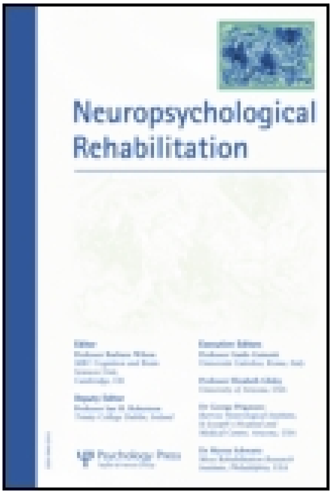


Neuropsychological Rehabilitation: An International Journal

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pnrh20

Visual analysis in single case experimental design studies: Brief review and guidelines

Justin D. Lane[a] & David L. Gast[a]

[a] Department of Special Education, The University of Georgia, Athens, GA, USA

Published online: 24 Jul 2013.

Neuropsychological Rehabilitation, 2014
Vol. 24, Nos. 3–4, 445–463, http://dx.doi.org/10.1080/09602011.2013.815636

Visual analysis in single case experimental design studies: Brief review and guidelines

Justin D. Lane and David L. Gast
Department of Special Education, The University of Georgia, Athens, GA, USA

(*Received 17 May 2013; accepted 10 June 2013*)

Visual analysis of graphic displays of data is a cornerstone of studies using a single case experimental design (SCED). Data are graphed for each participant during a study with trend, level, and stability of data assessed within and between conditions. Reliable interpretations of effects of an intervention are dependent on researchers' understanding and use of systematic procedures. The purpose of this paper is to provide readers with a rationale for visual analysis of data when using a SCED, a step-by-step guide for conducting a visual analysis of graphed data, as well as to highlight considerations for persons interested in using visual analysis to evaluate an intervention, especially the importance of collecting reliability data for dependent measures and fidelity of implementation of study procedures.

Keywords: Single case experimental design; Visual analysis.

INTRODUCTION

Visual analysis of graphic displays of data is the hallmark for interpreting effects of an intervention during studies using a single case experimental design (SCED; Kennedy, 2005). The independent variable is typically an intervention designed to reduce aberrant behaviour or increase pro-social or academic behaviours (Horner et al., 2005). The expectation is behaviours are graphed for each participant by session for all conditions of a study. SCED studies are based on baseline logic, meaning participants serve as their own control for evaluating change (Gast & Hammond, 2010). The most basic method for evaluating a behaviour-change programme is analysis

Correspondence should be addressed to Justin D. Lane, Department of Special Education, The University of Georgia, Athens, GA 30602, USA. E-mail:jdlane@uga.edu

of two adjacent conditions, also known as an A-B comparison. Each condition represents a set of specific variables under which a behaviour is measured continuously during baseline, intervention, and any subsequent conditions, with the first condition known as baseline (A) and the second condition known as intervention (B) (Lane, Wolery, Reichow, & Rogers, 2007). The expectation is that data collected during baseline are stable and change in a therapeutic direction upon introduction of the independent variable, with at least three replications of effect across behaviours, settings, or participants (Gast & Hammond, 2010; Kennedy, 2005; Lane et al., 2007). This is in contrast to traditional group design methods, which use at least one experimental group and one control group to compare effects of treatment through application of a statistical test (e.g., ANOVA) and a significance measure (e.g., $p < .05$; Kazdin, 2011). Traditional group design methods provide an opportunity for summative evaluation of effect for all participants. In contrast, SCED studies allow researchers to formatively evaluate a participant's performance session to session through continuous collection of individualised behaviour data. Researchers can modify an intervention if limited or no change in performance is observed throughout the study (Gast, 2010; Lieberman, Yoder, Reichow, & Wolery, 2010; Wolery & Harris, 1982).

SCED researchers traditionally code responses for each participant in vivo or following completion of a session using video records (Ayres & Gast, 2010). Each participant's performance is then calculated and transferred to a graph for purposes of visually analysing (a) trend, (b) level, and (c) stability. Researchers select a format for graphic display which best represents the purpose of the study based on the proposed hypotheses and research questions (Spriggs & Gast, 2010). There are various examples of graphic formats for presenting data, which include (a) cumulative records, (b) semi-logarithmic charts, (c) bar graphs, and (d) line graphs. Cumulative records are rooted in the field of experimental analysis of behaviour and originated from work by B. F. Skinner (Cooper, Heron, & Heward, 2007; Kennedy, 2005). When using a cumulative record, participants' responses are recorded and added across sessions. It is not uncommon for novice readers to report an increasing trend across sessions since responses are added session to session. Further inspection of a cumulative record may indicate minimal (e.g., addition of one response from previous session) to no change (flat line from session to session) in responding. Regardless of the magnitude of behaviour change, one of the benefits of a cumulative record is clear depiction of total responses within a study. Semi-logarithmic charts provide a format for presenting relative change in performance and are typically used to present rate of responding and proportional change in performance. In contrast to cumulative records and semi-logarithmic charts, bar graphs provide a simplified format for comparisons of discrete data or presentation of summative performance following completion of a study. Finally, line graphs are the most commonly used graphic format for presenting ongoing data collected during a study using a SCED. Performance within a session is plotted as a single data point and connected to subsequent data points as the study progresses (Cooper et al., 2007; Kennedy, 2005; Spriggs & Gast, 2010). For purposes of this paper, examples and information will focus on analysis of data using a line graph.

Researchers use A-B-C notation to differentiate conditions on a graphic display. A major tenet of SCED is that all conditions remain constant with exception of the introduction of one variable in the intervention condition. Additional components of an intervention may be introduced within a condition, also known as a phase change, but it is still recommended that any additional variables be introduced systematically for purposes of controlling threats to internal validity (Gast & Spriggs, 2010; Lane et al., 2007). If any modifications are made in a condition the expectation is to denote this alteration with a phase change line and a prime symbol (e.g., B' for one modification and B'' for an additional modification and so on). Also, if any subsequent conditions are introduced individual letters are added in alphabetical order to denote a novel independent variable. When two variables are combined in a single condition, two letters are used to denote this combination (e.g., BC to represent a combination of intervention B and C). This contributes to the reader's understanding of changes that occur during a single study for each participant (Gast & Spriggs, 2010).

Cooper et al. (2007) highlight the impact of graphic displays by indicating "an intervention that produces dramatic, replicable changes in behaviour that last over time are readily seen in a well-designed graphic display" (p. 149). While a "well-designed graphic display" involves multiple factors, ongoing analysis of data contributes to a better understanding of effects of interventions for researchers, as well as future readers. As stated above, visual analysis involves evaluation of (a) trend, (b) level, and (c) stability of data. Wolery and Harris (1982) define (a) trend, as "direction. . .the data pattern is progressing", (b) level, as the "relative value of the data pattern on the dependent variables", and (c) stability, as similarity "of scores in a given experimental condition" (pp. 446–447). Gast (2005) elaborates on these definitions by defining (a) trend, as "progress over time", (b) level, as "magnitude of the data", and (c) stability, as "variability or 'bounce' of the data" (pp. 1596–1597). Together, these three components provide the foundation for visual analysis of behaviours within and between conditions.

Within-condition analysis refers to evaluation of data patterns within a single condition during a study (Gast & Spriggs, 2010; Kennedy, 2005; Wolery & Harris, 1982). Beginning with baseline, researchers look for stability of data prior to implementation of an intervention, with data collected across a minimum of at least three to five sessions prior to introduction of an intervention (Horner et al., 2005). If variability is observed within a condition, the recommendation is to extend that condition until data are stable.

This also applies to behaviours changing in a therapeutic direction during baseline. In this instance, researchers also attend to the trend direction of the data. Participants may improve during a baseline condition due to maturation, response to specific environmental variables, or other factors outside of the study. Regardless of cause of change in a therapeutic direction, a researcher would be advised to wait until a clear pattern or stability is observed (Kennedy, 2005; Wolery & Harris, 1982). One method for evaluating trend direction is the split-middle method of trend estimation, which is a multi-step process and, as the name states, estimates trend direction within each condition based on calculations of median values (Wolery & Harris, 1982). The split-middle method provides researchers with a systematic process for estimating trend direction, but is less common in the literature when compared to other evaluation methods (Gast & Spriggs, 2010; White, 1972). Within-condition analysis begins during the first condition of the study and continues for the duration of the study, which is followed by between-conditions analysis of data as new conditions are introduced.

Between-conditions analysis of data refers to comparisons across adjacent conditions during a study (Gast & Spriggs, 2010; Kennedy, 2005; Wolery & Harris, 1982). A SCED researcher would look for an immediate and abrupt change in level and trend upon introduction of the independent variable. With consideration of trend and level, trend is considered more important for researchers conducting visual analysis of data (Gast & Spriggs, 2010). The researcher would also attend to variability in the data with consideration of overlap of data points across conditions. In the "how to" portion of this paper, specific methods for comparing adjacent conditions are provided. Consider the following example in the context of an A-B design: A participant enrolled in a study evaluating an intervention to increase percent responsivity to static presentation of various facial expressions by adults may display low and stable levels of attention (zero-celerating trend) during baseline and, upon introduction of the intervention, attending behaviours increase from 7% (final data point in baseline) to 15% (first intervention session). While there was an immediate level change in attention, the goal of the study was to increase attention over time. During subsequent sessions attending behaviours increase and stabilise at 55% across three consecutive sessions. This pattern in the data indicates an increasing trend in a therapeutic direction, as well as increase in level and stability of data. A replication of this effect across at least three participants within a study would provide the researcher data to support the likelihood of a functional relation between attending behaviours and the intervention for increasing attending behaviours for persons who display similar pre-treatment behaviours (Gast & Spriggs, 2010; Kennedy, 2005; Reichow, Volkmar, & Cicchetti, 2008).

With consideration of identifying a functional relation, measures of generality in SCED studies are evaluated through replication of effect with a single participant (intra-subject replication) and/or across participants (inter-subject replication) within a study and across studies (Kennedy, 2005). To assist researchers who want to replicate a study, detailed descriptions of procedures and participants' pre-treatment behaviours as they relate to the study are crucial (Horner et al., 2005; Lane et al., 2007; Reichow et al., 2008). While these general rules provide guidelines for researchers using a SCED design, attention to patterns of data and decisions based on within and between condition analyses should be the driving force behind a well-designed study (Horner et al., 2005; Odom et al., 2005). The remainder of this paper will provide details for visually analysing graphed data, followed by a discussion of strengths and challenges of visual analysis of SCED studies, as well as considerations and recommendations for researchers.

HOW TO VISUALLY ANALYSE DATA

The purpose of this section is to apply basic principles of visual analysis using the above hypothetical scenario (i.e., intervention to increase percent responsivity to static presentation of various facial expressions by adults). Gast and Spriggs (2010) provide a detailed description for visual analysis of graphed data using a SCED, which is the basis for the "how to" section of this paper. Each step is numbered and designed as an assistive tool for novice researchers who will visually analyse data when using a SCED. Steps are based on graphic display (see Figure 1 for hypothetical data for purposes of exemplifying visual analysis) and divided into (a) within-condition and (b) between-conditions analysis of data (see Figure 2). Figures 3–9 provide detailed guidelines for conducting each step of the visual analysis process.

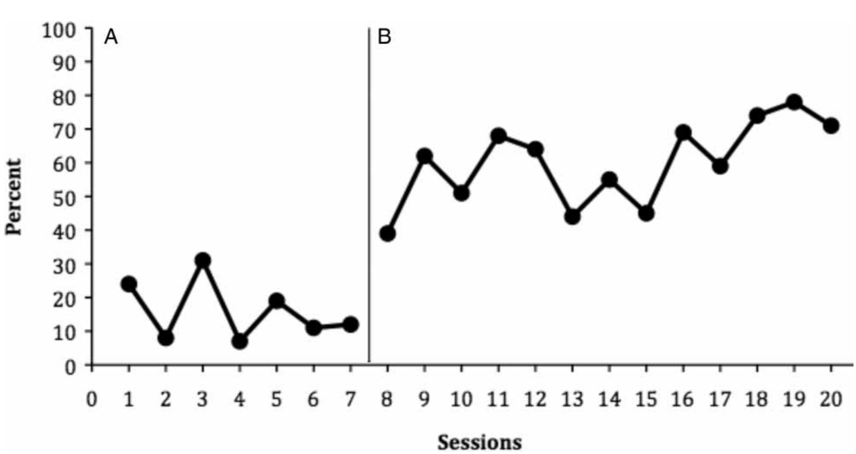


Figure 1. Graphic display using hypothetical data.

Within Condition Analysis of Graphed Data

Step 1. A-B-C Notation

Step 2. Number of Sessions by Condition

Step 3. Stability of Level and Range of Data by Condition

Step 4a. Level Change Within Each Condition

Step 4b. Absolute Level Change

Step 5. Estimate Trend

Step 6. Trend Stability

Step 7. Data Paths Within Trend

Between Conditions Analysis of Graphed Data

Step 1. Determine Number of Variables that Changed Between Conditions

Step 2. Change in Trend Direction Between Adjacent Conditions

Step 3. Change in Trend Stability

Step 4a-d. Level Change Between Conditions

Step 5a-b. Overlap of Data Between Conditions

Figure 2. Steps of visual analysis.

Within-condition analysis

Step 1 is assigning a letter to each condition (i.e., A-B-C notation) and Step 2 is counting the number of sessions for each condition.

Step 3 is calculating the mean, median, range, and stability envelope of data for each condition.

Step 4a is calculating level change within each condition and 4b is calculating the difference between the first and last value within each condition.

Step 5 is calculating trend using the split-middle method of trend estimation.

Step 6 is calculating percent of data points within the stability envelope for each condition and Step 7 is using the "freehand method" to evaluate data paths.

Between-condition analysis

Step 1 is determining the number of variables that changed between conditions. It should be noted that the ideal is only one change across conditions. Step 2 is identifying trend direction across adjacent conditions as accelerating, decelerating, or zero-celerating in a therapeutic or contra-therapeutic

Step 1:
A = Baseline, B = Intervention

Step 2.
Total sessions in A = 7,
Total sessions in B = 13

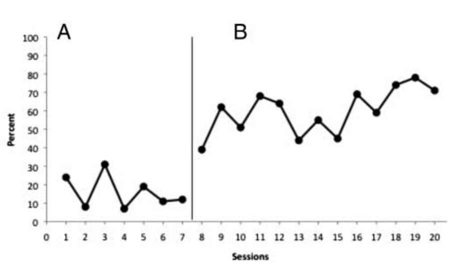


Figure 3. Within-condition analysis: Steps 1 and 2.

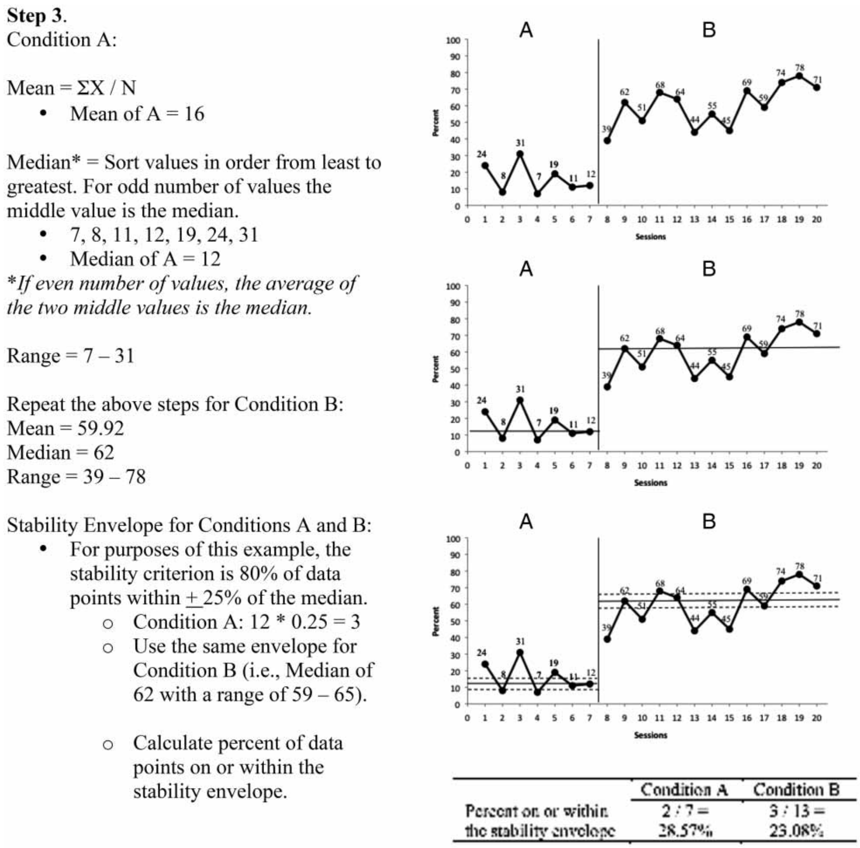


Figure 4. Within-condition analysis: Step 3.

direction. Step 3 is comparing the decision from Step 6 of from the within-condition analysis section to Step 2 of the between-condition analysis section.

Steps 4a–d are evaluating (a) relative, (b) absolute, (c) median, and (d) mean level change.

Steps 5a–b are calculating percent of non-overlapping data (PND) and percent of overlapping data (POD; Scruggs, Mastropieri, & Casto, 1987).

Summary of sample visual analysis

Within-condition analysis. Evaluation of each condition indicated data were variable during baseline and intervention (Step 3). Evaluation of level

Step 4a.
Relative level change:

- For each condition split data in half to identify the median value for each half*.
  - If even values remain after splitting data, calculate the average of the middle values for each half
- Subtract smallest value from largest value and note if the change is "improving" or "deteriorating".

**If odd number of values, the middle value will not be included.*

	Condition A	Condition B
Median of 1st half	24	56.5
Median of 2nd half	12	70
Relative Level Change	-12 Deteriorating	+13.5 Improving

Step 4b.
Absolute level change:

- For each condition identify the first value and last value.
- Subtract smallest value from largest value and note if the change is "improving" or "deteriorating"

	Condition A	Condition B
First Value	24	39
Last Value	12	71
Absolute Level Change	-12 Deteriorating	+32 Improving

Figure 5. Within-condition analysis: Steps 4a-b.

change within conditions indicated performance was deteriorating during baseline and improving during intervention (Steps 4a–b). Split-middle method of trend estimation was conducted and indicated there was decreasing contra-therapeutic trend during baseline and increasing trend in a therapeutic direction during intervention (Step 5), but data were considered variable following application of a stability envelope to trend lines (Step 6).

Between-condition analysis. Evaluation of behaviour change across conditions indicated only one variable was introduced across both conditions (Step 1). With consideration of within-condition analysis of trend, a change in performance across conditions went from a decelerating, deteriorating trend to an accelerating, improving trend (Steps 2 and 3). All level change measures indicated a positive (improving) change across conditions (Steps 4a–d). The limitations of calculating change in level across adjacent conditions should be considered. First, relative level change provides information regarding proportional change from the last half of baseline to the first half of the intervention condition using median values, but does not provide information regarding immediacy of change. Second, absolute level change only provides information regarding the immediacy of change from the last session of baseline to the first session during intervention without consideration of other data points. Third, mean level change may be influenced by outliers within either condition and thus would skew the mean value for the

Step 5.

- For each condition, split data in half (dashed line).
- Identify the mid-date (solid line) for each half*.
- The mid-date is the middle session for each half.
    - Condition A:
        - Mid-date for first half: Session 2
        - Mid-date for second half: Session 6
    - Condition B:
        - Mid-date for first half: Session 10.5**
        - Mid-date for second half: Session 17.5**

**If odd number of values, the middle value will not be included.*
***If even number of values, divide number of sessions for each half by two.*

- For each half of each condition, find mid-rate (horizontal solid line), which is the median value on the ordinate.
    - Condition A:
        - First-half = 24
        - Second-half = 12
    - Condition B:
        - First-half = 56.5*
        - Second-half = 70*

**Even number of values, which means add the two middle values and divide by 2 to obtain the median value.*

- Draw a line passing through the intersection of the mid-date and mid-rate for each condition (diagonal solid line).

- Final step is to move the trend line so there is an "equal number of data points above and below the line" (p. 208).

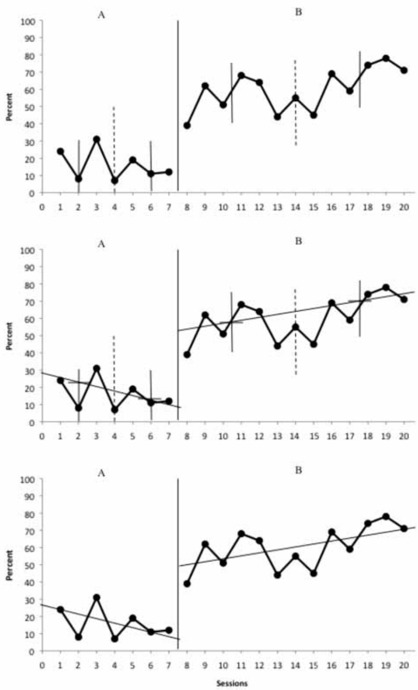
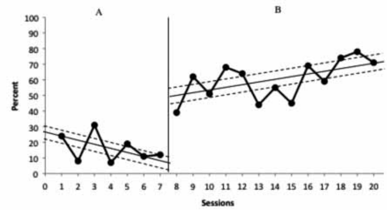


Figure 6. Within-condition analysis: Step 5.

corresponding condition. Finally, due to the limitations of calculating the mean value for each condition, calculating the median level change is recommended since median values are less likely to be influenced by outliers in the data. Based on this information, it is the recommendation of the

Step 6.
Place the stability envelope used in the final step of Step 3 over the trend line created in the final step of Step 5.

- Calculate percent of data points on or within the stability envelope.



	Condition A	Condition B
Percent on or within the stability envelope	3 / 7 = 42.86%	4 / 13 = 30.77%

Step 7.
Review Steps 1 – 6, as well as original graphic display, and determine the following about the trend of each condition:

- Direction.
- Stable or variable?
- Multiple paths within trend?

	Condition A	Condition B
Direction	Decelerating	Accelerating
Stable or variable?	Variable	Variable
Multiple paths within trend?	No	No

Figure 7. Within-condition analysis: Steps 6 and 7.

authors that none of these methods be used in isolation since there are potential limitations across all. Finally, calculations of PND and POD indicated there were 100% non-overlap and 0% overlap of behaviours observed during baseline and intervention (Step 5). If the results were replicated across participants, behaviours, or settings, investigators could report observation of a functional relation (Gast & Spriggs, 2010; Kennedy, 2005; Reichow et al., 2008).

DISCUSSION

Considerations of visual analysis of SCED studies

Visual analysis of graphic displays of data collected during SCED studies is a tradition in fields interested in ongoing evaluation of behaviour-change programmes for individuals or groups of individuals (Wolery & Harris, 1982). This individualised approach to evaluation of behaviours allows opportunity to systematically adapt or modify a programme based on characteristics observed during sessions (Gast, 2005). As with any approach to evaluation of data, limitations and recommendations which maximise understanding of results must be addressed. First, researchers using a SCED should evaluate data using multiple methods to better understand and improve confidence in findings (Gast, 2005). For example, PND provides a metric for measuring

Step 1.
Only one variable changed from baseline to introduction of the intervention during the intervention condition.

Step 2.
For each condition note the type of change in trend direction across adjacent conditions.

- Would summarize the trend across conditions A and B as follows:
  - "A change from a deteriorating-decelerating trend in baseline to accelerating-improving trend during intervention" (p. 230).

	Condition A	Condition B
Trend Direction	Decelerating	Accelerating
Deteriorating or Improving	Deteriorating	Improving

Step 3.
Review Step 6, specifically if data were variable or stable by condition.

- Would summarize stability across conditions as follows:
  - "A variable decelerating trend in a contra-therapeutic direction where, upon introduction of intervention, there was a change in trend direction to a variable accelerating trend in a therapeutic direction" (p. 230).

Step 4a.
Relative level change:

- Subtract the median value from the second half of baseline (A) from median value from the first half of intervention (B).

	Level Change A → B
Median 1st half of intervention (B)	56.5
Median 2nd half of baseline (A)	12
Relative level change	+44.5 Improving

Step 4b.
Absolute level change:

- Subtract the last value of condition A from the first value of condition B.

	Level Change A → B
First value of intervention (B)	39
Last value of baseline (A)	12
Absolute level change	+27 Improving

Figure 8. Between-condition analysis: Steps 1–4b.

improvement in target behaviours when compared to performance during baseline condition. The general idea of PND is higher percentages constitute a larger magnitude of change in a therapeutic direction, but PND is not without limitations. A participant may display change in a therapeutic direction during baseline, and upon introduction of an intervention this change continues. Using PND as the only metric would incorrectly identify a change in a therapeutic direction when in reality improvement across conditions was not influenced by the introduction of an intervention (Gast & Spriggs, 2010). For this reason, researchers should use multiple measures of evaluation, recognising the limitations of single approaches for interpretation. Second, each step listed in the "how to" section is not necessary for

Step 4c.
Median level change:
- Subtract the median value of condition A from the median value of condition B.

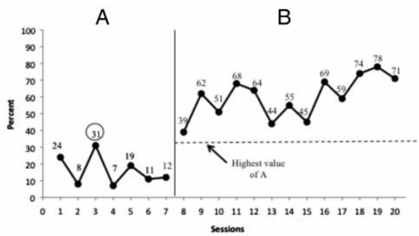
	Level Change A → B
Median value of intervention (B)	62
Median value of baseline (A)	12
Median level change	+50 Improving

Step 4d.
Mean level change:
- Subtract the mean value of condition A from the mean value of condition B.

	Level Change A → B
Mean value of intervention (B)	59.92
Mean value of baseline (A)	16
Mean level change	+43.92 Improving

Step 5a.
PND
- Mark the highest value of condition A on the graph*.
- PND: count the number of values that are above the highest value of A*.
- PND = Divide the number of values above the line by total number of sessions and multiply by 100 (%).
- PND = 13 / 13 x 100 = 100%

**If goal of intervention is to decrease a behavior, mark the lowest value of condition A and count number of values below the lowest value of condition A.*



Step 5b.
POD
- Mark the highest value of condition A on the graph*.
- POD: count the number of values that are same as or below the highest value of A*.
- POD = Divide the number of values on or below the line by total number of sessions and multiply by 100 (%).
- POD = 0 / 13 x 100 = 0%

**If goal of intervention is to decrease a behavior, mark the lowest value of condition A and count number of values above the lowest value of condition A.*

Figure 9. Between-condition analysis: Steps 4c–5b.

every study. For example, the split-middle method of trend estimation is typically conducted at the end of a study and is less common in the literature. In regard to trend estimation, the split-method is recommended over another trend method, such as ordinary-least squares regression due to issues related to auto-correlation of data and outliers. Ordinary-least squares regression is based on the assumption that data are independent and may also be influenced by outliers. In contrast, the split-middle method of trend estimation does not require independence of data and relies on median values, which are less sensitive to outliers than mean values (Good & Shinn, 1990). Measures included in the visual analysis process should be included for purposes of strengthening an unbiased interpretation of effects of an intervention (Gast & Spriggs, 2010).

A third consideration when using visual analysis with SCED data is the understanding all possible threats to the internal validity of a study. Internal validity refers to control and recognition of confounds during a study that could possibly provide an "alternative explanation of findings" (Gast, 2010, p. 4). Some of these threats have been addressed or will be addressed in more detail, but are listed here for readers consideration: (a) history, (b) maturation, (c) testing, (d) instrumentation (or reliability of measurement of the dependent variable), (e) fidelity of implementation, (f) attrition, (g) multi-treatment interference, (h) variability or instability in data, (i) adaptation, and (j) the Hawthorne effect (Cooper et al., 2007; Gast, 2010; Kennedy, 2005). Finally, researchers should recognise and adhere to quality indicators when conducting SCED studies. Horner et al. (2005) and Kratochwill et al. (2010) provide guidelines for researchers conducting SCED studies. It is recommended that guidelines by Horner et al. (2005) and Kratochwill et al. (2010) be reviewed prior to conducting a SCED study. Some of the key issues to consider are as follows: (a) report fidelity of implementation and reliability of measurement of the dependent variable, (b) with the goal of at least 80% agreement or higher. In addition, (c) collect a minimum of three to five data points for each condition, and (d) demonstrate at least three replications of effect before reporting a functional relation.

Reliability of findings and fidelity of implementation

Confidence in results is mediated by reliability of measurement of the dependent variable and fidelity of implementation of procedures during pre-intervention and intervention conditions, as well as any subsequent or intermediate conditions (e.g., maintenance, generalisation probe) that may occur during the study (Kennedy, 2005; Wolery, 2011). While there is not a rule regarding number or percent of sessions during which reliability and fidelity data should be collected, the general expectation and recommendation is at least 20% of conditions for each participant by at least one independent observer (i.e., person who is not implementing the intervention; Kazdin, 2011; Kennedy, 2005). The general idea regarding collection of reliability and fidelity data is "more is better". Baer, Wolf, and Risley (1968) emphasise the importance of detailed behavioural definitions and technological descriptions of behaviour-change programmes. In their seminal work on applied behaviour analysis (ABA), Baer et al. (1968) indicate well-written procedures are written as such a "[novice] reader could replicate…procedure[s] well enough to produce the same results" (p. 95). They continue, "explicit measurement of the reliability of human observers...becomes not merely good technique, but a prime criterion of whether the study was appropriately behavioural" (p. 93).

Lack of agreement regarding measurement of the dependent variable negates the results of visual analysis since disagreements exist about occurrence and/or non-occurrence of the target behaviour (Gast, 2005). For example, a researcher interested in decreasing perseverative speech of an adolescent with traumatic brain injury implements an intervention across environments (i.e., home, school, vocational training site), replicating effects of decreasing the target behaviour to a rate of 1 occurrence per two-hour observation for the final five intervention sessions. Another researcher collects data 20% of sessions across the study across environments and agrees with primary researchers' observations of the target behaviour on the average of 12% (e.g., both observers agreed that the behaviour occurred or did not occur only 12/100 times). This lack of agreement reduces the likelihood of replication since it is unclear what, if any, behaviour changed during the study. This uncertainty lessens trust that a proposed intervention is an appropriate choice for persons with similar pre-intervention behaviours.

Human error is expected during observation, and while exact agreement is ideal, it is not a "rule" that observers agree 100% across all observations. It should be stressed that researchers strive for agreement between 80 and 100%. While an arbitrary figure, agreement below 80% is considered a "red flag" when interpreting results of a study (Kennedy, 2005). Multiple factors may be responsible for low-levels of agreement and require consideration during design and implementation of a study. Prior to beginning a study, (a) detailed descriptions of target behaviours should be provided to all persons who will collect data and (b) training should occur until agreement is considered acceptable for purposes of the study. Complexity of behavioural descriptions may require researchers to further examine the purpose of the study and potentially create multiple target behaviours or remove unnecessary components of a behaviour-change programme. If, at any point during the study, there is low-level agreement between observers the study should cease until issues are identified and addressed (Ayres & Gast, 2010). Traditionally, agreement is reported as a percentage (and range) across a study by participants or across all participants (Artman, Wolery, & Yoder, 2012). In an effort to address limitations of reporting agreement data as a summative measure, some researchers advocate for graphing these data as a means for visually analysing agreement (or disagreement) between independent observers (Ledford & Wolery [in press]; Ledford, Wolery, Meeker, & Wehby, 2012).

Fidelity of implementation of procedures across conditions is another key factor to consider when interpreting results of a study. In a commentary on the importance of fidelity of measurement, Wolery (2011) describes the "essence of experimental work" as confidence that a behaviour-change programme was responsible for prosocial "shifts" in socially valued behaviours for purposes of improving experiences of participants' across their lifespan (p. 156). Adherence to specific procedural descriptions, in conjunction with data to support appropriate implementation of procedures, gives credence to interpretations of data using visual analysis and allows readers an opportunity to replicate these effects with participants with similar pre-treatment behaviours. It is the responsibility of researchers to report fidelity of implementation of each step of a behaviour-change programme by condition (Gast, 2010). This is especially important for researchers and practitioners attempting to replicate effects of an intervention. For example, a researcher attempts to replicate an intervention designed to improve recall for persons with dementia. In their report they indicate inability to replicate the magnitude of effect of a previous study, but only report percent and range of correct implementation collapsed across all steps. In this example, it is unknown if a step or multiple steps were excluded, thus impacting the magnitude of the behaviour-change programme. Even if the purpose of a study is evaluation of necessity of steps (i.e., addition or omission of steps) in a behaviour-change programme, researchers are urged to report adherence to procedures as written across each proposed component of an intervention. As with reliability of measurement of the dependent variable, detailed descriptions of procedures should be provided to all persons who will implement a study and collect data (Wolery, 1994). Specific criteria (e.g., 100% correct implementation of procedures across three consecutive opportunities) should be used as a guideline for mastery of procedures before beginning a study, and evaluated during a study (Ayres & Gast, 2010).

Statistics and SCED

Application of statistical methods to SCED data has received increased attention in the literature, especially as it relates to calculation of an effect size for interventions (Wolery, Busick, Reichow, & Barton, 2008). Campbell and Herzinger (2010) propose considerations regarding statistical analysis and SCED studies in that statistical tests may (a) add to the confidence of results of visual analysis of data, (b) "quantify strengths of outcomes", and (c) increase objectivity of analysis (pp. 421–422). The first author conducted an informal search of on-line databases (i.e., ERIC, PsycInfo, Education Research Complete, and Medline) for articles on visual analysis (i.e., keywords were *visual analysis, special education, psychology, education, single subject research, single subject research design, single case, single case experimental design, single subject, single subject experimental design*, in peer reviewed journals) and found 42 articles related to visual analysis and SCED. Of the 42 articles, 22 related to application of statistical methods to data from SCED studies with emphasis on calculations of effect size. The purpose of this informal search was to highlight the relatively large number of articles on various methods for calculating effect size, as well as to highlight that an optimal method for calculating effect size is not currently available, which corresponds with Campbell and Herzinger (2010) report that "little consensus exists regarding the appropriate calculation of effect sizes for single case designs" (p. 440). Campbell (2004) highlights the issues of application of statistical methods to SCED, specifically, observations are "usually not independent" (i.e., auto-correlated) and trend of data may "confound" results of effect size calculations (p. 235). Analysis of SCED data using statistical methods is a controversial topic that currently lacks a clear answer for researchers, but efforts are ongoing to identify appropriate procedures that summarise effects of multiple articles (Campbell, 2004; Wolery et al., 2008).

Conclusion

As highlighted by Cooper et al. (2007), visual analysis of SCED data answers two questions: "(1) Did behaviour change in a meaningful way, and (2) if so, to what extent can that change in behaviour be attributed to the independent variable" (p. 149). Baer et al. (1968) emphasise the importance and meaning of "applied" research when designing behaviour-change programmes in that an intervention should attempt to change behaviours which benefit individual participants. The primary goal of visual analysis is to identify if a functional relation exists between the introduction of an intervention and change in a socially desirable behaviour, as well as replicate effects across multiple participants. Visual analysis is sensitive to changes in behaviour and allows researchers to analyse each participant's behaviour through repeated measurement and evaluation, allowing observation of abrupt, as well as subtle changes over time. Challenges of visual analysis should also be considered when conducting a study. First, as seen through multiple attempts to apply statistical methods to SCED data, it can be difficult to summarise the effects of interventions across participants since various behaviours are measured, and individual modifications to an intervention may be made during a study. A second consideration is generality of findings to participants outside of studies. The issue of generality relies on researchers providing detailed descriptions of participant's pre-intervention behaviours to increase the likelihood of understanding for whom and under what conditions interventions may be effective. Third, researchers should evaluate agreement for results of visual analysis conducted by independent observers. Previous studies have found agreement can vary across persons for various reasons (Ottenbacher, 1993), but adherence to detailed procedures, as presented in this article, and familiarity of methods for increasing agreement (e.g., training) address possible variability that may arise during visual analysis of identical graphs across persons (Ledford et al., 2012). Finally, researchers using a comparison design (e.g., alternating treatments design) should be aware of specific considerations for conducting visual analysis of graphs. For more detail on comparison designs and visual analysis review Wolery, Gast, and Hammond (2010).

The purpose of this paper is to introduce readers to visual analysis of graphic displays of data collected during SCED studies. Visual analysis involves evaluation of performance within and across conditions using systematic procedures. Training using the above procedures should occur with multiple graphical displays prior to conducting visual analysis of data for publication. Agreement data should also be collected across persons training to use visual analysis. The guidelines and considerations presented should assist researchers in objectively evaluating behaviour-change programmes for individual participants or groups of participants across settings, behaviours, and participants. Confidence in findings is furthered by replication of effect using detailed procedures and understanding of results in regard to agreement of occurrence and non-occurrence of behaviours, as well as understanding implementation of procedures as designed. Visual analysis of data within a SCED framework offers researchers alternatives to understanding effect of behaviour-change programmes outside of special education and related fields.

REFERENCES

Artman, K., Wolery, M., & Yoder, P. (2012). Embracing our visual inspection and analysis tradition: Graphing interobserver agreement data. *Remedial and Special Education*, *33,* 71–77.

Ayres, K., & Gast, D. L. (2010). Dependent measures and measurement procedures. In D. L. Gast (Ed.), *Single subject research methodology in behavioral sciences* (pp. 129–165). New York, NY: Routledge.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *1,* 91–97.

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

Campbell, J. M., & Herzinger, C. V. (2010). Statistics and single subject research methodology. In D. L. Gast (Ed.), *Single subject research methodology in behavioral sciences* (pp. 91–109). New York, NY: Routledge.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Columbus, OH: Pearson.

Gast, D. L. (2005). Visual analysis of graphic data. In G. Sugai & R. Horner (Eds.), *Encyclopedia of behavior modification and cognitive behavior therapy: Educational applications* (Vol. 3, pp. 1595–1599). Thousand Oaks, CA: Sage.

Gast, D. L. (Ed.). (2010). General factors in measurement and evaluation. *Single subject research methodology in behavioral sciences* (pp. 91–109). New York, NY: Routledge.

Gast, D. L., & Hammond, D. (2010). Withdrawal and reversal designs. In D. L. Gast (Ed.), *Single subject research methodology in behavioral sciences* (pp. 234–275). New York, NY: Routledge.

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

Good, R. H., & Shinn, M. R. (1990). Forecasting accuracy of slope estimates for reading curriculum-based measurement: Empirical evidence. *Behavioral Assessment*, *12,* 179–193.

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

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

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

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: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Lane, K., Wolery, M., Reichow, B., & Rogers, L. (2007). Describing baseline conditions: Suggestions for study reports. *Journal of Behavioral Education*, *16,* 224–234.

Ledford, J. R., & Wolery, M. (in press). Effects of plotting a second observer's data on A-B-A-B graphs when observer disagreement is present. *Journal of Behavioral Education*.

Ledford, J. R., Wolery, M., Meeker, K. A., & Wehby, J. H. (2012). The effects of graphing a second observer's data on judgments of functional relations in A-B-A-B graphs. *Journal of Behavioral Education*, *21,* 350–364.

Lieberman, R. G., Yoder, P. J., Reichow, B., & Wolery, M. (2010). Visual analysis of multiple baseline across participants' graphs when change is delayed. *School Psychology Quarterly*, *25,* 28–44.

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children*, *71,* 137–148.

Ottenbacher, K. J. (1993). Interrater agreement of visual analysis in single-subject decisions: Quantitative review and analysis. *American Journal of Mental Retardation*, *98,* 135–142.

Reichow, B., Volkmar, F. R., & Cicchetti, D. V. (2008). Development of the evaluative method for evaluating and determining evidence-based practices in autism. *Journal of Autism and Developmental Disorders*, *38,* 1311–1319.

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

Spriggs, A. D., & Gast, D. L. (2010). Visual representation of data. In D. L. Gast (Ed.), *Single subject research methodology in behavioral sciences* (pp. 166–198). New York, NY: Routledge.

White, O. R. (1972). *The split-middle: A quickie method of trend analysis*. Eugene, OR: Regional Center for Handicapped Children.

Wolery, M. (1994). Procedural fidelity: A reminder of its functions. *Journal of Behavioral Education*, *4,* 381–386.

Wolery, M. (2011). Intervention research: The importance of fidelity measurement. *Topics in Early Childhood Special Education*, *31,* 155–157.

Wolery, M., Busick, M., Reichow, B., & Barton, E. E. (2008). Comparison of overlap methods for quantitatively synthesizing single-subject data. *Journal of Special Education*, *44,* 18–28.

Wolery, M., Gast, D. L., & Hammond, D. (2010). Comparative intervention designs. In D. L. Gast (Ed.), *Single subject research methodology in behavioral sciences* (pp. 329–381). New York, NY: Routledge.

Wolery, M., & Harris, S. R. (1982). Interpreting results of single-subject research designs. *Physical Therapy*, *62,* 445–452.