Educ. Treat. Child. (2021) 44:275–290
https://doi.org/10.1007/s43494-021-00053-3

ORIGINAL ARTICLE

Standards of Graph Construction in Special Education Research: A Review of their Use and Relevance

Richard M. Kubina Jr · Douglas E. Kostewicz · Seth A. King · Kaitlyn M. Brennan · Jennifer Wertalik · Karen Rizzo · Andy Markelz

Accepted: 2 August 2021 / Published online: 20 August 2021

Abstract Much of the special education literature features single-case experimental designs, which traditionally require researchers to determine functional relations among variables through the visual analyses of line graphs. Evidence suggests factors aside from the data influence visual analysis, including line graph construction. Fields including engineering, behavior analysis, and psychology have historically propagated standards related to the visual data displays to mitigate the effect of arbitrary graph construction on the interpretation of results. Although evidence suggests graphs featured in behavior analytic studies do not observe the standards, the extent to which researchers in special education adhere to longstanding graphing guidelines remains uncertain. The following article provides an overview of graphing standards and examines the adherence of line graphs from 532 issues of 28 distinct special education journals to traditional standards of visual display. Results indicated the majority of special education line graphs deviate from established line graph construction standards in important respects. The discussion centers on the need for updating and disseminating guidelines regarding line graph construction.

Keywords Line graphs · Single-case · Visual analysis · Graph construction · Graphing standards

The related disciplines of special education and applied behavior analysis (ABA) have traditionally relied on graphic displays of data to inform treatment decisions and evaluate interventions (Nelson et al., 2016). Graphical display represents a key element of single-case experimental designs (SCED), which evaluate the effect of interventions on a single unit of analysis over time (e.g., participant; Horner et al., 2005). In general, the presentation of SCED data involves a linear graph in

---

R. M. Kubina Jr, (✉)
Pennsylvania State University, Special Education Program, 209 CEDAR Building, University Park, PA 16802-3109, USA
e-mail: rmk11@psu.edu

D. E. Kostewicz · K. M. Brennan
University of Pittsburgh, Pittsburgh, PA, USA

S. A. King
University of Iowa, Iowa City, IA, USA

J. Wertalik
Georgia Southern University, Statesboro, GA, USA

K. Rizzo
Pennsylvania State University-Behrend, Erie, PA, USA

A. Markelz
Ball State University, Muncie, IN, USA

which the vertical axis (i.e., y-axis) depicts changes in an outcome and the horizontal axis (i.e., x-axis) depicts a unit of time. The detection and determination of experimental effects and functional relations occurs visually, without the assistance of formal significance tests (Kazdin, 2011). Thus, graphical displays complement the needs of clinicians and allow for the identification of effective practice in fields serving small populations (Wendel et al., 2015).

An emerging line of research suggests presentation characteristics may skew the visual analysis of graphically depicted data. Chart size and reduction in chart height negatively affect graphical perception and estimation of effects of line charts (Heer et al., 2009). A survey of experts in visual analysis ($n = 32$) conducted by Dart and Radley (2017) revealed that y-axes consisting of truncated percentage scales (e.g., 0–40%) resulted in the perception of larger treatment effects. A similar study (Radley et al., 2018) involving expert raters ($n = 29$) suggested that the number of data points per x- to y-axis ratio (DPPXYR), obtained by dividing the horizontal axis length by the vertical axis length and dividing the quotient by the number of data points permissible within the horizontal axis, also influences visual analysis. Graphs in which (1) the two axes approached a similar length, (2) large numbers of data points appeared on the horizontal axis, or (3) the horizontal axis included additional space for ungraphed data tended to result in the identification of larger treatment effects.

Standards and conventions of graphing have proliferated well in advance of research, affirming the influence of display characteristics on visual analysis (Kubina et al., 2017). Increased acknowledgement of the importance of graphic display has the potential to rekindle interest in existing standards of graphic display or encourage a transition toward new standards. As evidence continues to emerge, the rationale for existing standards and the extent to which such considerations need to appear in revised standards will remain compelling as well as the extent to which researchers in the field observed previous standards. The current article describes the standards of graphical display for the behavioral sciences that continue to appear in special education research texts and reviews the rationale for their use. We then identify problems with standards and potential revisions based on current research. Finally, the researchers systematically replicate a review of the adherence of graphing standards (e.g., Kubina et al., 2017) in special education.

Line Graph Construction Standards

Guidelines from the formative period of modern behavior analysis continue to govern multiple aspects of SCED (e.g., Sidman, 1960), including graphic display. A rich history detailing construction principles began to appear in books and articles in the early 1900s. Comprehensive books written by business leaders, engineers, and statisticians offered design rules (e.g., Brinton, 1914; Karsten, 1923). At the same time, a unified, authoritative source for line graph construction first took form in 1915 with the Joint Committee on Standards for Graphic Presentation (JCSGP). The JCSGP had the expressed purpose of studying methods used through diverse fields (e.g., psychology, mechanical engineering, political science) to establish standards that would advance the speed and accuracy of graphic communication (JCSGP, 1915). In 1936, a set of more detailed codes of practice emerged (Sectional Committee on Standards for Graphic Presentation [SCSGP], 1936). The graphic codes received a minor revision in 1938 (Committee on Standards for Graphic Presentation [CSGP], 1938). The next revision occurred in 1960, American Standard's *Time-Series Charts* (American Society of Mechanical Engineers, 1960). The final set of standards appeared in 1979 (American National Standards Committee Y15 [ANSC Y15.2], 1979) helping professionals prepare and use effective time-series charts such as the line graph.

All five versions of the aforementioned practices and detailed specifications, herein referred to as *Graphing Standards*, serve as a guide for the construction of line graphs. Related guidelines appear in well-regarded SCED textbooks (e.g., Ledford & Gast, 2018) and social and behavioral science style manuals such as the *Publication Manual of the American Psychological Association* (APA). The *Manual* directs graph makers to carefully construct their visual displays so that they can effectively communicate what the researchers have discovered (APA, 2020). Special education textbook authors echo similar sentiments, "The primary function of a graph is to communicate without assistance from the accompanying text" (Spriggs et al., 2018, p. 166).

Kubina et al. (2017) distilled many of the existing standards into general categories: *essential structure* and *quality features* (see Table 1). The *essential structure* of a line graph conveys its foremost purpose: examining values changing through time (Few, 2009). The slope of the line displayed across days, weeks, or months establishes prediction of behavior (Riley-Tillman & Burns,

Table 1 Scored graphing features for a line graph

Essential structure and function	Measured
Vertical and horizontal axes maintain a quantitative measure and time unit label, respectively, to show change in the measure over time (Harris, 1999)	Noted the labels on the vertical and horizontal axes.
Quality feature and function	Measured
Vertical axis length maintains a 2:3 or 3:4 ratio in relation to horizontal axis which limits data distortion (Cooper et al., 2020; Parsonson & Baer, 1978)	• Per graph, does the ratio of vertical to horizontal axes lengths fall between 5:8 to 3:4 (63% to 75%)? • Do all graphs on a page within the same figure align? • Do all graphs within the same article that maintain the same unit share the same physical length? • Are all graphs with the same unit on either axis scaled to the same minimum and maximum?
A minimal number of evenly spaced tick marks point outward to prevent graph clutter and confusion, while highlighting the data (Cleveland, 1994)*	• On both axes, are tick marks pointing outward for the entire length of the axes? • On both axes, are tick marks evenly spaced (i.e., at equal intervals)?
Tick marks have labels to show unit value (Robbins, 2005)	• On both axes, are tick marks numbered? • Are the scale counts correct?
The figure maintains clearly visible data points and paths to promoting clarity and data direction (Cooper et al., 2020; Robbins, 2005)	• Are data points on the figure clearly visible? • If the figure contains a data path, is it visible?
Labeling of condition change lines separates data into experimental conditions (Cooper et al., 2020)**	• If the figure contains a condition change line, is it visible? • If the figure contains a condition change line, does it have labels?
A figure caption conveys meaning (Cooper et al., 2020)	• Does the figure contain a caption?

Note. * = Measured based on Cleveland (1994), ** = Condition lines and labels not covered in Standards, but clearly used in single-case research and measurement based on Cooper et al. (2020).

2009). Thus, the *Graphing Standards* suggest line graphs have labeled axes that correspond with the dependent variable (vertical axis) and actual units of time (horizontal axis). *Quality features* encompass a range of graphical elements, such as the visibility of data points and the placement of tick marks. The guidelines attempt to ensure that features of the data themselves do not lead to misinterpretations due to vagaries of shading or ambiguous rendering of the scale.

Discerning an effect visually relates to a variety of quality features (Kubina et al., 2017). Scaling the vertical axis to different terminal values, for example, influences the ability of analysts to determine the magnitude of effects (Dart & Radley, 2017). As early as 1914, the dawning of line graphs, warnings appeared around scaling practices: “In general, it is unwise to compare the shapes of two curves unless they are plotted to the same scales, both horizontal and vertical” (Brinton, 1914, p. 79). The *equivalent scaling* rule attempts to mitigate the perception altering effects of improper terminal axis scaling between related graphs. Related graphs refer to figures within the same research article that share one or both axis labels. Figure 1 illustrates the equivalent scaling rule. Each shares unit labels for both axes with different maximum amounts. Level (top), trend (middle), and trend stability lines (bottom) appear in the second column corresponding to the data directly to the left. The third column shows column one data when applying the equivalent scaling rule. Scaling to each unit’s terminal amount across the three data sets (100 for the Quantitative Value and 32 for the Unit of Time) allows for commensurate visual comparisons. Applying the rule, however, produces clear changes to level, trend, and trend stability not seen in the first data sets.

“Single-case research involves the fine-grained analysis of change across time” (Horner & Spaulding, 2010, p. 1388). Thus, the standards place additional emphasis on the treatment of time in graphical displays. Clear, transparent depictions of time allow researchers to forecast any future changes or place behavioral changes in a meaningful context (Johnston et al., 2020). The use of opaque or imprecise ordinal units of time potentially impedes time series data analysis through issues of false equality and nonrepresentative data (Kubina et al., 2017). *False equality* occurs when observational measures vary throughout the study but appear as equal sessions on the graph.

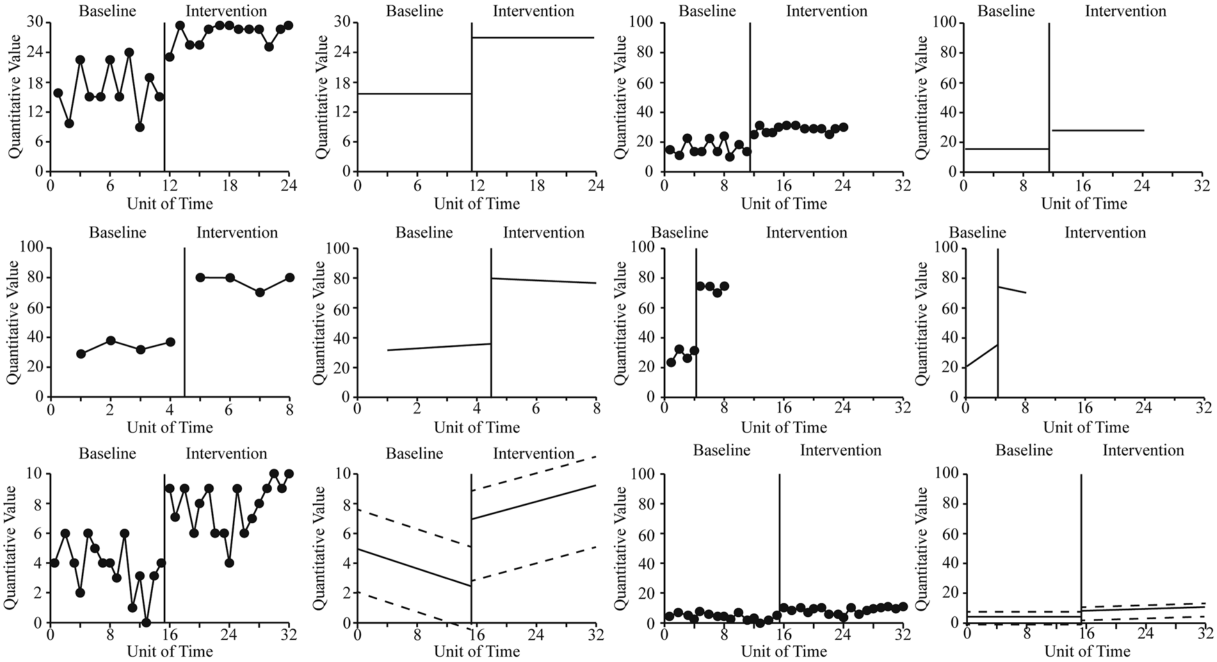

Fig. 1 Line graphs illustrating the equivalent scaling rule

Nonrepresentative data arises from incongruities between when sessions occur in relation to other sessions in time. For example, data collected sporadically and labeled as sessions misleadingly appear as contiguously plotted points (e.g., sessions 1–20).

Of the rules pertaining to essential structure, the guidelines related to the ideal proportion of the vertical and horizontal axes remain the most well-known. The *proportional construction rule* maintains that the ratio of the vertical to horizontal should fall between 5:8 to 3:4 (e.g., American National Standards Institute & American Society of Mechanical Engineers, 1960, American National Standards Committee Y15 [ANSC Y15.2], 1979; Parsonson & Baer, 1978; Poling et al., 1995; Cooper et al., 2020). The rule stems from the fact that changing the axes' proportions physically changes the position of the data. Slope judgment interpretations will unquestionably vary even though data remain the same. The slope of a trend, for instance, could adjust from 45° to 15° or 75°. The more divergent the slope, the greater probability interpretations will result in Type 1 or Type 2 errors.

Relevance of Graphing Standards

Despite the stated rationale for *Graphing Standards* (Cleveland, 1994; Cooper et al., 2020; Kubina et al., 2017), their relevance to current practice remains unclear. A survey of applied and experimental behavior analysis journals ($n = 11$; Kubina et al., 2017) found that in general graphs did not adhere to the *Graphing Standards*. For each journal, Kubina et al. randomly selected a single issue every 2 years from the start of the journal's publication to 2011. The authors identified 4,313 graphs across 191 issues and scored each graph based on their accordance with essential structure and quality features. Results suggest that, on average, authors adhered to standards regarding the use of captions, labels, and clear data paths (91%). Average observance of features related to the use of tick marks and consistent scale counts, though considerably lower (66%), suggest the majority of graphs in behavior analysis reflect standards concerning the features. The authors identified far lower levels of adherence to guidelines concerning the use of real units of time on the horizontal axis (31%), equivalent scaling (31%), and proportional construction (15%). Ledford et al. (2019) likewise found that articles published in 12 top special education journals ($n = 12$) over a 5-year period generally did not meet recommended standards for graphic display. Results of Peltier, McKenna, et al.'s (2021a); Peltier, Morano, et al.'s (2021b) reviews of special education journals ($n = 6$) from 2010 to 2019 suggest only 3% of graphs met proportional construction guidelines.

Some researchers have questioned the validity of specific *Graphing Standards* based on limited research conducted in this area (Dart & Radley, 2018). Several authors have noted that no evidence exists to support a connection between the use of real units of time on the horizontal axis and the interpretation of findings (Ledford et al., 2019; Shadish et al., 2015). Other researchers have provided evidence to support various alternatives to specific standards. Radley et al. (2018) suggest the proportional construction guidelines may contravene their intended purpose (e.g., result in false positives) and argue instead for the design specific DPPXYR as an alternative. Other researchers have likewise suggested using certain operations to drive graph construction. For example, Cleveland et al. (1988) researched the median-absolute slope criterion in constructing two-variable graphs. The procedure, based on a geometric algorithm, appears to promote accurate slope judgments.

From the *Graphing Standards* to individual research teams (e.g., Cleveland et al., 1988; Radley et al., 2018), there exist a wide range of suggestions for how best to construct line graphs. Despite specific differences, scholarship appears to verify the *Graphing Standards'* central premise: line graph construction has a profound influence on the interpretation of results. The relation between graphic displays and the interpretation of findings has obvious implications for the use of research, because inconsistently displayed SCED data potentially result in the misinterpretation of treatment efficacy data. Yet, the What Works Clearinghouse (WWC) and other standards for special education research (e.g., Cook et al., 2015), though they may have general guidelines related to the presence of line graphs in SCED, do not address line graph construction. Although many in the field continue to insist the visual analysis of graphic displays represents the standard of interpretation for SCED research (e.g., Manolov et al., 2016), the most recent standards of the WWC (2020) have explicitly limited the role of visual analysis in evaluating evidence. The WWC favors effect sizes and quantified results. The decision likely stems from the large body of evidence indicating factors independent of the data, including graphic display (Radley et al., 2018), may influence the results of visual analysis (Ninci et al., 2015).

Purpose

The degree to which line graphs in special education journals adhere to the *Graphing Standards* remains unclear. Regardless of whether one accepts the *Graphing Standards*, the observance of specific standards in the field provides information regarding how best to proceed in addressing issues related to graph construction. Adherence to the *Graphing Standards* would signify consensus regarding proper graph construction. However, such an accord would also require reeducation efforts in the event additional evidence demonstrates the unsuitability of previous conventions. On the other hand, limited adherence to established conventions may warrant the prioritization of SCED line graph construction standards by the WWC or other organizations. The current evaluation stems from agreement with the premise that graph construction remains a critical element of behavior analysis and our interest in the implications of adherence to the *Graphing Standards* for the field more generally. Therefore, the present review extends the methods featured in Kubina et al. (2017) in service of one primary question: To what extent do selected visual graphics adhere to the *essential structure* and *quality features* of line graph construction within special education journals?

Method

Journal Identification

The systematic review procedures appear in Fig. 2. The following section describes the review methods in greater detail. The current search first identified journal topic areas. The 13 special education categories (Individuals with Disabilities Education Act [IDEA], 2004), the Council for Exceptional Children's (CEC) Special Interest Divisions, and a previous study focused on special education journals (Kubina et al., 2010) formed the basis for creating 13 areas related to specific disability categories or topics of interest (Table 2).

Journal titles populating each of the areas came from one of two sources: CEC's website or the 2013 Web of Science (WOS) Journal Citation Reports® Social Sciences Edition (, 2014). CEC journals encompass the spectrum of materials available to special education practitioners and researchers. Eligible journals fit into one of the identified areas based on the focus of the publication's Division or through a review of the journal's mission statement. Excluding newsletters and magazines, 12 journals (Table 2, labeled CEC) met

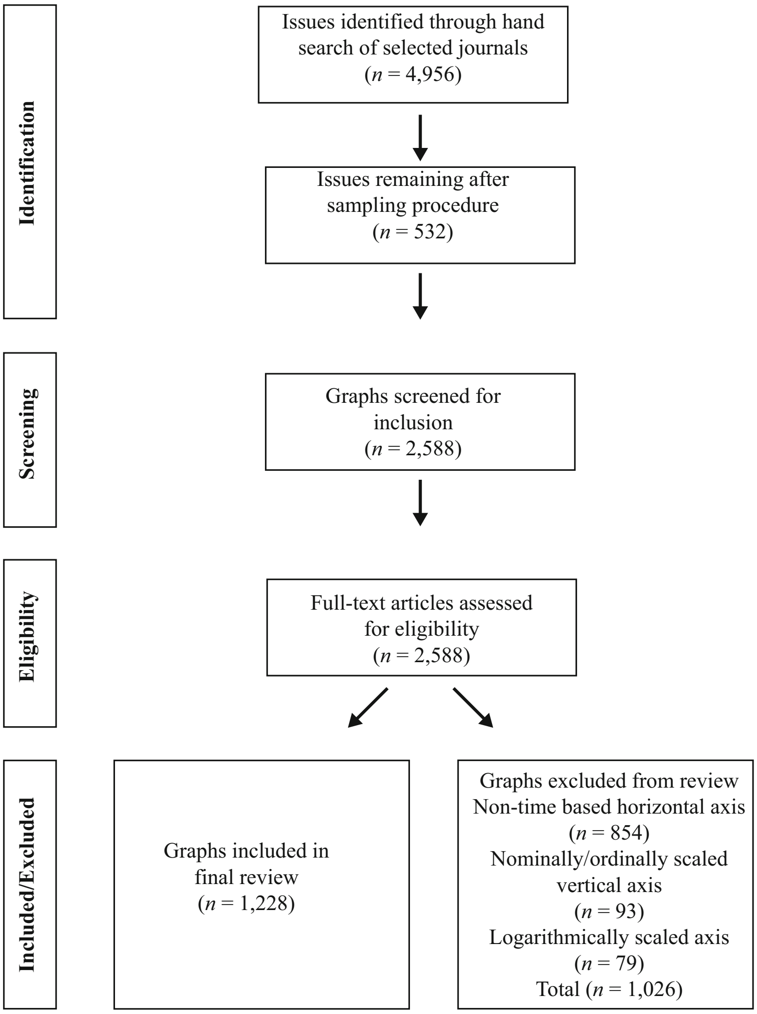

Fig. 2 A PRISMA flow diagram for the review

criteria. The JCR (WOS, 2014) served as the basis for the remaining journals.

The WOS Special Education Category (Web of Science (WOS), 2014) ranks special education journals by impact factor. An algorithm based on citation and consumption patterns serves as the basis for WOS subject categories (Wang & Waltman, 2016), which comprise the most widely accepted journal subject classification system in the world (Waltman, 2016). The current researchers reviewed the mission statements from all ranked special education journals (i.e., 37). Eligible journals had mission statements located within the journal or on its homepage with keywords matching 1 of the 13 identified areas. For areas with more than two titles identified from the JCR process, only the top two ranked by impact factor met criteria. Otherwise, all JCR identified journals met criteria. The JCR group increased the total number of journals by 15 (Table 2, labeled JCR).

Following the two-step process, no titles sorted into the Traumatic Brain Injury (TBI) area. After reviewing the mission statements of journals on JCR (Web of Science (WOS), 2014) Rehabilitation list, one journal, the *Journal of Head Trauma Rehabilitation*, met inclusion criteria (Table 2, labeled JCR*). In total, 28 journals across 13 areas met criteria for review.

Issue Selection

Issue selection occurred following journal identification. The researchers examined the publication history of each of the 28 journals from 1975, the generally

Table 2 Journals meeting criteria for inclusion in the study

Journal area	Journal title	Location	Pop. Issues/%	Sample Issues/%	Weight
Autism	Education and Training in Autism and Developmental Disabilities	JCR/CEC	176/3.77	22/4.37	.86
		CEC	140/3	17/3.37	.89
	Focus on Autism and Other Developmental Disabilities	JCR	79/1.69	6/1.19	1.42
	Research in Autism Spectrum Disorders				
Early Intervention	Journal of Early Intervention	CEC	134/2.87	19/3.77	.76
	Infants and Young Children	JCR	122/2.6	18/3.57	.73
Emotional Disturbance	Behavior Disorders	CEC	171/3.66	22/4.37	.84
	Journal of Emotional and Behavioral Disorders	JCR	104/2.23	13/2.58	.86
Gifted	Journal for the Education of the Gifted	CEC	144./3.08	18/3.57	.86
	Gifted Child Quarterly	JCR	176/3.77	22/4.37	.86
	High Ability Studies	JCR			
Hearing Impairments and Deafness	Journal of Deaf Studies and Deaf Education	JCR	100/2.14	12/2.38	.90
	American Annals of the Deaf	JCR	255/5.46	22/4.37	1.25
Learning Disabilities (Specific)	Learning Disabilities Research and Practice	JCR/CEC	132/2.82	14/2.77	1.02
	Journal of Learning Disabilities	JCR	348/7.44	22/4.37	1.71
	Annuals of Dyslexia	JCR	66/1.41	22/4.37	.32
Intellectual Disabilities	Journal of Intellectual Disability Research	JCR	328/7.02	22/4.37	1.61
	American Journal on Intellectual and Developmental Disabilities	JCR	176/3.77	22/4.37	.86
Orthopedic Impairments	Physical Disabilities Education and Related Services	CEC	48/1.02	13/2.58	.40
Speech or Language Impairments	Communication Disorders	JCR/CEC	244/5.22	22/4.37	1.20
	Child Language Teaching and Therapy	JCR	102/2.18	17/3.37	.65
	Journal of Fluency Disorders	JCR	180/3.85	22/4.37	.88
Traumatic Brain Injury	Journal of Head Trauma Rehabilitation	JCR*	180/3.85	17/3.37	1.14
Visual Impairments and Blindness	Visual Impairments	CEC	425/9.09	21/4.17	2.18
Multi-category	Exceptional Children	JCR/CEC	232/4.96	22/4.37	1.14
	Journal of Special Education	JCR/CEC	176/3.77	22/4.37	.86
	Research in Developmental Disabilities	JCR	212/4.54	19/3.77	1.20
	Topics in Early Childhood Special Education	JCR	152/3.25	19.3.77	.86
Deafblind	Deafblind Education Quarterly	CEC	16/.34	4/.79	.43

Note. CEC = Journal associated with a division of CEC; JCR = Journal found in Journal Citation Reports category: Special Education; JCR* = Journal found in Journal Citation Reports category: Rehabilitation. Column related to population describe actual number of issues available to sample from each journal and the proportion of issues relative to the entire population of selected journals. Sample column describes number of issues initially selected from review and the proportion of issues relative to the entire sample. Weight represents adjustment used to correct for disparities between population and the sample.

accepted inception of special education, until 2020. In accordance with previous literature surveys (e.g., Kostewicz et al., 2016; Kubina et al., 2010; Kubina et al., 2017), researchers randomly selected one issue every 2 years, for a maximum of 20 issues per title staggered across the journal's publication history. From a total 4,956 possible issues, the researchers selected 532, or 10.73%, of all issues. In survey research involving the estimation of proportions from known populations, researchers generally derive an acceptable sample size using Cochran's (1977) formula, where z accounts for the standardized confidence interval (CI; 95% CI = 1.96), p the proportion of a population expected to exhibit a specific response (set to .5 to provide the most conservative sample size requirement; Blair & Blair, 2015), e the margin of error, and N the population size:

$$Sample = \frac{\left[z^2 * p(1-p)/e^2\right]}{\left[1 + \left(z^2 * p(1-p)/(e^2 N)\right]\right.}$$

Assuming a confidence level of 95%, the sample size selected (n = 532) constituted a representative sample with an acceptable margin of error of approximately +4.1% (i.e., .041; Daniel & Cross, 2018).

Due to the issue selection method employed, the proportion of issues from each journal included in the sample approximated the proportion of issues actually published by each journal since 1975. However, we calculated weights for each journal to provide a more accurate estimate of graphing features appearing in the full range of published issues (Groves et al., 2009). Balancing entailed dividing the percentage of issues from each journal published by the percentage of issues from each journal featured in the sample. Following coding, we applied weights to all descriptive statistics. The actual number of issues per journal, issues included in the sample, and weighting variables appear in Table 2.

Graph Inclusion Criteria, Coding, and Analysis

Doctoral level professionals in behavior analysis and special education with experience in the use of visual analysis applied codes derived from Kubina et al. (2017). Scorers first examined each page within each of the identified issues for any graphs that maintained a horizontal and vertical axis with data moving left to right and then applied the inclusion criteria. A simple line graph meeting inclusion criteria had: (1) quantitative (i.e., interval or ratio) scaling on a single vertical axis, (2) a label on the horizontal axis of either a unit of time or sessions or trials, and (3) a maximum of 1 data point per data path on each horizontal interval. Graphs excluded from the sample that did not resemble line graphs included scatterplots and bar charts. As in Kubina et al. (2017), coders also eliminated graphs from consideration (n = 1,026) if the graphs displayed features inconsistent with time-series graphical display or that otherwise may have obligated authors to adopt unique graphing practices inconsistent with the quality standards. In particular, display excluded following initial identification exhibited: (1) nominally and/or ordinally scaled vertical axes (n = 93), (2) nontime-based horizontal axes (n = 854) like Miles or letters exclusively following a sequence (e.g., A, B, C, D, E . . .), or (3) dually or logarithmically scaled vertical axes (n = 79).

The first column on Table 1 displays the *essential structure* and *quality feature* components derived from the *Graphing Standards* and a rationale/function. The second column displays the questions scored for each line graph meeting criteria. All *essential structure* and *quality features* appear within the *Graphing Standards* except for the two noted with asterisks on Table 1. Guidance from Cleveland (1994) promotes tick marks facing outward rather than inward. In addition, the *Graphing Standards* do not address condition change lines and labels. Given that this practice clearly occurs on line graphs within single-case research, scoring of the features followed instructions from Cooper et al. (2020).

Using *Excel* spreadsheets, scorers noted labels for both axes per graph and within each article and a variety of questions regarding tick marks and axis scaling. Scorers used rulers to determine the length of each axis to define the length ratio and as straight edges for multiple graph alignment. Researchers continued to score the presence/absence of each of the remaining graphical components (e.g., data point visibility, figure caption, data path visibility). *Excel* calculated all descriptive statistics.

Scorer Calibration, Reliability, and Interobserver Agreement

A research team led by three doctoral-level behavior analysts with at least 5 years of experience in visual analysis and systematic literature reviews conducted all procedures. Additional scorers either held certification as a behavior analyst or had extensive experience using graphs as classroom teachers. During training, researchers reviewed all categories with scorers. Scorers then participated in a model-lead-test approach to scoring four graphs conducted by lead researchers. Finally, scorers evaluated a random graph and compared scores to those of a lead researcher. Scorers entered independent scoring upon reaching 100% exact agreement of a graph with a lead researcher.

Two measurement assessment techniques evaluated scoring: Reliability (i.e., intraobserver agreement) and interobserver agreement (IOA). For both reliability and IOA, a point-by-point approach (Cooper et al., 2020) determined the percent of reliability between both *Excel* sheets on each field for an identical issue. To determine reliability, each scorer rescored 20% of issues to establish a personal reliability score (Kostewicz et al., 2016). The average reliability across scorers totaled 93% with a range of 89%–100%. Both a second and third scorer assessed IOA by rescoring graphs from 20% of issues. Average IOA across both scorers totaled 93% with a range of 91%–100%.

Results

The journal title search identified 28 special education journals meeting criteria (Table 2). A description of the search and selection process appears in Fig. 2. Issue selection from the 28 journals produced 532 issues. The sample contained an average of 18 issues per title with a range of 5–23. After compensating for the disparity between issues published and issues selected for sampling, scorers identified 2,588 possible graphs. Several graphs (1,026) did not meet criteria due to scaling of the vertical (i.e., dual scaled, logarithmically, nominal, or ordinal scaling) or horizontal (i.e., label did not include a unit of time/sessions/trials) axes. Scorers coded the remaining 1,562 time series graphics, with approximately three per issue.

Essential Structure

Figure 3 contains two dot charts that display the labeling for each axis. Dots represent categorical instances and appear greatest to least from top to bottom. Although all graphs measured had a drawn vertical and horizontal axis, labeling of those axes differed. Within the 1,562 graphs, the vast majority of vertical labels appeared in three of the nine categories (Fig. 3). Percent (44%), count (i.e., behaviors observed per session; 28%), and frequency/rate (i.e., count per unit time; 17%) comprised approximately 89% of the total. Ratio and duration account for 6% of the labels with the remaining four variables rounding out the final 5%. Horizontal labels, on the other hand, had one category dominate. Sessions/trials accounted for 73%. Days made up approximately 13% of the remaining labels. The final 14% consisted of no label, other, minutes, and seconds.

Quality Features

Table 3 refers to numerous quality features of line graph construction. Researchers scored all possible categories for each graph. A maximum of 1,562 opportunities occurred in most categories. Some graphs or groups of graphs did not contain an opportunity to score certain features, which explains opportunities less than 1,562 (see Table 3). Researchers divided the total number of graphs successfully meeting each category by the total opportunities for that category.

Tick Marks and Scaling

An average of 88% and 74% adherence appeared across the vertical axis and horizontal axis categories, respectively (Table 3). The lowest adherence appeared in two categories on the horizontal axis: data (62%) and numbers (62%) on tick marks. Instead of data and numbers

Fig. 3 Dot charts showing vertical and horizontal axis categorical labels

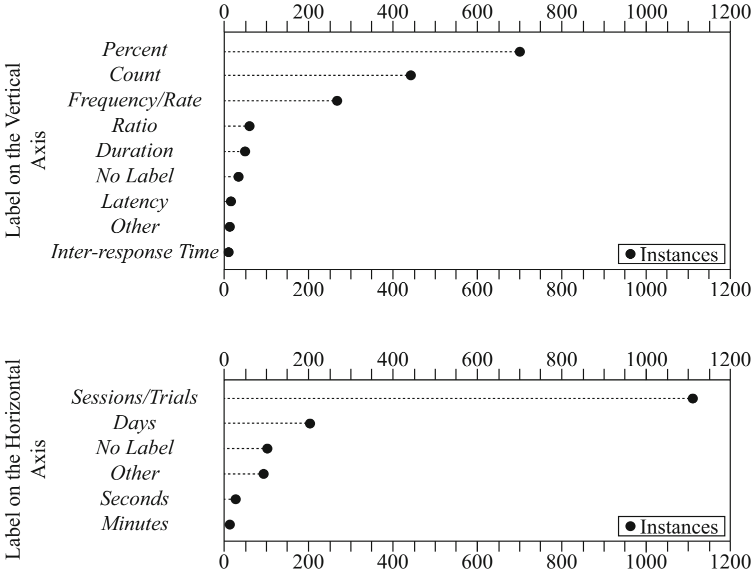

Table 3 Quality features: individual and comparisons of individual graph axes and axes on multiple graphs

Quality feature			Graphs meeting standard
Vertical Axis	Data Occur on Tick Marks		1,156/1,433 = 80%
	Full or Partial Tick Marks on Outside of Graph		1,135/1,433 = 79%
	Tick Marks Occur at Equal Intervals		1,295/1,1433 = 90%
	Numbers Occur on Tick Marks		1,310/1,433 = 91%
	Scale Count is Correct (e.g., 10, 20, 30)		1,394/1,433 = 97%
Horizontal Axis	Data Occur on Tick Marks		892/1,433 = 62%
	Full or Partial Tick Marks on Outside of Graph		1063/1,433 = 74%
	Tick Marks Occur at Equal Intervals		1,229/1,433 = 86%
	Numbers Occur on Tick Marks		878/1,433 = 61%
	Scale Count is Correct (e.g., 10, 20, 30)		1,252/1,433 = 87%
Data Points Clearly Visible			1,328/1,433 = 93%
Data Connected and Data Path Clearly Visible			1,392/1,392 = 100%
Figure Caption			1,1414/1,433 = 99%
Condition Change Labels Present			1,074/1,124= 96%
Data Path NOT Connected across Condition Change Line when Both Present			1010/1112 = 91%
Ratio of Vertical to Horizontal Axis Length: 5:8–3:4 (63%–75% difference)			185/1,433 = 13%
For multiple graphs within the same figure on the same page:	Stacked Vertically (Left/Right Sides of Graphs Align)		284/429 = 66%
	Side-By-Side (Top/Bottom of Graphs Align)		61/106 = 57%
For multiple graphs within the same article;	For vertical axes that share the same label;	Scaled to the same unit (min and max)	168/373 = 45%
		Drawn to the same physical length	138/373 = 37%
	For horizontal axes that share the same label;	Scaled to the same unit (min and max)	161/337 = 47%
		Drawn to the same physical length	154/337 = 46%

aligning with tick marks, authors placed both between visible tick marks.

Data Points and Paths, Condition Labels, and Figure Captions

The five categories encompassing data points/paths and condition/figure labels maintained a high average adherence percentage (96%) as compared to tick marks and scaling (Table 3). Ninety-two percent of figures did not connect data across condition change lines, and 98% and 96% of figures featured caption or condition change labels, respectively.

Axes and Axes Comparisons

Few reviewed figures observed the proportional construction rule (Table 3). In particular, only 12% of the vertical axes measured 5:8 to 3:4 (i.e., 63%–75%) of the length of the attached horizontal axis. On the other hand, graphs within the same figure on the same page aligned axes 66% (vertical) and 60% (horizontal) of the time. Scaling and axis length consistency occurred only in an average of 38% of instances for both the vertical and horizontal axes.

Discussion

The graphic display of data represents an important aspect of treatment and research in special education and related fields. The collected *Graphing Standards* constitute a series of expert guidelines designed to prevent graphic features from adversely influencing visual analysis. The current review examined the adherence of line graphs in special education to *essential structure* and *quality feature* standards. Results suggest that most graphs in special education describe the dependent variable along the vertical axis in accordance with the standards; however, horizontal axes often lack a label corresponding with time. The majority of graphs complied with *quality feature* standards, with notable exceptions. In particular, few graphs appeared to adhere to the proportional construction rule. Although the effect of the departures from current standards remains uncertain, they nonetheless have potential implications for special education.

Graphs appearing in special education journals closely reflect the limited adherence to the *Graphing Standards* found in behavior analytic journals (Kubina et al., 2017). For example, a unit of time appeared as a label on the horizontal axis for about 30% of graphs in both special education and behavior analytic journals (e.g., essential structure). Adherence for the quality feature of proportional construction came to approximately 15%. Like behavior analysis, the present review on special education journals suggests similar ramifications for the visual interpretation of data.

Findings of the present study revealed that special education graphs either (1) lack a horizontal axis label entirely or (2) use a label that does not actually refer to time (e.g., sessions). The effect of labeling the horizontal axis on visual analysis has not reached empirical validation (Ledford et al., 2019). However, if researchers intend to use line graphs as a time-series to display change through time, they must have a time unit (Few, 2009). In addition, the absence of time potentially inhibits the ability to forecast any future changes or place behavioral changes in a meaningful context (Johnston et al., 2020). Concepts such as immediacy and concurrence remain central to visual analysis, which provides a strong theoretical rationale for the depiction of time along a continuous scale.

Quality indicators for SCED research heavily emphasize the precise chronological order of events and the temporal relation between data points to establish a functional relation. WWC (2020) indicators require a series of consecutive data points and require observation sessions to occur within a certain period prior to the intervention. The guidelines (1) refer specifically to depiction of data rather than reports within the text and (2) lose much of their meaning when applied to an ordinal metric (i.e., sessions). Furthermore, sessions or "trials" cloud the ability to fully understand intervention intensity and duration (Romanczyk et al., 2014). The obfuscation of time identified in this review could have implications for the interpretation of research. Future work should confirm the importance of grounding time-series graphical presentations in units of time.

The *Graphing Standards* would further suggest lack of adherence to the 5:8 to 3:4 vertical to horizontal axis ratio in most studies potentially exaggerates or compresses trend and variability (*Graphing Standards*; Kubina et al., 2017); at minimum, inconsistent graphing practices would potentially result in widely varying interpretations of similar data presented in separate articles. The proportional construction rule provides a standardizing effect across graphs. In comparison, other fields have grappled with nonstandardization and the resulting variability of visual displays and the subsequent data. Researchers in the ophthalmic and medical community noted that a crucial element of peer-reviewed publications revolves around clear communication. The lack of consistency deepens complexity when journal readers evaluate and compare data and outcomes (Dupps et al., 2011).

Other ostensibly less critical quality features (i.e., "aesthetic-altering elements;" Dart & Radley, 2018 p. 351) of line graph construction may support effective data analysis and communication. For example, barely discernible data points could interfere with the identification of trends or patterns of behavior change. Departure from any other quality standard, such as guidelines regarding the positioning of data points on tick marks, potentially impede visual analysis and accompanying efforts to extract information for the calculation of effect sizes. The relevance of each individual standard to visual analysis remains unknown, however, due to the limited research in this area (Dart & Radley, 2018).

That the standards of graphic construction lack a basis in empirical work arguably mitigates the significance of the current findings. In general, we agree that research standards should have research evidence whenever possible. However, we do not think the lack of evidence should necessarily preempt consideration of the *Graphing Standards* in practice until such a time a research in this area provides robust support for alternative standards, for several reasons. The guidelines for graphic instruction bear some similarity to other standards of SCED research rooted in theoretical, rather than empirical, bases such as the commonly observed convention that experiments produce at least three demonstrations of effect (Lanovaz et al., 2019).

Other standards, such as those concerning the minimum number of experiments that support an effective practice (e.g., 5-3-20 criteria; Kratochwill et al., 2013) appear to lack a basis in theory or science (Lanovaz & Rapp, 2016).

In addition, socially enforced professional standards have historically established a subjective foundation for productive intradisciplinary communication in the absence of any clear empirical rationale (Danziger, 1985). As with other standards of presentation (e.g., APA style; Sigal & Pettit, 2012), cleaving to arbitrary conventions for graphic construction prevents distortions resulting from alterity. The last publication of the time-series standards explicitly addressed importance of consensus in the absence of evidence:

> This standard is a collection of preferred practices rather than a set of detailed specifications. It sets forth the best current usage, and thus offers standards "by general agreement" rather than "by scientific test." This concept of a standard based on practice implies and embodies gradual but noticeable changes over the years. As experience in the field of graphic presentation broadens and deepens, and as new problems occur, changing practices are inevitable. (American National Standards Committee Y15 [ANSC Y15.2], 1979, p. iii)

The extent to which empirical evidence, rather than “common sense” or the fiat of expert consensus, influences research standards is an important consideration. Yet research regarding the effect of construction guidelines remains at an early stage and invariably involves the discretion of researchers in specifying design criteria. For example, Cleveland et al. (1988) premise their support for the median-absolute-slope criterion on an extremely small sample (*n* = 16). Likewise, conclusions regarding the relative benefits of DPPXYR stem from a single study featuring a small “fairly homogenous sample” (*n* = 29; Radley et al., 2018, p. 321), and a search of 5 years of research in a small number of school psychology journals (*n* = 5) served as the basis of recommended DPPXYR values. It is worth noting that reviews conducted by Peltier, McKenna, et al. (2021a); Peltier, Morano, et al. (2021b) of journals in special education found different mean DPPXYR across graphs. The selection of school psychology journals as the basis for DPPXYR guidelines, rather than those found in special education journals, limits its generality.

In identifying the limitations of this scholarship, we do not intend to undermine the value of efforts to establish graphing conventions. The question remains, however: If evidence should serve as the foundation for quality standards, how much evidence should the field require? Such answers to the previous question and other methodological quandaries will likely assume provisional status and require some degree of subjective discretion. Though new evidence may inevitably promote revised standards, presentation guidelines (e.g., APA, 2020) assist in maintaining the consistency and integrity of the special education profession (Dart & Radley, 2018). We therefore support adherence to the *Graphing Standards* until compelling evidence emerges in support of an alternative.

Limitations

This study has four notable limitations. First, journals included in the search appeared in a widely accepted subject classification systems (i.e., JCR) or reflect the full range of subdisciplines explicitly acknowledged by the leading professional organization in special education (e.g., CEC). The present review, however, did not provide an exhaustive examination of potentially relevant journals. Second, criteria regarding the omission of certain graphs resulted in the exclusion of several unique visual displays. The intentional exclusion of atypical displays avoided presenting an excessively negative depiction of special education research. On the other hand, exclusion of displays for which the standards may not apply potentially masks the extent of revisions needed to make the standards relevant to practice. Third, aggregating the results, rather than presenting adherence to graphing standards over time or by discipline, may have obscured changes in research reporting. The summative view of the data remains appropriate given that efforts to identify evidence-based practices in special education do not exclusively rely on current research or discipline specific resources. Finally, findings represent an estimate of actual prevalence derived from a sample of included issues. Although more precise searches of narrower samples (e.g., Peltier, McKenna, et al., 2021a) partially corroborate findings from the current study, we urge caution when interpreting the results given the error inherent to such analyses. Alternate approaches to evaluating graphical outcomes would likely result in different outcomes (King et al., 2018).

Future Directions

Initial evidence suggests visually distorted data exaggerate visual impressions and lead to errors in the interpretation of data. The evidence provides a rationale for standards of graphic display (Dart & Radley, 2017). The present study does not directly provide evidence regarding the effect of nonstandard graphs on consumer interpretation. To ensure the validity of visual design standards, research regarding the impact of specific graphing features on visual analysis should continue. Such scholarship could include assessments of the influence of the ratio of axes (Radley et al., 2018), use of ordinal temporal dimensions (Ledford et al., 2019), and quality features (e.g., tick mark placement) on visual analysis and data extraction. Evidence that challenges existing standards should result in their revision, an exciting process with the potential to change the field. For example, additional studies suggesting DPPXYR represents a more critical design element than the ratio of axes (Radley et al., 2018) could give cause to modify or abandon the proportional construction rule. A more dramatic option could consist of jettisoning equal interval displays entirely and adopting standard ratio graphs due to their compatibility with data-based individualization and other instructional purposes (Datchuk & Kubina Jr., 2011; Jung et al., 2018; Kinney et al., 2020). Based on the limited adherence to existing graphing standards identified in the current review, research and development of graphing conventions should also emphasize dissemination and adherence to standards of scholarship.

Questions regarding the validity of extant graphing standards, however critical to the evolution of the discipline, elide an equally significant issue facing fields that rely on SCED: the inconsistency between special education graphs and existing standards. Widespread lack of adherence to professional standards effectively renders them irrelevant. Ensuring that visual analysis serves as a credible analytical tool requires greater attention to the elements of graphic display as well as a greater commitment to upholding and disseminating professional standards.

Raising awareness regarding the importance of graphing conventions may require concrete action within the field of special education. Guidelines for visual displays within prominent special education journals may address issues introduced through heterogenous, potentially misleading line graphs. Calls for articles underscoring the importance of graphing standards would likewise increase the salience of this issue (e.g., Dart & Radley, 2018). Reviews targeting reporting practices pertaining to the assessment of the dependent variable (e.g., IOA; Kostewicz et al., 2016) and implementation fidelity (Ledford & Wolery, 2013) suggest that such efforts have successfully changed publication practices. A multitiered approach to cultural change within the research community that encompasses resources, incentives (e.g., special recognition), and policy changes may also achieve gradual changes in the behavior of researchers (Nosek, 2019).

Incorporating standards for the visual display of SCEDs into general standards of evidence represents another important step. Recent research quality established by the CEC (Cook et al., 2015) simply require authors to provide a legible graph. Advocates of this approach suggest the fundamental transparency of graphical display provides some protection from wildly inaccurate interpretations (e.g., Wolery et al., 2010). Although construction standards could prevent graphical features from fostering missteps in judgment, supporting systematic approaches to visual analysis and statistical analyses could also limit the impact of suboptimal line graphs (e.g., Manolov et al., 2016). As an alternative, the field could follow the example set by the WWC (2020) and minimize the role of visual analysis in interpreting SCED. Given that visual analysis traditionally encompasses a range of relevant features resistant to quantification (e.g., trend, immediacy, consistency; Tanious et al., 2019), preserving its role in SCED would better serve researchers and practitioners. Eliminating issues in interpretation due to the vagaries of unstandardized graphic displays, whether through the adoption *Graphing Standards* or other guidelines for construction, could contribute to the credibility and persistence of visual analysis and SCED in special education.

Declarations

Conflicts of interest All authors declare that they have no Conflicts of interest.

Human participants or animals rights The present study did not involve human participants or animals.

Informed consent Due to the absence of participants, the research did not require informed consent.

References

American National Standards Committee Y15 [ANSC Y15.2]. (1979). *American Standard Time-Series Charts* (ASA Y15.2M-1979). American Society of Mechanical Engineers.

American National Standards Institute & American Society of Mechanical Engineers. (1960). *Time-series charts*. ASME.

American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). Author.

American Society of Mechanical Engineers. (1960). *American Standard Time-Series Charts* (ASA Y15.2).

Blair, E., & Blair, J. (2015). Setting sample size. In *Applied survey sampling* (pp. 91–112). Sage. https://doi.org/10.4135/9781483339402

Brinton, W. C. (1914). *Graphic methods for presenting facts*. Engineering Magazine Company.

Cleveland, W. S. (1994). *The elements of graphing data*. Hobart Press.

Cleveland, W. S., McGill, M. E., & McGill, R. (1988). The shape parameter of a two-variable graph. *Journal of the American Statistical Association, 83*(402), 289–300. https://doi.org/10.2307/2288843

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). John Wiley & Sons.

Committee on Standards for Graphic Presentation. (1938). *Time series charts: A manual of design and construction*. American Society of Mechanical Engineers.

Cook, B. G., Tankersley, M., Cook, L., & Landrum, T. J. (2015). Republication of "Evidence based practices in special education: Some practical considerations". *Intervention in School & Clinic, 50*, 310–315. https://doi.org/10.1177/1053451214532071

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis* (3rd ed.). Pearson Education.

Daniel, W. W., & Cross, C. L. (2018). *Biostatistics: A foundation for analysis in the health sciences* (11th ed.). John Wiley & Sons.

Danziger, K. (1985). The methodological imperative in psychology. *Philosophy of the Social Sciences, 15*(1), 1–13.

Dart, E. H., & Radley, K. C. (2017). The impact of ordinate scaling on the visual analysis of single-case data. *Journal of School Psychology, 63*, 105–118. https://doi.org/10.1016/j.jsp.2017.03.008

Dart, E. H., & Radley, K. C. (2018). Toward a standard assembly of linear graphs. *School Psychology Quarterly, 33*, 350–355. https://doi.org/10.1037/spq0000269

Datchuk, S. M., & Kubina Jr., R. M. (2011). Communicating experimental findings in single case design research: How to use celeration values and celeration multipliers to measure direction, magnitude, and change of slope. *Journal of Precision Teaching & Celeration, 27*, 3–17.

Dupps, W. J., Kohnen, T., Mamalis, N., Rosen, E. S., Koch, D. D., Obstbaum, S. A., Waring, G. O., Reinstein, D. Z., & Stulting, R. D. (2011). Standardized graphs and terms for refractive surgery results. *Journal of Cataract & Refractive Surgery, 37*, 1–3. https://doi.org/10.1016/j.jcrs.2010.11.010

Few, S. (2009). *Now you see it: Simple visualization techniques for quantitative analysis*. Analytics Press.

Groves, R. M., Fowler Jr., F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). John Wiley & Sons.

Harris, R. L. (1999). *Information graphics: a comprehensive illustrated reference*. Oxford University Press.

Heer, J., Kong, N., & Agrawala, M. (2009). Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1303–1312. https://doi.org/10.1145/1518701.1518897

Horner, R., & Spaulding, S. (2010). Single-case research designs. In N. J. Salkind (Ed.), *Encyclopedia of Research Design* (pp. 1386–1394). Sage.

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. https://doi.org/10.1177/001440290507100203

Individuals with Disabilities Education Act (2004) 20 U.S.C. § 1400.

Johnston, J. M., Pennypacker, H. S., & Green, G. (2020). *Strategies and tactics of behavioral research and practice* (4th ed.). Routledge.

Joint Committee on Standards for Graphic Presentation. (1915). Preliminary report published for the purpose of inviting suggestions for the benefit of the committee. *Publications of the American Statistical Association, 14*(112), 790–797. https://doi.org/10.2307/2965153

Jung, P. G., McMaster, K. L., Kunkel, A. K., Shin, J., & Stecker, P. M. (2018). Effects of data-based individualization for students with intensive learning needs: A meta-analysis. *Learning Disabilities Research & Practice, 33*(3), 144–155. https://doi.org/10.1111/ldrp.12172

Karsten, K. G. (1923). *Charts and graphs: An introduction to graphic methods in the control and analysis of statistics*. Prentice Hall.

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

King, S. A., Davidson, K., Chitiyo, A., & Apple, D. (2018). Evaluating Article Search and Selection Procedures in

Special Education Literature Reviews. *Remedial and Special Education*. Advance online publication available at:. https://doi.org/10.1177/0741932518813142.

Kinney, C. E., Begeny, J. C., Stage, S. A., Patterson, S., & Johnson, A. (2020). Three alternatives for graphing behavioral data: A comparison of usability and acceptability. *Behavior Modification*. https://doi.org/10.1177/0145445520946321

Kostewicz, D. E., King, S. A., Datchuk, S. M., Brennan, K. A., & Casey, S. D. (2016). Data collection and measurement assessment in behavioral research: 1958-2013. *Behavior Analysis: Research & Practice, 16*, 19–33. https://doi.org/10.1037/bar0000031

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34,* 26–38. https://doi.org/10.1177/0741932512452794.

Kubina, R. M., Kostewicz, D. E., & Datchuk, S. M. (2010). Graph and table use in special education: A review and analysis of the communication of data. *Evaluation & Research in Education, 23*, 105–119. https://doi.org/10.1080/09500791003734688

Kubina, R. M., Kostewicz, D. E., Brennan, K. M., & King, S. A. (2017). A critical review of line graphs in behavior analytic journals. *Educational Psychology Review, 29*, 583–598. https://doi.org/10.1007/s10648-015-9339-x

Lanovaz, M. J., & Rapp, J. T. (2016). Using single-case experiments to support evidence-based decisions: How much is enough? *Behavior Modification, 40*, 377–395. https://doi.org/10.1177/0145445515613584

Lanovaz, M. J., Turgeon, S., Cardinal, P., & Wheatley, T. L. (2019). Using single-case designs in practical settings: Is within-subject replication always necessary? *Perspectives on Behavior Science, 42*, 153–162. https://doi.org/10.1007/s40614-018-0138-9

Ledford, J. R., & Gast, D. L. (2018). *Single case research methodology: Applications in special education and behavioral sciences* (3rd ed.). Routledge.

Ledford, J. R., & Wolery, M. (2013). Procedural fidelity: An analysis of measurement and reporting practices. *Journal of Early Intervention, 35*, 173–193. https://doi.org/10.1177/1053815113515908

Ledford, J. R., Barton, E. E., Severini, K. E., Zimmerman, K. N., & Pokorski, E. A. (2019). Visual display of graphic data in single case design studies: Systematic review and expert preference analysis. *Education & Training in Autism & Developmental Disabilities, 54*(4), 315–327 https://eric.ed.gov/?id=EJ1234581.

Manolov, R., Losada, J. L., Chacón-Moscoso, S., & Sanduvete-Chaves, S. (2016). Analyzing two-phase single-case data with non-overlap and mean difference indices: Illustration, software tools, and alternatives. *Frontiers in Psychology, 7*, Article 32. https://doi.org/10.3389/fpsyg.2016.00032

Nelson, P. M., Van Normand, E. R., & Christ, T. J. (2016). Visual analysis among novices: Training and trend lines as graphic aids. *Contemporary School Psychology, 21*, 93–102. https://doi.org/10.1007/s40688-016-0107-9

Ninci, J., Vannest, K. J., Wilson, V., & Zhang, N. (2015). Interrater agreement between visual analysts of single-case data: A meta-analysis. *Behavior Modification, 39*, 510–541. https://doi.org/10.1177/0145445515581327

Nosek, B. (2019, June 11). *Strategy for culture change*. Center for Open Science. https://cos.io/blog/strategy-culture-change/

Parsonson, B., & Baer, D. (1978). The analysis and presentation of graphic data. In T. Kratochwill (Ed.), *Single subject research* (pp. 101–166). Academic Press.

Peltier, C., McKenna, J. W., Sinclair, T. E., Garwood, J., & Vannest, K. J. (2021a). Brief report: Ordinate scaling and axis proportions of single-case graphs in two prominent EBD journals from 2010 to 2019. *Behavioral Disorders*. https://doi.org/10.1177/0198742920982587

Peltier, C., Morano, S., Shin, M., Stevenson, N., & McKenna, J. W. (2021). A Decade Review of Single-Case Graph Construction in the Field of Learning Disabilities. *Learning Disabilities Research & Practice, 36,* 121–135. https://doi.org/10.1111/ldrp.12245.

Poling, A. D., Methot, L. L., & LeSage, M. G. (1995). *Fundamentals of behavior analytic research*. Plenum Press.

Radley, K. C., Dart, E. H., & Wright, S. J. (2018). The effect of data points per x- to y-axis ratio on visual analysts' evaluation of single-case graphs. *School Psychology Quarterly, 33*(2), 314–322. https://doi.org/10.1037/spq0000243

Riley-Tillman, T., & Burns, M. K. (2009). *Evaluating educational interventions: Single-case design for measuring response to intervention*. Guilford Press.

Robbins, N. B. (2005). *Creating more effective graphs*. John Wiley & Sons.

Romanczyk, R. G., Callahan, E. H., Turner, L. B., & Cavalari, R. N. S. (2014). Efficacy of behavioral interventions for young children with autism spectrum disorders: Public policy, the evidence base, and implementation parameters. *Review Journal of Autism & Developmental Disorders, 1*, 276–326. https://doi.org/10.1007/s40489-014-0025-6

Sectional Committee on Standards for Graphic Presentation. (1936). *Code of preferred practice for graphic presentation: Time series charts*. American Society of Mechanical Engineers.

Shadish, W. R., Hedges, L. V., Horner, R. H., & Odom, S. L. (2015). *The role of between-case effect size in conducting, interpreting, and summarizing single-case research.* (NCER 2015-002) National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncser/pubs/2015002/pdf/2015002.pdf. Accessed December 2015

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. Basic Books.

Sigal, M. J., & Pettit, M. (2012). Information overload, professionalization, and the origins of the publication manual of the American Psychological Association. *Review of General Psychology, 16*(4), 357–363.

Spriggs, A. D., Lane, J. D., & Gast, D. L. (2018). Visual representation of data. In J. R. Ledford & D. L. Gast (Eds.), *Single case research methodology: Applications in special education and behavioral sciences* (pp. 157–178). Routledge.

Tanious, R., Manolov, R., & Onghena, P. (2021). The Assessment of Consistency in Single-Case Experiments: Beyond A-B-A-B Designs. *Behavior Modification, 45*(4), 560-580. https://doi.org/10.1177/0145445519882889

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics, 10*, 365–391. https://doi.org/10.1016/j.joi.2016.02.007

Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics, 10*, 347–364. https://doi.org/10.1016/j.joi.2016.02.003

Web of Science. (2014). *2013 journal citation reports® social sciences edition.* Thompson Reuters. https://www.jcr-incites-thomsonreuters-com.

Wendel, E., Cawthon, S. W., Ge, J. J., & Beretvas, S. N. (2015). Alignment of single-case design (SCD) research with individuals who are deaf or hard of hearing with the What Works Clearinghouse Standards for SCD research. *Journal of Deaf Studies & Deaf Education, 20*(2), 103–114. https://doi.org/10.1093/deafed/enu049

What Works Clearinghouse. (2020). *What Works Clearinghouse standards handbook* (Vers. 4.1). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf. Accessed October 2020

Wolery, M., Busick, M., Reichow, B., & Barton, E. E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *Journal of Special Education, 44*, 18–28. https://doi.org/10.1177/0022466908328009