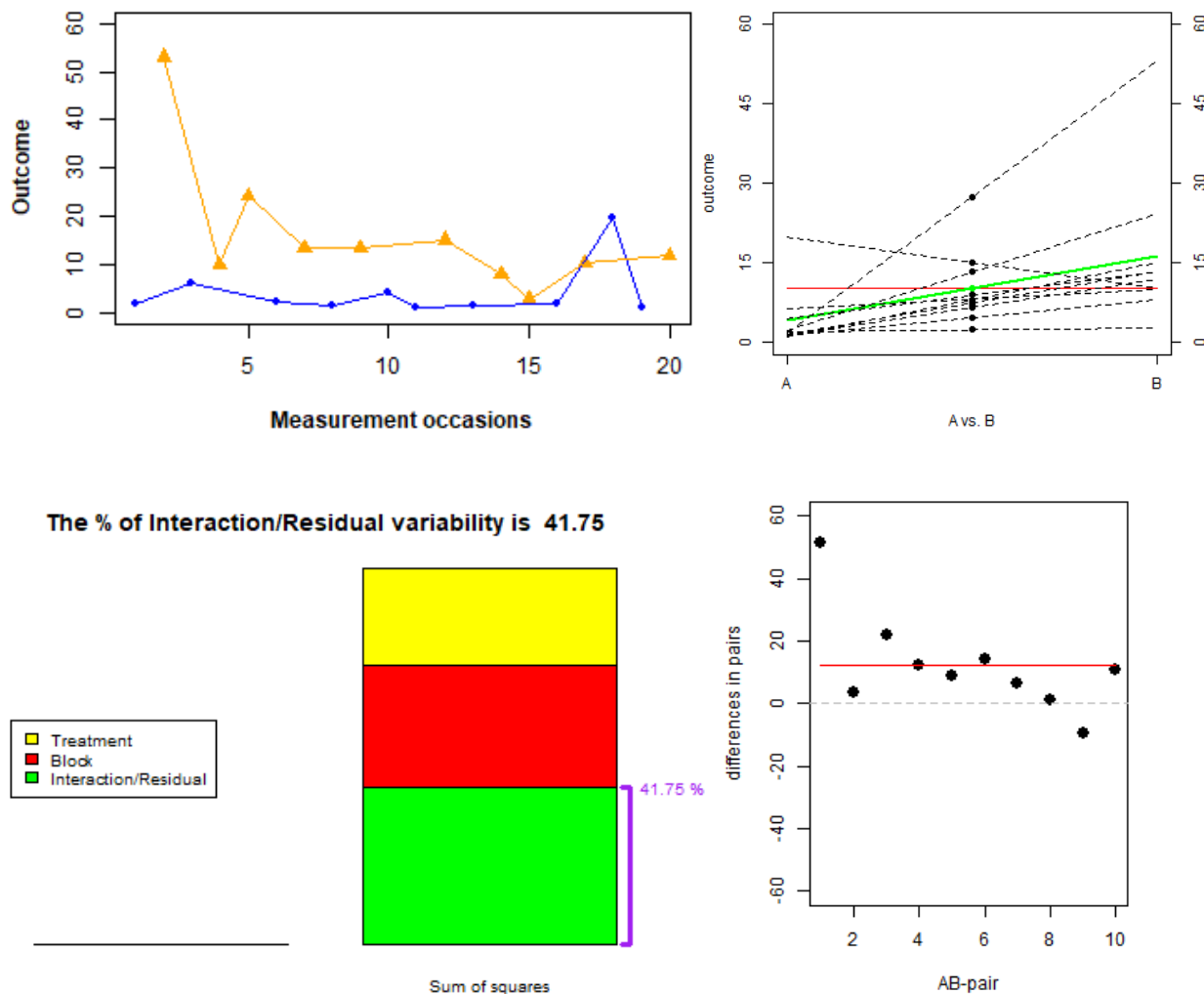


Figure 9. Lloyd et al. (2018) data: duration of hand-flapping – CEAB=58.25%. The upper left panel is a time series plot. The upper right panel represents the A values on the left Y-axis, connected with a dashed line to the corresponding B-values on the right Y-axis. The horizontal red line represents the grand mean of the outcomes, whereas the thick green line connects the mean of the A-values to the mean of the B-values. The lower left panel represents the proportions of variability attributed to treatment (yellow area), blocking effect (red area) and interaction/residual (green area): the percentage to the left is the residual/interaction variability in relation to block plus residual variability, whereas the percentage to the right is the residual/interaction variability in relation to the total variability (including the treatment effect). The lower right panel represents the difference between the A and B measurement in each block, as compared to the average mean difference (horizontal red line).

CONSISTENCY IN ALTERNATION DESIGNS

Application to AATDs

In an AATD, the level of responding is not expected to be consistent across blocks, because the participant is expected to improve in both conditions (probably faster in one of them). The lack of a consistent average level of responding across blocks would be quantified by the effect of the blocking variable. With CEAB (the consistency of effects across blocks), we can assess whether the difference between conditions is consistent in size. Three scenarios are possible: (a) one of the conditions is not consistently and clearly superior to the other throughout the whole data series (e.g., Cihak et al., 2006, data for Group 1; Klingbeil et al., 2019, data for participant Carlos); (b) the difference between the conditions is of a very similar size for all measurement occasions (e.g., Klingbeil et al., 2019, data for participant Zoe; Savaiano et al., 2016, data for participant Helen); or (c) the difference between conditions increases with time (Coleman et al., 2015, data for participant Alice; Klingbeil et al., 2019, data for participant Daniela).

The data for Carlos (Klingbeil et al., 2019) are represented on Figure 10. The difference between the conditions is not very clear, neither in terms of differentiation nor in terms of efficiency (speed of improvement or slope of the trend line). In that sense, the variability attributed to the intervention is practically zero, whereas the variability attributed to blocking is very high (see the distance between the lines on the Y-axis in the upper right panel of Figure 10). When the effect is almost null (i.e., differences within blocks very close to zero, some positive, some negative – as per the lower right panel of Figure 10) it does not make sense to discuss the consistency of effect across blocks.

CONSISTENCY IN ALTERNATION DESIGNS

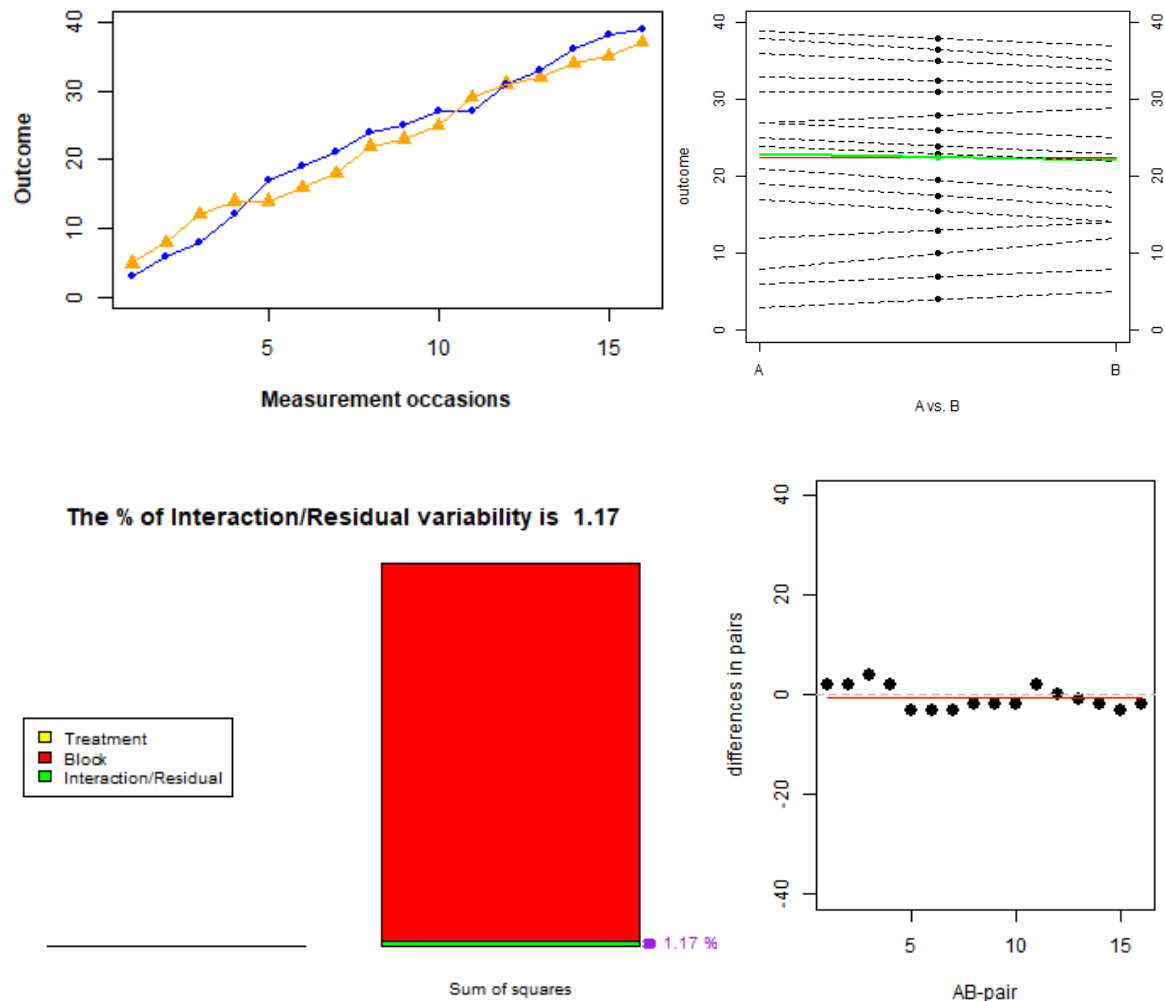


Figure 10. Klingbeil et al. (2018) data for Carlos: CEAB=98.83%. The upper left panel is a time series plot. The upper right panel represents the A values on the left Y-axis, connected with a dashed line to the corresponding B-values on the right Y-axis. The horizontal red line represents the grand mean of the outcomes, whereas the thick green line connects the mean of the A-values to the mean of the B-values. The lower left panel represents the proportions of variability attributed to treatment (yellow area – not shown as it is practically zero), blocking effect (red area) and interaction/residual (green area): the percentage to the left is the residual/interaction variability in relation to block plus residual variability, whereas the percentage to the right is the residual/interaction variability in relation to the total variability (including the treatment effect). The lower right panel represents the difference between the A and B measurement in each block, as compared to the average mean difference (horizontal red line).

CONSISTENCY IN ALTERNATION DESIGNS

The data for Zoe (Klingbeil et al., 2019) are represented in Figure 11. There is a differentiation between the two conditions, which is practically the same throughout the whole series. In that sense, the slopes of the trend lines are very similar (i.e., the trend lines are practically parallel, as per the upper right panel of Figure 11). This indicates similar efficiency of the two interventions. This can also be understood as consistency of the effect across blocks (CEAB=98.76%), but this is usually not the desired result in an AATD. Due to the presence of similar trends, the variability attributed to blocking is high (86%).

CONSISTENCY IN ALTERNATION DESIGNS

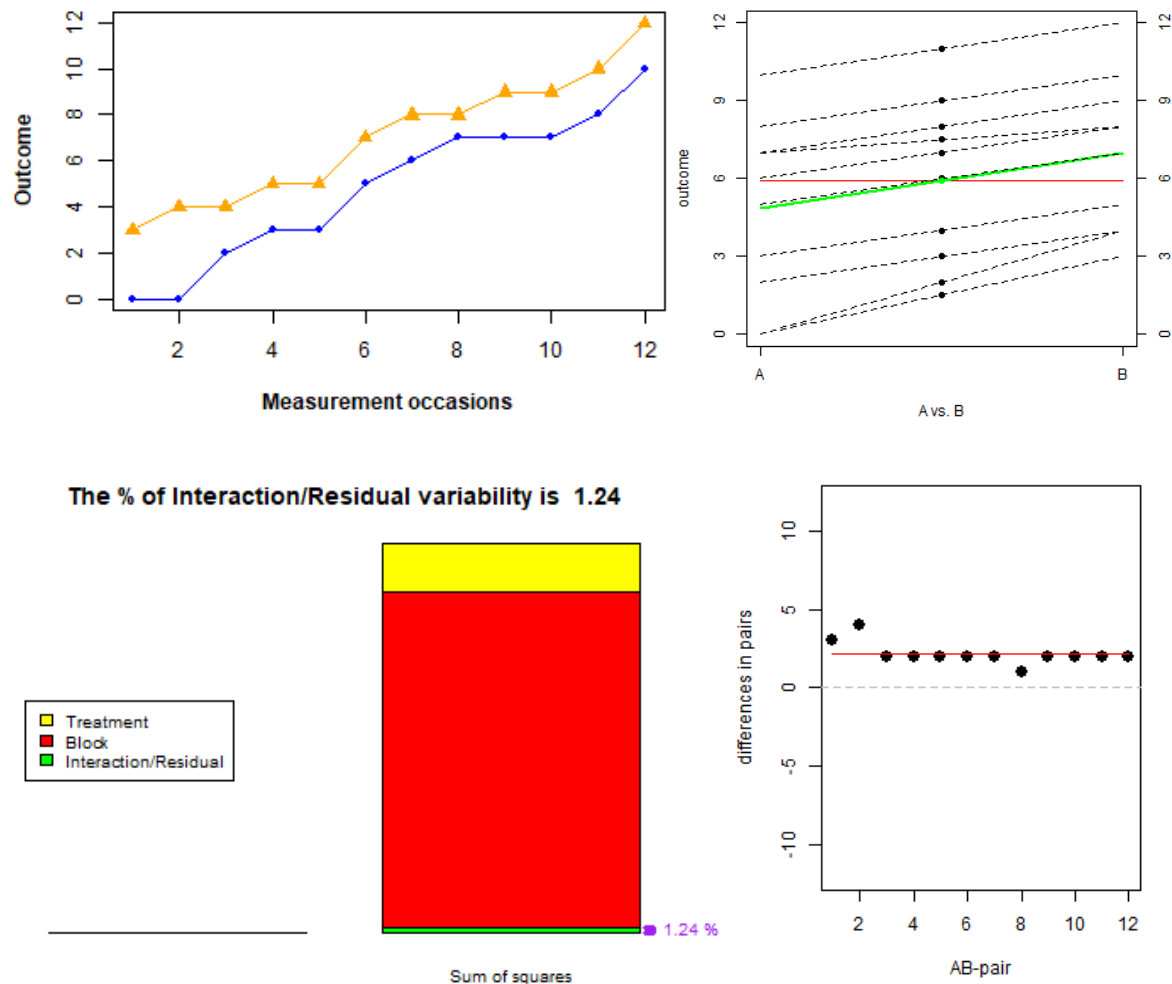


Figure 11. Klingbeil et al. (2018) data for Zoe: CEAB=98.76%. The upper left panel is a time series plot. The upper right panel represents the A values on the left Y-axis, connected with a dashed line to the corresponding B-values on the right Y-axis. The horizontal red line represents the grand mean of the outcomes, whereas the thick green line connects the mean of the A-values to the mean of the B-values. The lower left panel represents the proportions of variability attributed to treatment (yellow area), blocking effect (red area) and interaction/residual (green area): the percentage to the left is the residual/interaction variability in relation to block plus residual variability, whereas the percentage to the right is the residual/interaction variability in relation to the total variability (including the treatment effect). The lower right panel represents the difference between the A and B measurement in each block, as compared to the average mean difference (horizontal red line).

CONSISTENCY IN ALTERNATION DESIGNS

The data for Daniela (Klingbeil et al., 2019) are represented on Figure 12. In this case, there is not only differentiation, but also difference in efficiency. In other words, the difference between the conditions becomes larger as time passes (see the lower right panel of Figure 12). This can be interpreted as smaller consistency of effects across blocks as compared to the Zoe (here CEAB=85.63%), but it is also indicative that one of the interventions leads to achieving the final goal faster.

CONSISTENCY IN ALTERNATION DESIGNS

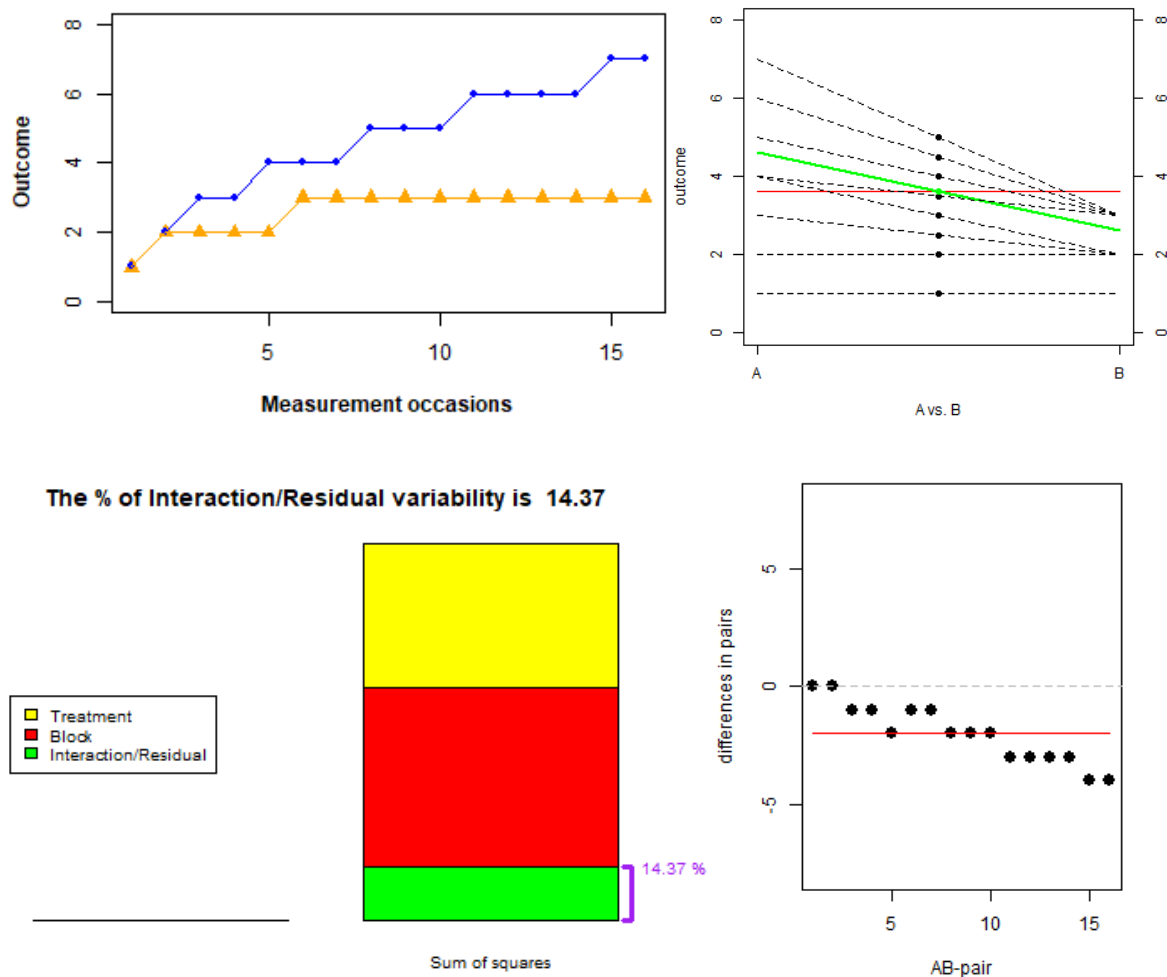


Figure 12. Klingbeil et al. (2018) data for Daniela: CEAB=85.63%. The upper left panel is a time series plot. The upper right panel represents the A values on the left Y-axis, connected with a dashed line to the corresponding B-values on the right Y-axis. The horizontal red line represents the grand mean of the outcomes, whereas the thick green line connects the mean of the A-values to the mean of the B-values. The lower left panel represents the proportions of variability attributed to treatment (yellow area), blocking effect (red area) and interaction/residual (green area): the percentage to the left is the residual/interaction variability in relation to block plus residual variability, whereas the percentage to the right is the residual/interaction variability in relation to the total variability (including the treatment effect). The lower right panel represents the difference between the A and B measurement in each block, as compared to the average mean difference (horizontal red line).

Discussion

Contributions

In the current text, we first propose a quantification of the consistency of effects across blocks for an alternation design with block randomization. This quantification, CEAB, is based on a solid statistical model such as the analysis of variance. It should be noted that we are not recommending here the use of analysis of variance as a primary method for evaluating intervention effectiveness (e.g., Gentile, Roden, & Klein, 1972). In contrast, we only use the variance partitioning performed by the analysis of variance, with no reference to statistical significance, which is likely to be affected by serial dependence (Toothaker, Banz, Noble, Camp, & Davis, 1983). Thus, only descriptive, but not inferential information is used.

For obtaining easily the results of the variance partitioning and the quantifications of consistency, a web-based application was created (<https://manolov.shinyapps.io/ConsistencyRBD/>). This application also provides several graphical representations: (a) a time series line graph; (b) a plot superimposing the pairs of measurements obtained in the different blocks (as the upper right panels presented throughout the current text); (c) a representation of the proportion of variability explained by the intervention, by the blocks and the residual / interaction variability (see the lower left panels); and (d) a representation of differences between conditions for each block, represented in a time sequence and compared to the mean difference between conditions (see the lower right panels).

Implications for Applied Researchers

One approach to the analysis of data obtained from alternation designs is visual inspection (Wolery et al., 2018). When a quantitative analysis is actually performed complementing the

CONSISTENCY IN ALTERNATION DESIGNS

visual inspection, the data are typically analyzed by reporting means and ranges per condition (Manolov & Onghena, 2018). However, relying on averages is not sufficient, as they may hide relevant variability (Normand, 2016). A nonzero effect, on average considering the tiers in an MBD, the phases in an ABAB design or the alternations of the conditions in an alternation design does not necessarily entail consistency. Actually, a lack of consistency could be indicative of an excess of uncontrolled sources of variation, which would suggest that the underlying mechanism of the intervention (or the variables controlling the behavior of interest) is not sufficiently understood.

Before an intervention can be recommended for certain situations (problematic behaviors and personal characteristics), there should be some information available regarding the expected direction and magnitude of the effect of this intervention. Otherwise, erroneous conclusions about treatment efficacy in applied research can have severe consequences. This may for example result in administering ineffective treatments to patients or misallocation of scarce financial resources. Given the importance of assessing consistency (Kratochwill et al., 2010; Lane et al., 2017; Ledford, 2018), the current text, with its focus on alternation designs, fills a relevant gap in the literature, complementing the previous work (Tanious, De, et al., 2019a; Tanious, Manolov, et al., 2019).

Thanks to the web-application developed, the quantifications proposed can be easily complemented with several visual representations of the data. These visual representations enable a better interpretation of the numerical values, because some of them directly represent the degree of consistency of the effect across blocks (i.e., the degree to which the dashed lines cross in upper right panels of the Figures included here), as well as the degree of consistency of the average level across blocks (i.e., the distances between the dashed lines on the Y-axis in

CONSISTENCY IN ALTERNATION DESIGNS

these upper right panels). For instance, the upper right panel of Figure 6 shows that the average level across blocks is the same, but the effect is not completely consistent across blocks. In contrast, the upper right panel of Figure 2 shows that the effect is perfectly consistent across blocks, but the average level is very different. Furthermore, another graphical representation (i.e., the lower right panels of the Figures included here) represents the size of the difference between conditions, preserving their temporal order. This is very important, given that the measurement time or the order of the blocks is not taken into consideration in the ANOVA partitioning of the variance. For instance, the lower right panel of Figure 12 illustrates how the effect is getting larger for later blocks and measurement occasions, whereas the lower right panel of Figure 7 shows that the average effect is zero (and thus consistency of effect need not be assessed) and positive and negative differences are alternated in time.

Our recommendation is to interpret the numerical results (e.g., the CEAB value) alongside a visual analysis of the raw time series graph just as it has been recommended when assessing the magnitude of effect (Fisher, Kelley, & Lomas, 2003; Harrington & Velicer, 2015). Visual inspection is necessary in order to know whether the lack of consistency in the effect is due to excessive unexplained variability or due to trends with different slopes (i.e., the conditions becoming more dissimilar with time). The former case (e.g., Figure 8) is indicative of an insufficient experimental control and would indicate a problematic lack of consistency. In contrast, the latter data pattern (e.g., Figure 12) could be desirable if the effect of the difference between conditions is expected to become more pronounced with time. In this latter case, a certain degree of lack of consistency across blocks can be expected and not be considered detrimental. In that sense, we consider that the quantitative results (such as CEAB) should be

CONSISTENCY IN ALTERNATION DESIGNS

interpreted always in relation to the expected data pattern, considering whether an ATD or an AATD is used.

A kind of unexplained variability that can introduce lack of consistency is the presence of outliers (see Figure 9). An outlier can be expected to reduce the consistency of effects across blocks. For the data depicted in Figure 9, $CEAB=58.25\%$. If the value for the second measurement occasion (a high outlier in the condition marked with a filled triangle) is set to be equal to the value for the fourth measurement occasion, belonging to the same condition but not as outlying, $CEAB$ would be 63.04% . Similarly, if the value for 18th measurement occasion (a high outlier in the condition marked with a filled circle) is set to be equal to the value for the 16th measurement occasion, belonging to the same condition but not outlying, $CEAB$ would be 66.44% . If both outliers are replaced by their not so extreme “neighbors” from the same condition, then $CEAB$ would be 78.32% . Visual analysis can help identifying whether an outlier is the likely cause of a relatively low value of $CEAB$.

Visual inspection can also be useful when the variability attributed to the intervention is zero (i.e., there is no main effect of the intervention, on average). In such a case, interpreting $CEAB$ will not make sense, but a visual inspection of the data can be useful for determining whether the lack of average effect is due to (a) the two conditions giving identical scores, (b) rapid alternation in the superiority of one condition over another (similar to the data represented on Figure 7), or (c) one condition being superior in the beginning and the other in the end of the time series (similar to the data represented on Figure 10).

Finally, the interpretation of the results in terms of a causal relation between the intervention and the target behavior can be aided by the introduction of randomization in the design, for instance, block randomization in an alternation design (Ledford, 2018). Apart from boosting

CONSISTENCY IN ALTERNATION DESIGNS

internal validity, the use of randomization makes possible the application of a randomization test as quantification of the degree to which the effect size observed can be expected by chance (Levin, Kratochwill, & Ferron, 2019; Onghena & Edgington, 1994).

Limitations and Future Research Directions

The current text presents and illustrates a proposal for quantifying consistency of effects across blocks. Our aim was to offer a didactical demonstration that is easy to follow, on the basis of several specific examples of different data patterns and different degrees of consistency.

Nonetheless, a simulation study would still be useful for providing evidence on the performance of the proposal made. Specifically, generated data could be used to explore how different data patterns (e.g., including linear and nonlinear trends in similar or different directions, outliers in one or the two conditions) are reflected in the quantification of consistency of effects across blocks (proposed here) or in the quantification of the consistency of superiority such as the Percentage of nonoverlapping data (Wolery et al., 2014). A simulation study would be especially relevant in case another measure of consistency is proposed for alternation designs with block randomization, in order to compare the performance and informative value of the quantifications.

A second relevant line of future research would be to perform a field test, for establishing interpretative benchmarks for CEAB. Specifically, such a field test can follow the approach for obtaining benchmarks for CONDAP (Tanious, De, et al., 2019b). Finally, it would be important to continue developing measures of consistency, specifically for alternating treatments designs with restricted randomization (Onghena & Edgington, 1994). Such designs are challenging for two reasons. First, the variance partitioning cannot be obtained on the basis of blocks (as such are absent) and, therefore, a different kind of measure of consistency is called for. Second, these designs can lead to unequal number of measurements per condition in certain alternating

CONSISTENCY IN ALTERNATION DESIGNS

sequences (e.g., Eilers & Hayes, 2015; Maitland & Gaynor, 2016), which makes less straightforward even the assessment of consistency in superiority using the Percentage of nonoverlapping data (Wolery et al., 2014).

CONSISTENCY IN ALTERNATION DESIGNS

References

- Blampied, N. M. (2017). Analyzing therapeutic change using modified Brinley plots: History, construction, and interpretation. *Behavior Therapy, 48*(1), 115–127.
<https://doi.org/10.1016/j.beth.2016.09.002>
- Byiers, B. J., Reichle, J., & Symons, F. J. (2012). Single-subject experimental design for evidence-based practice. *American Journal of Speech-Language Pathology, 21*(4), 397–414.
[https://doi.org/10.1044/1058-0360\(2012/11-0036\)](https://doi.org/10.1044/1058-0360(2012/11-0036))
- 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*(4), 387–400.
<https://doi.org/10.1177/002246698501900404>
- Cihak, D., Alberto, P. A., Taber-Doughty, T., & Gama, R. I. (2006). A comparison of static picture prompting and video prompting simulation strategies using group instructional procedures. *Focus on Autism and Other Developmental Disabilities, 21*(2), 89–99.
<https://doi.org/10.1177/10883576060210020601>
- Coleman, M. B., Cherry, R. A., Moore, T. C., Park, Y., & Cihak, D. F. (2015). Teaching sight words to elementary students with intellectual disability and autism: A comparison of teacher-directed versus computer assisted simultaneous prompting. *Intellectual and Developmental Disabilities, 53*(3), 196–210. <https://doi.org/10.1352/1934-9556-53.3.196>
- Edgington, E. S. (1996). Randomized single-subject experimental designs. *Behaviour Research and Therapy, 34*(7), 567–574. [https://doi.org/10.1016/0005-7967\(96\)00012-5](https://doi.org/10.1016/0005-7967(96)00012-5)
- Eilers, H. J., & Hayes, S. C. (2015). Exposure and response prevention therapy with cognitive defusion exercises to reduce repetitive and restrictive behaviors displayed by children with

CONSISTENCY IN ALTERNATION DESIGNS

autism spectrum disorder. *Research in Autism Spectrum Disorders*, 19(Nov), 18–31.

<https://doi.org/10.1016/j.rasd.2014.12.014>

Fahmie, T. A., & Hanley, G. P. (2008). Progressing toward data intimacy: A review of within-session data analysis. *Journal of Applied Behavior Analysis*, 41(3), 319–331.

<https://doi.org/10.1901/jaba.2008.41-319>

Fisher, W. W., Kelley, M. E., & Lomas, J. E. (2003). Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. *Journal of Applied Behavior Analysis*, 36(3), 387–406. <https://doi.org/10.1901/jaba.2003.36-387>

Ganz, J. B., & Ayres, K. M. (2018). Methodological standards in single-case experimental design: Raising the bar. *Research in Developmental Disabilities*, 79(1), 3–9.

<https://doi.org/10.1016/j.ridd.2018.03.003>

Geist, K., & Hitchcock, J. H. (2014). Single case design studies in music therapy: Resurrecting experimental evidence in small group and individual music therapy clinical settings. *Journal of Music Therapy*, 51(4), 293–309. <https://doi.org/10.1093/jmt/thu032>

Gentile, J. R., Roden, A. H., & Klein, R. D. (1972). An analysis-of-variance model for the intrasubject replication design. *Journal of Applied Behavior Analysis*, 5(2), 193–198.

<https://doi.org/10.1901/jaba.1972.5-193>

Hammond, D., & Gast, D. L. (2010). Descriptive analysis of single subject research designs: 1983-2007. *Education and Training in Autism and Developmental Disabilities*, 45(2), 187–202. <https://www.jstor.org/stable/23879806>

Harrington, M., & Velicer, W. F. (2015). Comparing visual and statistical analysis in single-case studies using published studies. *Multivariate Behavioral Research*, 50(2), 162–183.

<https://doi.org/10.1080/00273171.2014.973989>

CONSISTENCY IN ALTERNATION DESIGNS

Hays, W. L. (1994). *Statistics* (5th ed.). Orlando, FL: Harcourt Brace College Publishers.

Heyvaert, M., Wendt, O., Van den Noortgate, W., & Onghena, P. (2015). Randomization and data-analysis items in quality standards for single-case experimental studies. *Journal of Special Education, 49*(3), 146–156. <https://doi.org/10.1177/0022466914525239>

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165–179. <https://doi.org/10.1177/001440290507100203>

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

Kirk, R. E. (2013). *Experimental design: Procedures for the behavioral sciences* (4th ed.). Thousand Oaks, CA: Sage.

Klingbeil, D. A., January, S. A. A., & Ardoin, S. P. (2019, May 25). Comparative efficacy and generalization of two word-reading interventions with English learners in elementary school. *Journal of Behavioral Education*. Advance online publication. <https://doi.org/10.1007/s10864-019-09331-y>

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 What Works Clearinghouse website: https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_scd.pdf

Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation, 24*(3-4), 445–463. <https://doi.org/10.1080/09602011.2013.815636>

CONSISTENCY IN ALTERNATION DESIGNS

- Lane, J. D., Ledford, J. R., & Gast, D. L. (2017). Single-case experimental design: current standards and applications in occupational therapy. *American Journal of Occupational Therapy, 71*(2), 7102300010p1–7102300010p9. <https://doi.org/10.5014/ajot.2017.022210>
- Lane, J. D., Shepley, C., & Spriggs, A. D. (2019, September 27). Issues and improvements in the visual analysis of A-Bb single-case graphs by pre-service professionals. *Remedial and Special Education*. Advance online publication. <https://doi.org/10.1177/0741932519873120>
- Lanovaz, M., Cardinal, P., & Francis, M. (2019). Using a visual structured criterion for the analysis of alternating-treatment designs. *Behavior Modification, 43*(1), 115–131. <https://doi.org/10.1177/0145445517739278>
- Ledford, J. R. (2018). No randomization? No problem: Experimental control and random assignment in single case research. *American Journal of Evaluation, 39*(1), 71–90. <https://doi.org/10.1177/1098214017723110>
- Ledford, J. R., Barton, E. E., Severini, K. E., & Zimmerman, K. N. (2019). A primer on single-case research designs: Contemporary use and analysis. *American Journal on Intellectual and Developmental Disabilities, 124*(1), 35–56. <https://doi.org/10.1352/1944-7558-124.1.35>
- Levin, J. R., Kratochwill, T. R., & Ferron, J. M. (2019). Randomization procedures in single-case intervention research contexts: (Some of)“the rest of the story”. *Journal of the Experimental Analysis of Behavior, 112*(3), 334-348. <https://doi.org/10.1002/jeab.558>
- Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766. <https://doi.org/10.1016/j.jesp.2013.03.013>
- Lloyd, B. P., Finley, C. I., & Weaver, E. S. (2018). Experimental analysis of stereotypy with applications of nonparametric statistical tests for alternating treatments designs.

CONSISTENCY IN ALTERNATION DESIGNS

Developmental Neurorehabilitation, 21(4), 212–222.

<https://doi.org/10.3109/17518423.2015.1091043>

Maggin, D. M., Briesch, A. M., Chafouleas, S. M., Ferguson, T. D., & Clark, C. (2014). A comparison of rubrics for identifying empirically supported practices with single-case research. *Journal of Behavioral Education*, 23(2), 287–311. <https://doi.org/10.1007/s10864-013-9187-z>

Maggin, D. M., Briesch, A. M., & Chafouleas, S. M. (2013). An application of the What Works Clearinghouse standards for evaluating single-subject research: Synthesis of the self-management literature base. *Remedial and Special Education*, 34(1), 44–58. <https://doi.org/10.1177/0741932511435176>

Maggin, D. M., Cook, B. G., & Cook, L. (2018). Using single-case research designs to examine the effects of interventions in special education. *Learning Disabilities Research & Practice*, 33(4), 182–191. <https://doi.org/10.1111/ldrp.12184>

Manolov, R. (2019). A simulation study on two analytical techniques for alternating treatments designs. *Behavior Modification*, 43(4), 544–563. <https://doi.org/10.1177/0145445518777875>

Manolov, R., & Onghena, P. (2018). Analyzing data from single-case alternating treatments designs. *Psychological Methods*, 23(3), 480–504. <https://doi.org/10.1037/met0000133>

Maitland, D. W. M., & Gaynor, S. T. (2016). Functional analytic psychotherapy compared with supportive listening: An alternating treatments design examining distinctiveness, session evaluations, and interpersonal functioning. *Behavior Analysis: Research and Practice*, 16(2), 52–64. <https://doi.org/10.1037/bar0000037>

CONSISTENCY IN ALTERNATION DESIGNS

Mengersen, K., McGree, J. M., & Schmid, C. H. (2015). Statistical analysis of N-of-1 trials. In J. Nikles & G. Mitchell (Eds.), *The essential guide to N-of-1 trials in health* (pp.135-153).

Dordrecht: Springer.

Michiels, B., & Onghena, P. (2019). Randomized single-case AB phase designs: Prospects and pitfalls. *Behavior Research Methods*, *51*(6), 2454–2476. <https://doi.org/10.3758/s13428-018-1084-x>

Miller, M. J. (1985). Analyzing client change graphically. *Journal of Counseling and Development*, *63*(8), 491–494. <http://dx.doi.org/10.1002/j.1556-6676.1985.tb02743.x>

Natesan, P., & Hedges, L. V. (2017). Bayesian unknown change-point models to investigate immediacy in single case designs. *Psychological Methods*, *22*(4), 743–759.

<https://doi.org/10.1037/met0000134>

Nikles, J. & Mitchell, G. (Eds.) (2015). *The essential guide to N-of-1 trials in health*. Dordrecht: Springer.

Normand, M. P. (2016). Less is more: Psychologists can learn more by studying fewer people.

Frontiers in Psychology, *7*, e934. <https://doi.org/10.3389/fpsyg.2016.00934>

Olive, M. L., & Smith, B. W. (2005). Effect size calculations and single subject designs.

Educational Psychology, *25*(2-3), 313–324. <https://doi.org/10.1080/0144341042000301238>

Onghena, P., & Edgington, E. S. (1994). Randomization tests for restricted alternating treatments designs. *Behaviour Research and Therapy*, *32*(7), 783–786. [https://doi.org/10.1016/0005-](https://doi.org/10.1016/0005-7967(94)90036-1)

[7967\(94\)90036-1](https://doi.org/10.1016/0005-7967(94)90036-1)

Onghena, P., & Edgington, E. S. (2005). Customization of pain treatments: Single-case design and analysis. *Clinical Journal of Pain*, *21*(1), 56–68. [https://doi.org/10.1097/00002508-](https://doi.org/10.1097/00002508-200501000-00007)

[200501000-00007](https://doi.org/10.1097/00002508-200501000-00007)

CONSISTENCY IN ALTERNATION DESIGNS

Parker, R. I., Cryer, J., & Byrns, G. (2006). Controlling baseline trend in single-case research.

School Psychology Quarterly, 21(4), 418–443. <https://doi.org/10.1037/h0084131>

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(2), 284–299.

<https://doi.org/10.1016/j.beth.2010.08.006>

Parsonson, B. S., & Baer, D. M. (1978). The analysis and presentation of graphic data. In T. R. Kratochwill (Ed.), *Single-subject research: Strategies for evaluating change* (pp. 101–165). New York: Academic Press.

Petursdottir, A. I., & Carr, J. E. (2018). Applying the taxonomy of validity threats from mainstream research design to single-case experiments in applied behavior analysis. *Behavior Analysis in Practice*, 11(3), 228–240. <https://doi.org/10.1007/s40617-018-00294-6>

Reichow, B., Barton, E. E., & Maggin, D. M. (2018). Development and applications of the single-case design risk of bias tool for evaluating single-case design research study reports. *Research in Developmental Disabilities*, 79(1), 53–64.

<https://doi.org/10.1016/j.ridd.2018.05.008>

Savaiano, M. E., Compton, D. L., Hatton, D. D., & Lloyd, B. P. (2016). Vocabulary word instruction for students who read braille. *Exceptional Children*, 82(3), 337–353.

<https://doi.org/10.1177/0014402915598774>

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(2), 123–147. <https://doi.org/10.1016/j.jsp.2013.11.005>

CONSISTENCY IN ALTERNATION DESIGNS

Shadish, W. R., & Sullivan, K. J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods*, 43(4), 971–980.

<https://doi.org/10.3758/s13428-011-0111-y>

Shepley, C., Ault, M. J., Ortiz, K., Vogler, J. C., & McGee, M. (2019, February 22). An exploratory analysis of quality indicators in adapted alternating treatments designs. *Topics in Early Childhood Special Education*. Advance online publication.

<https://doi.org/10.1177/0271121418820429>

Sidman, M. (1960). *Tactics of scientific research*. New York, NY: Basic Books.

Smith, J. D. (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods*, 17(4), 510–550.

<https://doi.org/10.1037/a0029312>

Solanas, A., Manolov, R., & Onghena, P. (2010). Estimating slope and level change in N=1 designs. *Behavior Modification*, 34(3), 195–218. <https://doi.org/10.1177/0145445510363306>

Swaminathan, H., Rogers, H. J., Horner, R., Sugai, G., & Smolkowski, K. (2014). Regression models for the analysis of single case designs. *Neuropsychological Rehabilitation*, 24(3-4), 554–571. <https://doi.org/10.1080/09602011.2014.887586>

Tanious, R., De, T. K., Michiels, B., Van den Noortgate, W., & Onghena, P. (2019a). Assessing consistency in single-case A-B-A-B phase designs. *Behavior Modification*. Advance online publication. <https://doi.org/10.1177/0145445519837726>

Tanious, R., De, T. K., Michiels, B., Van den Noortgate, W., & Onghena, P. (2019b).

Consistency in single-case ABAB phase designs: A systematic review. *Behavior*

Modification. Advance online publication. <https://doi.org/10.1177/0145445519853793>

CONSISTENCY IN ALTERNATION DESIGNS

- Tanious, R., Manolov, R., & Onghena, P. (2019). The assessment of consistency in single-case experiments: Beyond A-B-A-B designs. *Behavior Modification*. Advance online publication. <https://doi.org/10.1177/0145445519882889>
- Tate, R. L., Perdices, M., Rosenkoetter, U., McDonald, S., Togher, L., ..., Vohra, S. (2016). The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016: Explanation and elaboration. *Archives of Scientific Psychology*, 4(1), 10–31. <https://doi.org/10.1037/arc0000027>
- Toothaker, L. E., Banz, M., Noble, C., Camp, J., & Davis, D. (1983). N = 1 designs: The failure of ANOVA-based tests. *Journal of Educational Statistics*, 8(4), 289–309. <https://doi.org/10.3102/10769986008004289>
- Tukey, J. W. (1977). *Exploratory data analysis*. London, UK: Addison-Wesley.
- Warren, T., Cagliani, R. R., Whiteside, E., & Ayres, K. M. (2019, November 13). Effect of task sequence and preference on on-task behavior. *Journal of Behavioral Education*. Advance online publication. <https://doi.org/10.1007/s10864-019-09358-1>
- Wolery, M., Gast, D. L., & Ledford, J. R. (2014). Comparison designs. In D. L. Gast & J. R. Ledford (Eds.), *Single case research methodology: Applications in special education and behavioral sciences* (2ⁿd ed.) (pp. 297–345). London, UK: Routledge.
- Wolery, M., Gast, D. L., & Ledford, J. R. (2018). Comparative designs. In D. L. Gast & J. R. Ledford (Eds.), *Single case research methodology: Applications in special education and behavioral sciences* (3rd ed.) (pp. 283–334). London, UK: Routledge.
- Wolfe, K., Barton, E. E., & Meadan, H. (2019). Systematic protocols for the visual analysis of single-case research data. *Behavior Analysis in Practice*, 12(2), 491–502. <https://doi.org/10.1007/s40617-019-00336-7>

CONSISTENCY IN ALTERNATION DESIGNS

Zimmerman, K. N., Ledford, J. R., Severini, K. E., Pustejovsky, J. E., Barton, E. E., & Lloyd, B.

P. (2018). Single-case synthesis tools I: Comparing tools to evaluate SCD quality and rigor.

Research in Developmental Disabilities, 79(Aug), 19–32.

<https://doi.org/10.1016/j.ridd.2018.02.003>