Applied hybrid single-case experiments published between 2016 and 2020: A systematic review

René Tanious and Patrick Onghena

Abstract
Single-case experimental designs (SCEDs) are frequently used research designs in psychology, (special) education, and related fields. Hybrid designs are formed by combining two or more of the basic SCED forms (i.e. phase designs, alternation designs, multiple baseline designs, and changing criterion designs). Hybrid designs have the potential to tackle complex research questions and increase internal validity, but relatively little is known about their use in actual research practice. Therefore, we systematically reviewed SCED hybrid designs published between 2016 and 2020. The systematic review of 67 studies indicates that a hybrid of phase designs and multiple baseline designs is most popular. Hybrid designs are most frequently analyzed by means of visual analysis paired with descriptive statistics. Randomization in the study design is common only for one particular kind of hybrid design. Examples of hybrid studies reveal that these designs are particularly popular in educational research. We compare some of the results of the systematic review to those obtained by Hammond and Gast, Shadish and Sullivan, and Tanious and Onghena. Finally, we discuss the results of the present systematic review in light of the need for specific guidelines for hybrid designs, including analytical methods, design specific randomization and reporting, and the need for terminological clarification.

Keywords
Single-case experimental designs, single-case hybrid designs, single-case embedded designs, systematic review

Introduction
Single-case experimental designs (SCEDs) are a class of experimental designs used to assess intervention effects at an individual level. One case (e.g. a student or a classroom) is observed repeatedly over time under different conditions of at least one independent variable (e.g. teacher praise) so that the entity serves as its own control for comparison purposes (Barlow et al., 2009). SCEDs are broadly applicable to many research questions and can be particularly valuable when resources are scarce, when dealing with participants with low incidence disabilities, and when assignment of participants to a no treatment group is ethically unfeasible (McDonnell and O’Neill, 2003). It has been found that in fields such as disability and education, SCEDs are not only common, but even more frequently used than other forms of experimentation (Shadish and Sullivan, 2011). SCEDs can produce insights about treatment effectiveness on an individual level unmatched by group designs. In addition, if desired, SCEDs can be used to assess the feasibility of an intervention before implementing it in large-scale group studies.

The basic characteristics of SCEDs as outlined above can be implemented in various design forms. In the following paragraphs, we will first give an overview of the available SCED forms. We will then outline how these forms can be combined to form hybrid designs. To date, little is known about the use of these hybrid designs in applied research, both from a theoretical (e.g. which specific design combination) and practical (e.g. typical number of participants and measurements) point of view. Knowledge about these characteristics may help
A classification of SCEDs

The basic procedure of repeatedly observing a single entity under different conditions of at least one independent variable can be implemented in many different design forms. Typically, there are five overarching categories that each have subcategories on their basic form (cf. classification by Onghena and Edgington, 2005): Phase, alternation, multiple baseline, changing criterion, and hybrid. In this section, we will shortly discuss the first four basic categories, and also add the possibilities for incorporating randomization in the design as an important feature for strengthening the internal validity (Tate et al., 2013). The next sections will then go into greater detail about how these basic categories can be combined to form hybrid designs.

Phase designs. The first main category of SCEDs is phase designs. As the name suggests, phase designs are marked by distinct experimental phases. The most basic variant within this category is the AB design, where A stands for measurements taken within the baseline condition and B stands for measurements taken within the experimental condition. This basic variant can be replicated within cases several times (e.g. ABAB, ABABAB), replicated across cases, or include more than two distinct conditions (e.g. ABAC, A-B-C-BC-A). In educational settings, phase designs including a withdrawal or reversal (e.g. ABA and ABAB) can be of particular interest when it is ethically feasible to withdraw treatment for a certain period of time to investigate whether the dependent variable (e.g. disruptive behavior) reverts back to pre-intervention levels. An example may be the use of praise to reduce disruptive behavior. Praise may be introduced as the intervention to observe if disruptive behavior decreases. Subsequently, the intervention may be withdrawn to assess whether disruptive behavior reverses to baseline levels. Randomization can be included in phase designs by randomly determining the moments of phase change between adjacent conditions (Onghena, 1992).

Alternation designs. In alternation designs, condition changes occur rapidly. In the completely randomized version, each measurement time has an equal chance of being assigned to one of the available treatments (e.g. A or B when two treatments or a treatment and control condition are compared) (Edgington, 1980a). A modification to this basic version is to place limits on how often each condition can occur consecutively to avoid administering the same treatment too many times in a row (Onghena and Edgington, 1994). Finally, it is also possible to alternate the treatments in blocks. For example, if two treatments are compared, the treatments could be presented in blocks of two and the order of treatment presentation is then randomized within each block (Edgington, 1980b).

Multiple baseline designs. Similar to phase designs, multiple baseline designs consist of distinct experimental phases with A–B comparisons. The introduction of the intervention is temporally staggered to assess whether the intervention affects only the cases, behaviors, or settings that are in the B-phase while the other cases, behaviors, or settings remain in baseline (Baer et al., 1968). There are several ways in which randomization can be included in multiple baseline designs. An overview of available randomization techniques for the multiple baseline design is available in Kratochwill and Levin (2010). The first way is identical to phase designs where the moment of phase change is randomly determined for each case, behavior, or setting (Marascuilo and Busk, 1988; Onghena, 1992). The second way to include randomization is to determine different baseline phase lengths in advance and assign the cases, behaviors, or settings randomly to the different baselines (Wampold and Worsham, 1986). Multiple baseline designs can be used when it is of interest to compare the effect of an intervention across different units (e.g. classrooms), behaviors (e.g. reading and algebra), or settings (e.g. digital and on-campus). For example, if the goal of the study is to assess the effectiveness of a new curriculum in learning a second language, the effectiveness of the curriculum may be compared across different classrooms with a staggered introduction of the new curriculum across classrooms.

Changing criterion designs. Changing criterion designs are another class of SCEDs that consist of different phases. After an initial baseline phase the intervention is introduced and usually remains in place for the remainder of the experiment. The intervention is presented in subphases during which the participant has to meet pre-specified criteria which change between adjacent phases. Each criterion can either be a single value (Hartmann and Hall, 1976) or a range defined by an acceptable upper and lower limit (McDougall, 2005). Randomization in changing criterion designs can be incorporated in the same way as for phase designs (i.e. randomly determining the moment of criterion change between adjacent phases) (Ferron et al., 2019; Onghena et al., 2019). Changing criterion designs are recommended when it is ethically or practically unfeasible to withdraw the intervention and/or when changes in behavior are expected to occur gradually. For example, if the goal is to gradually increase the number of bites of non-preferred food items for a child with autism...
spectrum disorder through rewards (e.g., access to toys), the number of bites accepted may be gradually increased to achieve access to rewards for the child.

A note on terminology

As we noted elsewhere (Tanious and Onghena, 2021), SCED hybrid designs—like other forms of SCEDs—go by many names. Ledford and Gast (2018) refer to these designs as “combination designs” and Moeyaert et al. (2020) call them “combined designs,” highlighting the fact that two or more of the basic SCED forms are combined in a single study. Pustejovsky and Ferron (2017) as well as Swan et al. (2020) use the term hybrid designs, highlighting that neither of the basic SCED forms stay in its pure form, creating a hybrid version of two designs. These different terms are interchangeable as they all refer to the same class of SCED designs. However, in this manuscript we follow the terminology used in Pustejovsky and Ferron as well as Swan et al., because we believe the term “hybrid” more accurately captures the peculiarities of this design category.

SCED hybrid designs: Prevalence, definition, and rationale

In a recent systematic review of 423 applied SCED studies published between 2016 and 2018 hybrid designs accounted for roughly 9% (N=38) of all included studies (Tanious and Onghena, 2021). However, little is known about the methodological characteristics of applied hybrid designs because recent systematic reviews are lacking. In addition, recent guidelines on conduct and analysis of SCEDs (What Works Clearinghouse, 2020a, 2020b) do not include hybrid designs as a separate category and therefore offer little guidance to applied researchers wishing to conduct SCED hybrid studies.

However, this does not mean that hybrid designs have never been formally reviewed as part of broader systematic reviews. In their systematic review of applied SCED research published between 1983 and 2007, Hammond and Gast (2010) offer the following definition of hybrid designs: “Researchers could also combine two or more single subject designs to evaluate a functional relation between the independent and dependent variable [. . . ]” (p. 190). Ledford and Gast (2018) assert that hybrid designs typically contain one dominant or primary design that will lead decision making processes, with the other design(s) being incorporated in the light of perceived shortcomings of the primarily chosen design. Hammond and Gast found that the frequency of published hybrid designs was relatively low overall, but increased over the years. In total, they found 36 hybrid designs published over 25 years, with the most frequent combination of designs being multiple baseline designs combined with alternation designs, and multiple baseline designs combined with phase designs. In their systematic review of applied SCED research published in 2008, Shadish and Sullivan (2011) found that over a quarter of all studies used hybrid designs with the most frequent being a combination of multiple baseline designs and phase designs and multiple baseline and alternation designs. An example of a hybrid design consisting of a multiple baseline design and ABAB phase design is the study by Radley et al. (2016), investigating the effectiveness of the Quit Classroom Game for increasing academically engaged behavior across three first-grade classrooms. Regarding the reasons for choosing this particular hybrid design, the authors state that it allows for “sequential application, comparison of intervention efficacy, and replication of intervention effects both within and across participants” (p. 100). In other words the ABAB element allows for comparisons of intervention effectiveness by withdrawing and reintroducing the intervention and the multiple baseline component allows for assessment of intervention effectiveness between participants. A graphical display of the results can be found in Figure 1.

In general, there may be many reasons for combining two or more of the basic SCED forms to form a hybrid: To answer more than one research question in one experiment, to compensate for weaknesses in a single design option (i.e. increasing experimental control), to respond to certain circumstances in the experiment such as covariation of behaviors (Ledford and Gast, 2018); to obtain more intra- and intersubject replications (Hammond and Gast, 2010); to rule out alternative explanations for behavior change (Plavnick and Ferreri, 2013); and to address complex and more sophisticated research questions (Moeyaert et al., 2020).

Thus, while hybrid designs can be a valuable tool for assessing intervention effectiveness, relatively little is known about their use apart from the specific design combinations, especially in more recent years.

Research questions

By means of a systematic review, the present study therefore attempts to answer the following research questions about hybrid designs:

1. Which designs are combined to form hybrid designs?
2. What is the frequency of use for each hybrid design?
3. How many data aspects are analyzed in different hybrid designs?
4. Which data analytical techniques are used?
5. How common is randomization in hybrid designs?
6. Which forms of randomization are used and what is their prevalence?
7. How common are randomization tests in randomized hybrid SCEDs?
8. What is the typical number of participants and measurements?
9. What terms are used by applied researchers to label hybrid designs?
Method

Search methods

In Tanious and Onghena (2021), we presented a systematic review of applied SCED research published between 2016 and 2018. Thirty-eight hybrid studies were identified as part of that systematic review. Those studies are included in the present systematic review for further in-depth analyses. In addition, we expanded the search to include the years 2019 and 2020 as well as adding new search terms specifically aimed at finding hybrid designs for all included years. Thus, the date range was set to 2016–2020. We consider 5 years a sufficiently long period to collect a large enough sample of studies without losing focus of the recency. All searches were performed using PubMed and Web of Science because these search engines fulfill the necessary quality criteria to use them as primary search engines in systematic reviews (Gusenbauer and Haddaway, 2020). Table 1 offers an overview of the used search terms.

Inclusion and exclusion criteria

The first inclusion criterion was that all studies had to present applied hybrid SCED studies as defined in the introduction. Studies using only a single form of SCEDs (e.g. multiple...
baseline across participants) were therefore excluded. A related criterion was that all studies had to present primary applied SCED research as opposed to purely methodological articles (e.g. simulation studies) and systematic reviews or meta-analyses. Another inclusion criterion concerned the period of publication. To assess the most recent trends in the publication of hybrid designs, while at the same time obtaining a large enough sample size, all studies had to be published between 2016 and 2020. Whenever possible, this was checked with the date of first online publication. The print date was only checked when the date of first only publication was unavailable. Studies that were published online outside the date range were excluded even when the print date fell within the inclusion period to avoid studies being included that were potentially conducted long before the inclusion period. In addition, articles were only considered admissible if they were written in English and available through KU Leuven library without additional payment. The study selection process was verified by the second author for five studies with an interrater agreement of 80%.

**Coding criteria**

**Design combination.** First, we coded the combination of designs according to the four basic categories presented in the introduction to answer the first two research questions. For coding the design, we followed the description given by the authors. If the combination of designs was not clearly named, the figure(s) depicting the results was checked. When coding the combination of designs, it was also checked which design was the “dominant” one according to the description given by the authors. For example, if the authors stated that they had used a multiple baseline design with alternating treatments, then the multiple baseline design would be the dominant design.

**Data aspects.** Based on previous advances in the field, the 2010 What Works Clearinghouse guidelines (Kratochwill et al., 2010) recommended taking into account six data aspects when analyzing SCED data: level, trend, variability, over, immediacy of effect, and consistency of data patterns. Originally, these guidelines were presented in the context of visual analysis. However, they have also had substantial influence on the development of statistical methods for analyzing SCEDs. These data aspects are therefore applicable to all forms of analytical methods as discussed in the next section. The data aspects were coded to answer research question 3. Level is the central tendency in an experimental condition, which can for example be operationalized as the mean or median score. Trend refers to the best fitting straight line which can for example be calculated via ordinary least squares regression or split-middle trend estimation. Variability refers to the dispersion of the scores around the central tendency and is usually expressed in terms of standard deviations. Overlap refers to the proportion of measurements that are overlapping between adjacent phases. Immediacy refers to how rapid changes in the dependent variable occur after changing experimental conditions and the What Works Clearinghouse guideline suggest operationalizing it as the mean difference between the last three baseline phase measurements and the first three intervention phase measurements. Finally, consistency of data patterns refers to how consistent the data patterns from experimentally similar phases are and it has been proposed to measure this with a standardized Euclidean distance measure (Tanious et al., 2020). For studies analyzing several data aspects, the specific combination of data aspects was coded. It was also possible that studies did not analyze any of these data aspects.

**Data analysis.** To answer research question 4, we identified which analytical methods were used to assess intervention effectiveness in hybrid designs. To that end, we distinguished two basic forms of analysis common in SCEDs: visual and statistical analyses (Manolov and Moeyaert, 2017a). Visual analysis refers to the inspection of a graphical representation of the data—usually in the form of a time-series graph depicting the dependent variable(s) under the different experimental conditions to draw conclusions about intervention effectiveness. Visual analysis can be performed with or without the addition of visual clues such as mean or trend lines, according to the data aspects outlined in the previous section. Guidelines for the visual analysis of SCEDs are for example available in Lane and Gast (2014) as well as the previously discussed What Works Clearinghouse guidelines. Statistical analysis involves the calculation of summary measures of the raw data to determine intervention effectiveness. This can be done in a purely descriptive way (e.g. mean

### Table 1. Search terms used in the present study.

| Search term |
|-------------|
| single-case hybrid design |
| single-subject hybrid design |
| single-case combined design |
| single-subject combined design |
| single-case combination design |
| single-subject combination design |
| “multiple baseline design” |
| “reversal design” |
| “withdrawal design” |
| “AB design” |
| “ABAB design” |
| “changing criterion design” |
| “alternating treatments design” |
| “single-case” AND “randomized block design” |
| “single-subject” AND “randomized block design” |
| “single-case” AND “completely randomized design” |
| “single-subject” AND “completely randomized design” |
| “Single-case experimental design” |
| “Single-subject experimental design” |
differences or other effect size measures) or with estimates of some unknown parameter value. In the case of inferential statistical analyses, the degree of uncertainty in the parameter estimation is reflected in p-values and/or confidence intervals. It is of course possible that studies use a combination of these analytical methods, for example, visual analysis with descriptive statistics. In those cases, the specific combination of analytical methods was coded. Finally, it was also possible that studies did not use any analytical methods. Several available statistical analysis techniques for SCEDs, and how they can be used jointly with visual analysis, are discussed in Manolov and Moeyaert (2017a, 2017b).

Randomization. Some of the design specific ways of incorporating randomization have been described in the introduction. Given that hybrid designs consist of two or more of the basic designs, one or several forms of randomization can be incorporated into the design. They were coded in the following manner to answer research question 5 and 6. Randomization can either be restricted (i.e. limiting the maximum number of consecutive measurements of the same conditions) or unrestricted (i.e. each measurement can be taken under any of the experimental conditions). A variation on this is block randomization, where the different conditions are grouped into blocks of measurement occasions. The order of conditions is then randomized within each block with as many blocks as the researcher planned in the design stage. Furthermore, randomization can be incorporated according to the phase change moments, meaning that the moment of phase change between two adjacent conditions is determined randomly. Another variant is that the phase lengths are specified in advance and that the different cases, behaviors, or settings are then randomly assigned to the different phase lengths. Finally, it is also possible that randomization is either not present or has been incorporated into the design, but the authors of the study do not specify the exact manner in which this has been achieved. It should be emphasized that all forms of randomization sketched here are part of the study design itself. Therefore, other forms of randomization that were not part of the study design were coded as “randomization not present.” An example of such a form of randomization is a situation in which the order of treatments was fixed, but the stimuli presentation within experimental conditions was randomized.

Randomization test. Randomization tests are a class of non-parametric significance tests applicable to randomized SCEDs. Randomization tests are valid and powerful significance tests under the assumption that “in experiments in which randomization is performed, the actual arrangement of treatments [...] is one chosen at random from a predetermined set of possible arrangements” (Welch, 1937, p. 47). The obtained test statistic of the experiment is then compared to said other possible arrangements to assess how many other arrangements would have led to a larger treatment effect. The steps involved in conducting a randomized SCED and analyzing the obtained data with a randomization test are explained in Heyvaert and Onghena (2014). The presence of randomization tests was coded dichotomously (i.e. either present or not present) to answer research question 7, following the authors’ description of the analyses performed. The coding of randomization tests is by definition only applicable to the subset of studies that used randomization in the design.

Number of participants and measurement occasions. To answer research question 8, we first checked the method section of each article to find these numbers. If one or both were not specified in the running text, we proceeded to check the x-axis of the results figure(s) for the number of measurements and the number of graphs for the number of participants, with the exception of multiple baseline designs where one graph can include several participants. The number of participants was coded separately for each study when an article presented more than one study. The number of “participants” was calculated for the unit of analysis as defined in the original article. For example, a multiple baseline design across two classrooms would be coded as having two “participants” even if each classroom contained several students. The number of measurement occasions was defined as the number of sessions in which a measurement could have taken place, meaning that missing data were not subtracted. If the number of measurements differed for studies with more than one participant, the median number of measurement occasions was taken with the exception of multiple baseline designs where the overall lengths of the study was coded. If the number of participants and/or the number of measurements could not be identified, the code “n/a” (for “not available”) was used.

Term. As discussed in the introduction, hybrid designs go by different names. To gain a better understanding of which terms are used by applied researchers, we coded the terms used to describe the design from the abstract to answer research question 9. If the design was not specified in the abstract, the term used in the methods section was coded. This coding category was continuously updated until four major categories emerged: combination, combined, embedded, and other.2

Results

General results

In total, 67 studies were identified that met all inclusion criteria. Figure 2 shows the study inclusion flowchart. Figure 3 provides an overview of the frequency of articles per year: 2016 (n = 5, 7.45%), 2017 (n = 20, 29.85%), 2018 (n = 16, 23.88%), 2019 (n = 15, 22.38%), and 2020 (n = 11, 16.42%). The most popular design was a hybrid of multiple baseline
and phase designs, which accounted for 35.82% \((n=24)\) of all included studies. Across all studies, the median number of participants and measurements was three (IQR=2) and 31 (IQR=22) respectively. Over a quarter of all studies included some form of randomization in the design \((n=18, 26.87\%)\), but no study conducted a randomization test. The average number of data aspects analyzed was 2.95 \((SD=1.77)\). The most frequent method of data analysis was visual analysis paired with descriptive statistics \((n=36, 53.73\%)\). Table 2 provides an in-depth overview of the results per hybrid design.

Hybrid designs

**Multiple baseline and phase designs.** This was the most frequently used hybrid design \((n=24, 35.82\%)\). Given that multiple baseline designs always contain an AB phase change, all hybrid designs in this category consisted of multiple baseline designs and other phase designs (e.g. ABAB, ABC). In the majority of this hybrid design, the multiple baseline design was named as the dominant design \((n=21, 87.5\%)\). Randomization was implemented in 20.83% \((n=5)\) of the studies with random allocation of participants, behaviors, or settings to predetermined baseline lengths being the most frequent mode of randomizing, occurring in 60% \((n=3)\) of randomized studies. The remaining two randomized studies did not specify the exact manner in which randomization was implemented. The median number of participants was three with a median of 28 measurements. The most frequently used method of data analysis was visual analysis paired with descriptive statistics which was used in just over half of this hybrid design \((n=13, 54.17\%)\). This hybrid design showed the second highest number of data aspects analyzed \((M=3.33, SD=1.52)\).

**Multiple baseline and alternation designs.** This was the second most frequently used hybrid design \((n=19, 28.35\%)\). The alternation design was named as the dominant design in nine (47.37%) and the multiple baseline design in 10 studies (52.63%). This hybrid design showed by far the highest proportion of randomized studies with over half of the studies...
incorporating some form of randomization \( n = 10, 52.63\% \). The exact proportion of the different forms of randomization can be found in Table 2, but the most common variant was that no specification was given about the way in which randomization had been implemented \( n = 3, 30.00\% \). The median number of participants and measurements was four (IQR = 1.5) and 34 (IQR = 20.5) respectively. Despite the high proportion of randomized studies within this hybrid design, the most frequent modes of analysis were visual analysis alone \( n = 7, 36.84\% \) and visual analysis paired with descriptive statistics \( n = 8, 42.11\% \). The mean number of data aspects analyzed was 2.74 (SD = 2.00).

**Phase and alternation designs.** This hybrid design was used in 16.42\% \( n = 11 \) of the studies. The alternation design was most frequently named as the dominant design \( n = 8, 72.72\% \). Randomization occurred in two studies (18.20\%) with one using restricted randomization and one using blocked randomization. The median number of participants and measurements was two (IQR = 1) and 34.5 (IQR = 16.25). All studies in this category were analyzed using either visual analysis alone \( n = 1, 9.09\% \) or paired with descriptive statistics \( n = 10, 90.90\% \). On average, this hybrid design showed the highest number of data aspects analyzed \( M = 3.36, SD = 1.21 \).

**Multiple baseline and changing criterion designs.** Seven studies (10.45\%) used this hybrid design with all studies mentioning the multiple baseline design as the dominant design. None of the studies incorporated randomization. The most frequent mode of analysis for this hybrid design was a combination of visual analysis with descriptive and inferential statistics \( n = 3, 42.86\% \). The median number of participants and measurements was three (IQR = 0) and 67 (IQR = 64) respectively. Compared to the previous three discussed hybrid designs, the multiple baseline and changing criterion designs

### Table 2. Results per hybrid design.

| Hybrid Design                        | Randomization (n, %) | Participants (Median, IQR) | Measurements (Median, IQR) | Analysis (n, %) | Number of data aspects (mean, SD) | Terms (n, %) |
|--------------------------------------|----------------------|----------------------------|---------------------------|----------------|-----------------------------------|-------------|
| Multiple baseline and phase          | Phase lengths (3, 12.5) Not specified (2, 8.33) Not randomized (19, 79.17) | 3, 2 | 28, 20.5 | Visual (2, 8.33) Inferential stats (1, 4.17) Visual + descriptive stats (13, 54.17) Visual + descriptive + inferential (8, 33.33) | 3.33, 1.52 | Combined (2, 8.33) Combination (0) Embedded (12, 50.00) Other (10, 41.67) |
| (24, 35.82%)                         |                      |                            |                           |                |                                   |             |
| Multiple baseline and alternation    | Unrestricted (1, 5.26) Restricted (2, 10.52) Blocked (1, 10.52) Phase lengths (1, 5.26) Blocked and phase lengths (1, 5.26) Not specified (3, 15.79) Not randomized (9, 47.37) | 4, 1.5 | 34, 20.5 | Visual (7, 36.84) Visual + descriptive stats (8, 42.11) Visual + descriptive + inferential (4, 21.05) | 2.74, 2.00 | Combined (1, 5.26) Combination (3, 15.79) Embedded (7, 36.84) Other (8, 42.11) |
| (19, 28.35%)                         |                      |                            |                           |                |                                   |             |
| Phase and alternation                | Restricted (1, 9.10) Blocked (1, 9.10) Not randomized (9, 81.18) | 2, 1 | 34.5, 16.25 | Visual (1, 9.09) Visual + descriptive stats (10, 90.90) | 3.36, 1.21 | Combination (1, 9.10) Embedded (4, 36.36) Other (6, 54.54) |
| (11, 16.42%)                         |                      |                            |                           |                |                                   |             |
| Multiple baseline and changing       | Not randomized (7, 100) | 3, 0 | 67, 64 | Visual (2, 28.57) Visual + descriptive stats (2, 28.57) Visual + descriptive + inferential (3, 42.86) | 2.29, 2.69 | Combined (4, 57.14) Embedded (2, 28.57) Other (1, 14.29) |
| criterion (7, 10.45)                 |                      |                            |                           |                |                                   |             |
| More than two                        | Restricted (1, 33.33) Not randomized (2, 66.66) | 6, 2.5 | 41, 5.33 | Descriptive stats (1, 33.33) Visual + descriptive stats (1, 33.33) Visual + descriptive + inferential (1, 33.33) | 3.00, 1.73 | Combined (1, 33.33) Embedded (1, 33.33) Other (1, 33.33) |
| (3, 4.48)                            |                      |                            |                           |                |                                   |             |
| Other                                | Not randomized (3, 100) | 2, 1 | 25, 11 | Visual (1, 33.33) Visual + descriptive stats (2, 66.67) | 1.33, 0.58 | Embedded (1, 33.33) Other (2, 66.67) |
| (3, 4.48)                            |                      |                            |                           |                |                                   |             |

IQR = interquartile range.
showed the lowest average number of data aspects analyzed \((M=2.26, SD=2.69)\).

**More than two.** Three studies (4.48\%) used hybrid designs consisting of more than two of the basic designs. Two studies used a hybrid consisting of phase, alternation, and multiple baseline designs. One of these two studies implemented restricted randomization. One study used a hybrid consisting of phase, multiple baseline, and changing criterion designs. The number of participants for the three studies was three, six, and eight with 41, 23, and 48 measurements. One study was analyzed with descriptive statistics, one with visual analysis and descriptive statistics, and one with visual analysis, descriptive, and inferential statistics. The number of data aspects analyzed was one, four, and four.

**Other.** Three studies (4.48\%) used a hybrid design not falling under the previous categories. One study used a hybrid consisting of changing criterion and alternation designs and two studies used hybrid designs consisting of changing criterion and phase designs (other than the regular phase structure of changing criterion design). Randomization did not occur in any of these studies. The number of participants was three, one, and two with 27, 25, and 5 measurements. One study was analyzed with visual analysis alone and two studies with visual analysis paired with descriptive statistics. The number of data aspects analyzed was two, one, and one.

**Terms used**

Table 2 offers an overview of the terms used per design. The overall results indicate that hybrid designs go by many different terms. The majority of the terms used fell under the category “other” \((n=28, 41.79\%)\). However, while many different terms were used, one term occurred nearly as frequently as those other terms combined: embedded \((n=27, 40.30\%)\). The terms combined \((n=8, 11.94\%)\) and combination \((n=4, 5.97\%)\) accounted for a much smaller proportion and the term hybrid design was not used at all.

**Design examples**

Table 3 gives an overview of selected studies included in the systematic review per hybrid design with the goal of the study and references for the interested reader.

**Discussion**

The goal of the present systematic review was to assess recent publication trends in hybrid SCED studies. After a clear increase between the years 2016 and 2017, there was a decelerating trend in the frequency of hybrid SCED publications. The results align with the ones of Shadish and Sullivan (2011) and Hammond and Gast (2010) regarding the two most frequently used hybrid designs: multiple baseline and phase designs, and multiple baseline and alternation designs. While the present systematic review and the results of Shadish and Sullivan found that the multiple baseline and phase design hybrid was used most frequently, Hammond and Gast found that the multiple baseline and alternation hybrid was used most frequently. While both Shadish and Sullivan as well as Hammond and Gast found that the phase and alternation hybrid accounted only for a small fraction, the present systematic review indicates that it is the third most frequently used hybrid which may be indicative of a growing popularity of that particular design. Hammond and Gast did not report any hybrid designs consisting of more than two of the basic designs, but the results of the present study and Shadish and Sullivan indicate a low prevalence of such hybrid designs. This is however not indicative of the value of such designs, but could rather point to a lack of familiarity of hybrid designs consisting of more than two of the basic designs.

What stands out further is the low prevalence of the changing criterion design as part of hybrid designs. In comparison to the other three basic designs, the changing criterion design was only found in a small proportion of hybrid designs. This is in line with the overall low prevalence of this design as evidenced in other systematic reviews (Hammond and Gast, 2010; Shadish and Sullivan, 2011; Tانious and Onghena, 2021). However, it should be emphasized that the changing criterion design has many desirable features, especially in educational settings. These features include that the intervention is never withdrawn and that changes in the dependent variable(s) are expected to occur gradually. Therefore, it is not entirely clear why the changing criterion design, either as a standalone design or as part of a hybrid design, continues to be used somewhat infrequently. Researchers wishing to utilize this design are referred to the guidelines by Klein et al. (2017).

Regarding the descriptive study characteristics, several comparisons can be made. The median number of three participants for hybrid designs is identical to the one reported by Shadish and Sullivan and Tαnious and Onghena (2021) for all SCED designs. The median number of 31 measurements per time series is somewhat higher than the median number of 21 measurements reported by Shadish and Sullivan. However, both are subject to large variability: The present study found a range of 5–190 and Shadish and Sullivan found a range of 2–160 measurements. The average number of data aspects analyzed \((M=2.95, SD=1.77)\) was somewhat higher than the one reported in Tαnious and Onghena for all SCED designs \((M=2.61, SD=1.63)\). Similarly, the proportion of randomized hybrid design studies (26.87\%) was somewhat higher than the proportion reported in Tαnious and Onghena (22.46\%).

In terms of randomization, it should further be noted that the largest proportion of randomized studies stemmed from the multiple baseline and alternation hybrid. When looking at each hybrid design separately, it is noteworthy that the other designs had a relatively low prevalence of randomized
studies. Particularly surprising is the result that randomiza-
tion was quite uncommon in the phase and alternation hybrid, which was the most frequently used hybrid design overall. This result is surprising for at least two reasons. First, the multiple baseline and alternation design showed, by far, the highest proportion of randomized studies. Given that multiple baseline designs also rely on a phase structure, it would be evident to assume that the alternation and phase design would show similarly high proportions of randomized studies. Second, the alternation design may be considered the most logical design to incorporate randomization because it relies on the rapid alternation of treatments (Barlow and Hayes, 1979). It would therefore be straightforward to apply randomization to any form of hybrid design consisting in part of alternation designs.

Apart from these methodological study characteristics, several other observations are noteworthy. Even though we did not formally quantify these observations, it stood out that nearly all studies shared some commonalities. Usually the participants were children or adolescents and studies took place in an educational and/or therapeutical setting. Oftentimes, the participants had (developmental) disabili-
ties, for example autism spectrum disorder or dyslexia. Some commonly observed dependent variables were on-task behavior, academic engagement, disruptive behavior, and other learning outcomes.

Table 3. Examples of published studies per hybrid design.

| Hybrid design                              | Goal                                                                 | Reference                      |
|--------------------------------------------|---------------------------------------------------------------------|--------------------------------|
| Multiple baseline and phase                | Increase academic engagement and reduce disruptive behavior in three first-grade classrooms | Radley et al. (2016)           |
|                                            | Investigate the effects of Class-Wide Function-Related Intervention Teams (CW-FIT) on on-task behavior, teacher praise and teacher reprimand rates | Jolstead et al. (2017)         |
|                                            | Increase the percentage of tasks completed independently by college students with intellectual disability and autism spectrum disorders by implementing a wearable smartwatch-based intervention | Wright et al. (2020)           |
| Multiple baseline and alternation          | Investigate whether the use of simultaneous multisensory structured language instruction promoted better letter name and sound production, word reading, and word spelling for second grade children with typical development or with dyslexia than structured language instruction alone. | Schlesinger and Gray (2017)    |
|                                            | Evaluate the effectiveness of implementation planning and participant modeling on teachers' implementation of an evidence-based classroom management plan. | Hagermoser Sanetti et al. (2018) |
|                                            | Compare the efficiency of two prompt strategies on establishing the intraverbal behavior (i.e. question answering) for three 4-year-old children with autism spectrum disorder in China | Xu et al. (2020)               |
| Phase and alternation                      | Compare the effects of an independent group contingency and a randomized dependent group contingency on compliance with assigned literacy worksheets in a first-grade general education classroom | Deshais et al. (2019)         |
|                                            | Evaluate the effect of demand fading on escape-maintained challenging behavior, mands, and task completion for two children with autism spectrum disorder | Gerow et al. (2020)           |
|                                            | Evaluate the effects of instructional prompts on letter identification for two kindergarten students who experienced difficulties with letter identification skills | Hathaway et al. (2021)*       |
| Multiple baseline and changing criterion   | Evaluate the effects of an interdependent group contingency for decreasing disruptive behavior in adolescents across two therapy groups | Brogan et al. (2017)          |
|                                            | Examining the effects of backward chaining in combination with least-to-most prompting, time delay, and differential reinforcement on the acquisition of functional communication skills in children with autism spectrum disorder and developmental disabilities | Muharib et al. (2019)         |
|                                            | Assess the effectiveness of a paraphrasing intervention on the problem-solving performance for nine third-grade students who are English learners and at risk of mathematical disabilities | Kong and Swanson (2019)       |
| Phase and changing criterion              | Treating elopement without blocking in a child with autism          | Boyle et al. (2019)           |
| Phase, alternation, and multiple baseline  | Examine two aspects of treatment intensity in treatment for childhood apraxia of speech: practice amount and practice distribution. | Maas et al. (2019)            |

*First published online in 2020.
Revisited: A note on terminology

As stated in the introduction, SCED hybrid designs go by many names. For the purposes of this paper, we used the term “hybrid design” which is commonly used in the SCED methodology literature. However, the results of the present systematic review indicate that the term is not known by applied researchers. While many different terms were found, the single most popular term was “embedded.” Examples of the use of this term are “multiple-baseline design with an embedded ABAB condition sequence” (Radley et al., 2016), “reversal design with an embedded multielement design” (Deshais et al., 2019), and “adapted alternating treatments design embedded in a multiple baseline across participants design” (Yuan and Zhu, 2020). Some examples of terms coded in the category “other” are “nested,” “and,” and “with.” The results further indicate that the terms “combined” and “combination,” while common in SCED methodology literature, were also rather uncommon. For future research synthesis and bridging the divide between applied researchers and SCED methodologists, it seems beneficial to settle on one term to describe this class of SCED studies. The results of the present systematic review clearly favor the term “embedded.”

Simulation studies and development of guidelines

As stated in the introduction, specific guidelines on the conduct and analyses of SCED hybrid designs are lacking. We hope that the present study can give some impetus toward the development of these guidelines. A first step in the direction of analytical methods for hybrid designs specifically has been made by Moeyaert et al. (2020) who proposed regression-based effect size estimations and hierarchical linear modeling. As Moeyaert et al. however note, simulation studies can be used to further investigate the power of certain analytical methods for particular design conditions. The specification of these design conditions is of course of great importance to the validity of any results obtained via simulation studies. The results of the present systematic review may guide the use of design characteristics in simulation studies in terms of the specific design, the typical number of participants and measurements, and the specific forms of randomization. Regarding the use of analytical methods, the results indicate that guidelines on the use of inferential statistical methods for hybrid designs are badly needed. Within each hybrid design, by far the most frequently used analytical method was visual analysis, either alone or paired with descriptive statistics. This held true even for hybrid designs with a relatively high proportion of randomized studies (i.e. multiple baseline and alternation).

Accordingly, the results further suggest that the development of hybrid specific guidelines should focus on proposing design specific randomization schemes. In hybrid designs it is always possible to incorporate several ways of randomization, namely at least one per basic design. However, only one of all randomized studies did use more than one form of randomization. With the exception of the multiple baseline and alternation hybrid, the proportion of randomized hybrid designs was relatively low or even zero for some designs (i.e. multiple baseline and changing criterion, phase and changing criterion, alternation and changing criterion). Given that all nonrandomized hybrids include a changing criterion design component, the randomization procedures proposed by Ferron et al. (2019) and Onghena et al. (2019) for changing criterion designs may be of particular interest. Another noteworthy result regarding randomization is that in studies that did use randomization, a considerable proportion did not specify the exact manner in which randomization was used. Future reporting guidelines should therefore emphasize the need for precise reporting of the manner in which randomization was implemented, similar to item 8 in the single-case reporting guidelines proposed by Tate et al. (2016).

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

René Tanious https://orcid.org/0000-0002-5466-1002

Notes

1. We use the term “case” to refer more broadly to the participants in SCDEs. Cases can be aggregated at different levels to form the unit of analysis: Individual (e.g. a student), group (e.g. a classroom), or organizational (e.g. a school).
2. The term hybrid was not found in the sample. Implications for future terminology of this design are discussed in the section “Revisited: A note on terminology.”

References

Baer DM, Wolf MM and Risley TR (1968) Some current dimensions of applied behavior analysis. Journal of Applied Behavior Analysis 1: 91–97.
Barlow DH and Hayes SC (1979) Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. Journal of Applied Behavior Analysis 12: 199–210.
Barlow DH, Nock M and Hersen M (2009) Single-Case Experimental Designs: Strategies for Studying Behavior for Change, 3rd edn. Boston, MA: Pearson.
Boyle MA, Keenan G, Forck KL, et al. (2019) Treatment of elopement without blocking with a child with autism. Behavior Modification 43(1): 132–145.
Brogan KM, Falligant JM and Rapp JT (2017) Interdependent group contingencies decrease adolescents’ disruptive behaviors during group therapy: A practitioner’s demonstration. Behavior Modification 41(3): 405–421.
Deshais MA, Fisher AB and Kahng S (2019) A comparison of group contingencies on academic compliance. *Journal of Applied Behavior Analysis* 52(1): 116–131.

Edgington ES (1980a) Overcoming obstacles to single-subject experimentation. *Journal of Educational Statistics* 5(3): 261–267.

Edgington ES (1980b) Validity of randomization tests for one-subject experiments. *Journal of Educational Statistics* 5(3): 235–251.

Ferron J, Rohrer LL and Levin JR (2019) Randomization procedures for changing criterion designs. *Behavior Modification*. 145445519847627. DOI: 10.1177/0145445519847627

Gerow S, Radhakrishnan S, Davis TN, et al. (2020) A comparison of demand fading and a dense schedule of reinforcement during functional communication training. *Behavior Analysis in Practice* 13: 90–103.

Gusenbauer M and Haddaway NR (2020) Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of google scholar, PubMed, and 26 other resources. *Research Synthesis Methods* 11(2): 181–217.

Hagermoser Sanetti LM, Williamson KM, Long ACJ, et al. (2018) Increasing in-service teacher implementation of classroom management practices through consultation, implementation planning, and participant modeling. *Journal of Positive Behavior Interventions* 20(1): 43–59.

Hammond D and Gast DL (2010) Descriptive analysis of single-subject research designs: 1983-2007. *Education and Training in Autism and Developmental Disabilities* 45(2): 187–202. Available at: https://www.jstor.org/stable/23879806 (accessed 17 May 2021).

Hartmann DP and Hall RV (1976) The changing criterion design. *Journal of Applied Behavior Analysis* 9: 527–532.

Hathaway KL, Schielzelt KM and Dietrick J (2021) Evaluating the effects of instructional prompts and strategic incremental rehearsal on the letter identification mastery of two typically developing kindergarteners. *Behavior Analysis in Practice* 14: 20–35.

Heyvaert M and Onghena P (2014) Analysis of single-case data: Randomisation tests for measures of effect size. *Neuropsychological Rehabilitation* 24(3–4): 507–527.

Jolstead KA, Caldarella P, Hansen B, et al. (2017) Implementing positive behavior support in preschools: An exploratory study of CW-FIT Tier I. *Journal of Positive Behavior Interventions* 19(1): 48–60.

Klein LA, Houlihan D, Vincent JL, et al. (2017) Best practices in utilizing the changing criterion design. *Behavior Analysis in Practice* 10: 52–61.

Kong JE and Swanson HL (2019) The effects of a paraphrasing intervention on word problem-solving accuracy of English learners at risk of mathematic disabilities. *Learning Disability Quarterly* 42(2): 92–104.

Kratochwill TR, Hitchcock J, Horner RH, et al. (2010) Single-case designs technical documentation. What Works Clearinghouse. Available at: https://files.eric.ed.gov/fulltext/ED510743.pdf (accessed 20 April 2021)

Kratochwill TR and Levin JR (2010) Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods* 15(2): 124–144.

Lane JD and Gast DL (2014) Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation* 24: 445–463.

Ledford JR and Gast DL (eds) (2018) *Single Case Research Methodology: Applications in Special Education and Behavioral Sciences*, 3rd edn. New York, London: Routledge.

Maas E, Gilderselee-Neumann C, Jakielski K, et al. (2019) Bang for your buck: A single-case experimental design study of practice amount and distribution in treatment for childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research* 62(9): 3160–3182.

Manolov R and Moeyaert M (2017a) How can single-case data be analyzed? Software resources, tutorial, and reflections on analysis. *Behavior Modification* 41(2): 179–228.

Manolov R and Moeyaert M (2017b) Recommendations for choosing single-case data analytical techniques. *Behavior Therapy* 48(1): 97–114.

Marasculo LA and Busk PL (1988) Combining statistics for multiple-baseline ab and replicated ABAB designs across subjects. *Behavioral Assessment* 10(1): 1–28.

McDonnell J and O’Neill R (2003) A perspective on single/within subject research methods and “scientifically based research”. *Research and Practice for Persons with Severe Disabilities* 28(3): 138–142.

McDougall D (2005) The range-bound changing criterion design. *Behavioral Interventions* 20: 129–137.

Moeyaert M, Akhmedjanova D, Ferron J, et al. (2020) Effect size estimation for combined single-case experimental designs. *Evidence-Based Communication Assessment and Intervention* 14(1–2): 28–51.

Muharib R, Alzrayer NM, Wood CL, et al. (2019) Backward chaining and speech-output technologies to enhance functional communication skills of children with autism spectrum disorder and developmental disabilities. *Augmentative and Alternative Communication* 35(4): 251–262.

Onghena P (1992) Randomization tests for extensions and variations of ABAB single-case experimental designs: A rejoinder. *Behavioral Assessment* 14: 153–171.

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

Onghena P and Edgington ES (2005) Customization of pain treatments: Single-case design and analysis. *Clinical Journal of Pain* 21: 56–68; discussion 69.

Onghena P, Tanious R, De TK, et al. (2019) Randomization tests for changing criterion designs. *Behaviour Research and Therapy* 117: 18–27.

Plavnick JB and Ferreri SJ (2013) Single-case experimental designs in educational research: A methodology for causal analyses in teaching and learning. *Educational Psychology Review* 25: 549–569.

Pustejovsky JE and Ferron JM (2017) Research synthesis and meta-analysis of single-case designs. In: Kaufmann JM, Hallahan DP and Pullen PC (eds) *Handbook of Special Education*. New York: Routledge, pp.168–185.

Radley KC, Dart EH and O’Handley RD (2016) The quiet classroom game: A class-wide intervention to increase academic engagement and reduce disruptive behavior. *School Psychology Review* 45(1): 93–108.
Schlesinger NW and Gray S (2017) The impact of multisensory instruction on learning letter names and sounds, word reading, and spelling. *Annals of Dyslexia* 67: 219–258.

Shadish WR and Sullivan KJ (2011) Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods* 43: 971–980.

Swan DM, Pustejovsky JE and Beretvas SN (2020) The impact of response-guided designs on count outcomes in single-case experimental design baselines. *Evidence-Based Communication Assessment and Intervention* 14(1-2): 82–107.

Tanious R, De TK, Michiels B, et al. (2020) Assessing consistency in single-case A-B-A-B phase designs. *Behavior Modification* 44(4): 518–551.

Tanious R and Onghena P (2021) A systematic review of applied single-case research published between 2016 and 2018: Study designs, randomization, data aspects, and data analysis. *Behavior Research Methods* 53: 1371–1384.

Tate RL, Perdices M, Rosenkoetter U, et al. (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.

Tate RL, Perdices M, Rosenkoetter U, et al. (2016) The single-case reporting guideline In BEhavioural interventions (SCRIBE) 2016 statement. *Aphasiology* 30(7): 862–876.

Wampold BE and Worsham NL (1986) Randomization tests for multiple-baseline designs. *Behavioral Assessment* 8(2): 135–143.

Welch BL (1937) On the z-test in randomized blocks and Latin squares. *Biometrika* 29(1/2): 21–52.

What Works Clearinghouse (2020a) *Procedures Handbook (Version 4.1).* Institute of Education Sciences. Available at: https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Procedures-Handbook-v4-1-508.pdf (accessed 20 April 2021)

What Works Clearinghouse (2020b) *Standards Handbook, version 4.1.* Institute of Education Sciences. Available at: https://ies.ed.gov/ncee/wwc/Docs/referenceresources/W WC-Standards-Handbook-v4-1-508.pdf (accessed 20 April 2021)

Wright RE, McMahon DD, Cihak DF, et al. (2020) Smartwatch executive function supports for students with ID and ASD. *Journal of Special Education Technology.* Epub ahead of print 15 September 2020. DOI: 10.1177/0162643420950027

Xu S, Niu R, Lee GT, et al. (2020) Comparing the effects of echoic prompts and echoic plus picture prompts on establishing intraverbal behavior for children with Autism. *Education and Training in Autism and Developmental Disabilities* 55(1): 89–100.

Yuan C and Zhu J (2020) An evaluation of prompting procedures in error correction for children with autism. *Behavioral Interventions* 35(4): 581–594.

**Author biographies**

René Tanious is a postdoctoral researcher at the Faculty of Psychology and Educational Sciences at KU Leuven, Belgium. His current research interests include single-case experimental designs, randomization, systematic reviews, and combining statistical and visual analysis for single-case experimental designs.

Patrick Onghena is a professor of Research Methodology and Statistics at the Faculty of Psychology and Educational Sciences, KU Leuven, Belgium. His major research interests include single-case experimental designs, distribution-free statistical inference, the methodology of systematic reviews, mixed methods research, and research on the teaching of statistics.