Educ Psychol Rev (2017) 29:583–598
DOI 10.1007/s10648-015-9339-x

REVIEW ARTICLE

A Critical Review of Line Graphs in Behavior Analytic Journals

Richard M. Kubina Jr.[1] **· Douglas E. Kostewicz**[2] **·**
Kaitlyn M. Brennan[2] **· Seth A. King**[3]

Published online: 3 September 2015

Abstract Visual displays such as graphs have played an instrumental role in psychology. One discipline relies almost exclusively on graphs in both applied and basic settings, behavior analysis. The most common graphic used in behavior analysis falls under the category of time series. The line graph represents the most frequently used display for visual analysis and subsequent interpretation and communication of experimental findings. Behavior analysis, like the rest of psychology, has opted to use non-standard line graphs. Therefore, the degree to which graphical quality occurs remains unknown. The current article surveys the *essential structure* and *quality features* of line graphs in behavioral journals. Four thousand three hundred and thirteen graphs from 11 journals served as the sample. Results of the survey indicate a high degree of deviation from standards of graph construction and proper labeling. A discussion of the problems associated with graphing errors, future directions for graphing in the field of behavior analysis, and the need for standards adopted for line graphs follows.

Keywords Line graphs · Time series · Graphical construction guidelines · Graphing standards

Behavior analysis, a subfield of psychology, owes a great debt to the visual display of data. For example, the cumulative recorder offered a standard visual display of an organism's performance data. The distinctive visual patterns of behavior led to the discoveries such as schedules of reinforcement (Lattal 2004). As behavior analysis moved forward in time, the visual displays shifted from cumulative recorders to line graphs. Data show that cumulative records in the

✉ Richard M. Kubina, Jr.
rmk11@psu.edu

[1] Special Education Program, The Pennsylvania State University, 209 CEDAR, Building University Park, State College, PA 16802-3109, USA

[2] University of Pittsburgh, Pittsburgh, PA, USA

[3] Tennessee Technological University, Cookeville, TN, USA

Journal of the Experimental Analysis of Behavior continue to appear infrequently and in other years not at all (Kangas and Cassidy 2010).

The shift away from cumulative records to line graphs coincided with the advent of an emphasis on applied work. The oft-cited paper of Baer et al. (1968) laid the foundation for discerning the facets of behavior analysis. Three of seven characteristics of applied behavior analysis have a direct link to visual displays of data. First, *Analytic* refers to a convincing demonstration of an experimental effect. The preferred medium for all analysis of data occurs through graphs. Second, *Effective* conveys the requirement for the intervention to produce a practical and meaningful magnitude of behavior change. Line graphs allow for the determination as well as the public documentation and communication of the significance of behavioral improvements (Spriggs and Gast 2010). And third, *Generality* means that the behavior persists across time, environments, and operant responses within a class. The line graph, part of the family of time series graphs, directly portrays the extent to which behavior does or does not persist.

Principles of graphic presentation for line graphs have quality standards necessary for the accurate representation of data. A number of publications have described the standards for proper construction for line graphs (American National Standards Institute and American Society of Mechanical Engineers 1960, 1979; American Standards Association 1938; American Statistical Association 1915; Department of the Army 2010). For example, the publication of *Time series charts: a manual of design and construction* set forth agreed upon standards for line graphs (American Standards Association 1938). The committee provided guidance on many specific features ranging from scale rulings and graph dimensions to the weight of lines and use of reference symbols. Through time, many professional organizations and researchers have continued to offer principles of design and procedure for constructing high-caliber line graphs (e.g., Behavior Analysis—Cooper et al. 2007; Statistics—Cleveland 1993, 1994; General Science—Scientific Illustration Committee 1988; Technical Drawing, Drafting, and Mechanical Engineering—Giesecke et al. 2012). Table 1 lists major *quality features* of line graphs tailored toward use in the behavioral sciences.

An analysis of the following basic behavior analysis (Alberto and Troutman 2013; Catania 1998; Cooper et al. 2007; Malott and Shane 2014; Mayer et al. 2014; Pierce and Cheney 2013; Vargas 2013) and single case design books (Barlow et al. 2009; Gast 2010; Johnston and Pennypacker 2009; Kazdin 2011; Kennedy 2005) corresponds to the graphical standards for a line graph previously listed. In addition to *quality standards* line graphs have an *essential structure* consisting of two axes, the horizontal and vertical, representing a time unit and a quantitative value, respectively. Time units can cover minutes, hours, days, weeks, and years based on the second (National Institute of Standards and Technology 2014). The range of behavior on the vertical axis spans dimensionless quantities like percentages and ratios to dimensional quantities such as repeatability and temporal extent measured with frequency and duration, respectively (Johnston and Pennypacker 2009).

Not adhering to the *essential structure* may yield distorted, exaggerated, or imprecise information. The *essential structure* shows change over time. Figure 1 shows three line graphs with the same data. The first line graph made following the "proportional construction ratio," discussed later, displays a series of data with a moderately increasing variable trend. The line graph has an extended vertical axis changing the variability from moderate to low. The trend also increases when compared to the previous graph. Stretching the horizontal axis in the third line graph depresses the trend and decreases variability.

Table 1 Quality features of a line graph and measurement

Essential structure	Function	Measured
Vertical axis labeled with quantitative measure; horizontal axis labeled with time unit	To show the change in the measure over time (Harris 1999)	What label does the vertical and horizontal axes maintain? Does the figure maintain a line for each axis?
Quality feature		
Vertical axis length has a 2:3 ratio to the horizontal axis	To properly display data variability and limit distortion (Cooper et al. 2007; Parsonson and Baer 1978)	Is the ratio of vertical to horizontal axis 5:8 to 3:4 (63 to 75 % difference)? Do axes line up in all multiple baseline graphs? Are axis lengths the same for all figures with the same unit within each article? Are all figures with the same unit scaled to the same minimum and maximum?
Tick marks point outward	To prevent or minimize data obfuscation (Cleveland 1994)	On both axes, are tick marks pointing outward for the entire length of the axes?
A minimal number of evenly spaced tick marks	To decrease graph clutter emphasizing the most prominent feature, the data (Cleveland 1994)	On both axes, are tick marks evenly spaced (i.e., at equal intervals)?
Tick marks have labels	Delineates the value of the axes' units (Robbins 2005)	On both axes, are tick marks numbered? Are the scale counts correct?
Data points clearly visible	Enhances data clarity (Robbins 2005)	Are data points on the figure clearly visible?
Data paths clearly visible (if used)	Shows the direction of data clearly (Cooper et al. 2007)	If the figure contains a data path, is it visible?
Condition change lines (if used)	Visually separates data between conditions (Cooper et al. 2007)	If the figure contains a condition change line, is it visible?
Condition labels	Identification of experimental conditions (Cooper et al. 2007)	If the figure contains a condition change line, does it have labels?
Figure caption	When combined with other graphic elements, conveys meaning (Cooper et al. 2007)	Does the figure contain a caption?

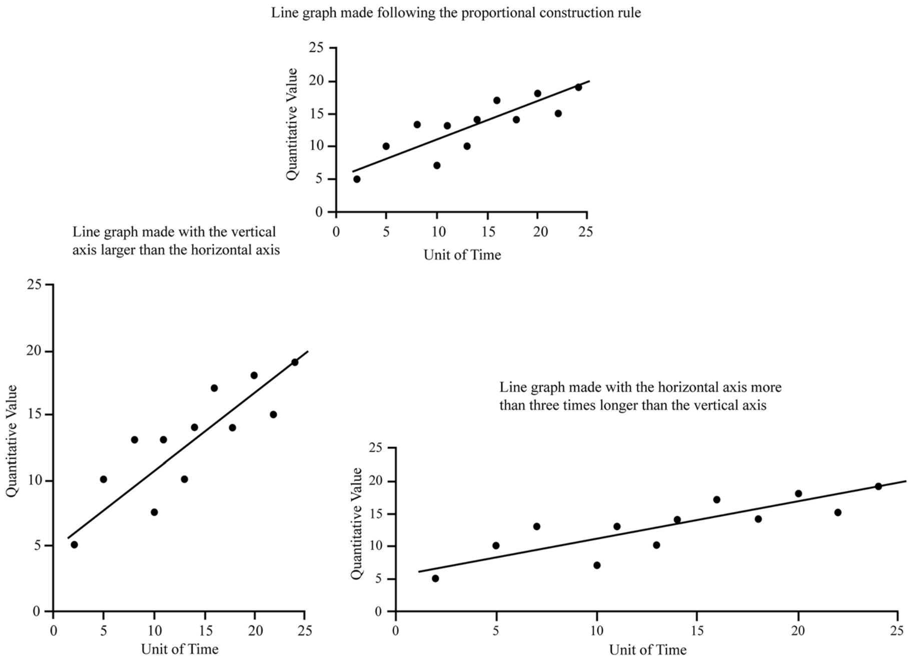

Fig. 1 Sample graphs containing the same scaling with variable length axes

The *Behavior of Organisms* of Skinner (1938), "Some Current Dimensions of Applied Behavior Analysis" of Baer et al. (1968), and chapters of Cooper et al. (2007) on the construction and interpretation of graphical displays represent examples for the use, rationale, and creation of behavior analytic line graphs. As a result, line graphs have become the primary visual display for presenting behavioral data in fieldwork, theses, dissertations, lectures, conference presentations, and journal articles (Cooper et al. 2007; Mayer et al. 2014; Poling et al. 1995; Spriggs and Gast 2010). How well the field of behavior analysis attends to *essential structure* and *quality features* of line graphs, however, remains unknown. The current survey examines the quality of line graphs contained in behavioral journals and attempts to answer two questions. First, how well do selected visual graphics follow the *essential structure* of line graph construction? Namely, to what extent do selected line charts have time units on the horizontal axis and quantitative units on the vertical axis? Second, how well do selected visual graphics follow the *quality features* of line graphs (Table 1)?

Method

Initial selection followed criteria established in previous surveys for the identification of prominent behavioral journals (Carr and Britton 2003; Critchfield 2002; Kubina et al. 2008). Journals had to explicitly pertain to behavior analysis and have at least a 10-year publication record. The survey sampled a variety of behavior analytic foci (e.g., education, cognitive behavior modification, experimental analysis). Eleven journals met criteria.

Six journals covered technical applications, practices, and issues related to the field of behavior analysis (*Behavior Modification*, *Behavior Therapy*, *Child and Family Behavior*

Therapy, *Cognitive and Behavioral Practice*, *Journal of Applied Behavior Analysis*, and *Journal of Behavior Therapy and Experimental* Psychiatry). Five additional journals discussed behavior analysis in relation to education (*Education and Treatment of Children*, *Journal of Behavioral Education*), experimental behavior analysis (*Journal of the Experimental Analysis of Behavior*, *Learning and Behavior*), and the analysis of verbal behavior (*The Analysis of Verbal Behavior*).

After journal identification, one random issue from every 2-year block served as the basis for selecting graphs. The process began at each journal's inception date and concluded in 2011. The investigators examined all graphs that had a vertical and horizontal axis with data moving left to right. First, the graph must have contained a maximum of one data point per data series on the horizontal axis interval excluding scatterplot graphs (i.e., multiple data points can occur on the same horizontal interval) and bar charts. Second, a unit of time or sessions and a quantitative value must have occurred on the horizontal and vertical axis, respectively. Graphs scaled with nominal or ordinal vertical axes and/or non-time based horizontal axes did not meet inclusion criteria. Third, graphs with dually and/or logarithmically scaled vertical axes, as line graph variants, also failed to satisfy inclusion criteria. Investigators analyzed each included graph individually whether appearing alone or in the context of other graphs (i.e., multiple baselines).

Investigators scored each graph for components of line graph *essential structure* and the presence or absence of line graph *quality features* (specific questions appear on Table 1). Scorers initially determined presence or absence of axes and labels. Using rulers or straight edges, scorers then determined the ratio of the length of the vertical to the horizontal axis and the axes scaling and alignment to other graphs within the same figure (i.e., multiple baseline figures) and/or article. Scorers continued to evaluate each graph according to the remaining questions noted on Table 1 and entered all data on an accompanying *Excel* file. The process repeated for each graph meeting criteria.

Scorer Calibration, Reliability, and Interobserver Agreement

Scorers received instruction on all procedures. Instruction consisted of review and guided practice of scoring and entering data on the *Excel* spreadsheet for three graphs. At the conclusion of the instructional sessions, experimenters had scorers evaluate a random, but previously scored issue and compared results. Scorers had to meet 100 % agreement prior to independent scoring.

Two measurement assessment techniques evaluated scoring: reliability and interobserver agreement. For reliability (Johnston and Pennypacker 2009), each scorer rescored 20 % of issues. A comparison occurred between the second examination and the initial score. An exact agreement approach (Kennedy 2005) determined the percent of agreement between individual cells of data in Excel sheets. The average reliability totaled 95 % with a range of 94–100 %. Interobserver agreement followed the same procedure but took place between different scorers on 20 % (28) of issues. The average interobserver agreement equaled 91 % with a range of 89–100 %.

Results

A total of 11 behavioral journals served as the basis of the graph analysis. A random sampling produced 191 issues, an average of 17 issues per journal with a range of 9–27. A potential 5989 time series graphics occurred in 1622 articles. Removing graphs that contained logarithmically scaled (114) or dually scaled axes (203), the number of graphs came to 5672. Graphs with nominal (60) or ordinal (227) scaled vertical axes (i.e., qualitatively scaled) or included a label for the horizontal axis that did not fall into a unit of time (1072) failed to meet review criteria resulting in a total of 4313 coded graphs or approximately 23 per issue.

Essential Structure

The *essential structure* of a simple line graph starts with two drawn axes, a vertical and horizontal. Of the 4313 graphs meeting criteria, 98 % (4206) and 97 % (4200) of graphs contained a drawn line on the vertical and horizontal axis, respectively. Two additional *essential structure* criteria require the labeling of the two axes with a quantitative value on the vertical axis and time along the horizontal axis. Two dot charts (Cleveland 1984b), in Figs. 2 and 3, show the breakdown that authors used for labeling axes. On both figures, dots represent categorical instances and appear from greatest to least.

Figure 2 shows that, of the eight vertical label categories, five contained 98 % of instances. Authors used a percent label 27 % (1159) of the time on the vertical axes, more often than any other classification. Count (20 % or 865), ratio (18 % or 782), no label (17 % or 753), and frequency/rate (16 % or 677) groupings round out the initial 98 %. The remaining 2 % consist of latency (40), duration (36), and interresponse time (1).

On Fig. 3, labels for the horizontal axes revealed less diffusion as compared to the vertical axis breakdown. No label (1486) and sessions/trials (1469) accounted for 69 %. A total of 459 (11 %) graphs incorporated seconds, 268 (6 %) days, and 262 (6 %) minutes along the horizontal axis. The final 9 % include other unit of time (172), multiple units on the same axis (160), and hours (37).

Quality Features

Tables 2 and 3 show a variety of *quality features* coded for each graph and groupings of graphs. Graphs meeting each instance receive a comparison to the total number of opportunities resulting in a percent occurrence. Coded graphics not meeting the *quality feature* (except

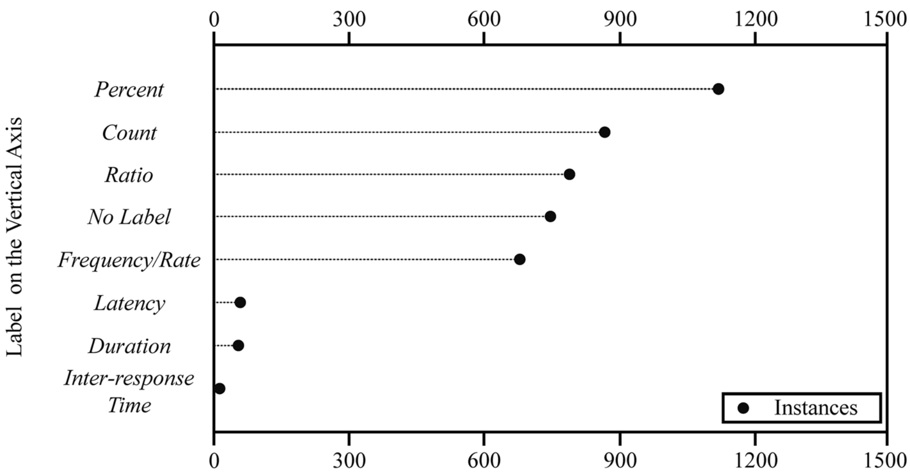

Fig. 2 A dot chart showing the instances of vertical axis labels

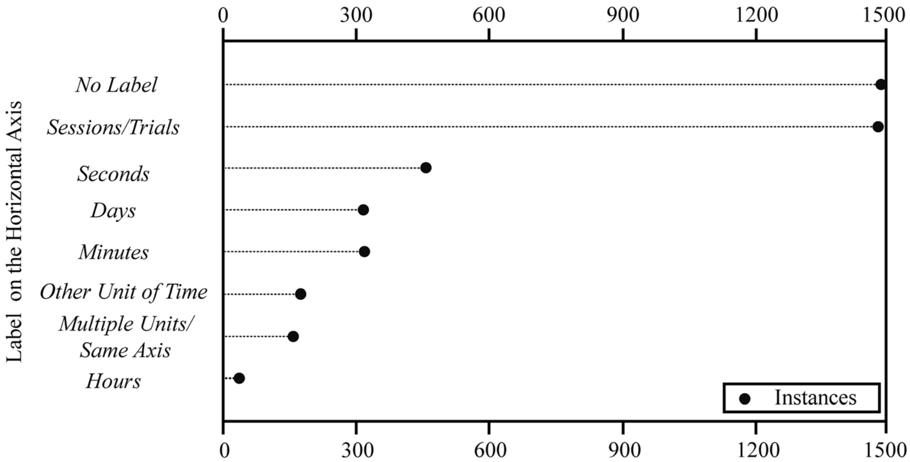

Fig. 3 A dot chart showing the instances of horizontal axis labels

for data connected across condition change line) result in an error shown as a remaining percentage in the final column. Each instance of the category *data connected across condition change line* already constitutes an error; thus, the same percentage occurs in both columns.

Tick marks and scaling Considerable differences in quality occurred between tick marks and scaling of the vertical and horizontal axes. Table 2 shows that, across the five tick mark

Table 2 Quality features

Graphic quality feature		Instances per opportunity	Percent of error (%)
Vertical axis tick marks	Data occur on tick marks	3481/4313=81 %	19
	Full or partial tick marks on outside of graph	2460/4313=57 %	43
	Tick marks occur at equal intervals	3768/4313=87 %	13
	Numbers occur on tick marks	3493/4313=81 %	19
	Scale count is correct (e.g., 10, 20, 30, etc.)	3702/4313=86 %	14
Horizontal axis tick marks	Data occur on tick marks	2866/4313=66 %	34
	Full or partial tick marks on outside of graph	2115/4313=49 %	51
	Tick marks occur at equal intervals	2714/4313=63 %	37
	Numbers occur on tick marks	2214/4313=51 %	49
	Scale count is correct (e.g., 10, 20, 30, etc.)	2304/4313=53 %	47
Data points clearly visible		3703/4313 =86 %	14
Data connected and data path clearly visible		4193/4313=97 %	3
Figure caption		4253/4313=99 %	1
Condition change labels present		1774/2172=82 %	18
Data connected across condition change line		158/2151=7 %	7

and scaling features, half as much average error occurred on the vertical (22 %) rather than horizontal axis (44 %). On both axes, the highest percentages (43 and 51 %) of error related to fully or partially placing tick marks on the outside of the figure. Instead, graphs either had no tick marks or tick marks on the inside of the graph.

Data points and paths, condition labels, and figure captions A decrease in error appeared when coding *quality features* associated with data points and paths and condition and figure labels (Table 2). Only 1 and 3 % of the sampled graphs failed to contain a figure caption and had connected data paths clearly visible. Although, 18 % of the time graphs containing condition change lines failed to have a label for each condition and 7 % had data paths connected across condition change lines.

Axes and axis comparisons Table 3 highlights *quality features* associated with axis comparisons. The first main comparison occurred between the vertical and horizontal axes with 15 % meeting a ratio of 5:8 to 3:4 or 63 to 75 % difference in length. For graphs on the same page, both vertical and horizontal axes aligned, when possible, 79 % of instances. Expanding the analysis to each article, graphs received coding for maintaining the same physical and scaling structure. On both axes, approximately 70 % of instances failed to maintain similar scaling for the same unit across the article and 40 % had variable axis length.

Discussion

Graphs have one fundamental purpose: to affect the interpretative behavior of the graph reader (Johnston and Pennypacker 2009). Information in graphs includes documenting performance, analyzing intervention effects, interpreting experimental and applied outcomes, and predicting the future course of behavior. Graphs generate meaning based on physical distinctions of

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

Graphic quality feature			Instances per opportunity	Percent of error (%)
Ratio of vertical to horizontal axis length: 5:8 to 3:4 (63 to 75 % difference)			637/4313=15 %	85
For multiple graphs within the same figure on the same page:		Vertical axes align	1016/1285=79 %	21
		Horizontal axes align	1053/1331=79 %	21
For multiple graphs within the same article	Vertical axes that share the same label:	Scaled to same unit (min and max)	262/838=31 %	69
		Drawn to the same physical length	467/838=56 %	44
	Horizontal axes that share the same label:	Scaled to same unit (min and max)	222/710=31 %	69
		Drawn to the same physical length	436/710=61 %	39

shape, size, color, positioning, and symbols (Cleveland and McGill 1985; Tufte 1990). While all graphed data convey a message, "A graphical method is successful only if the decoding is effective. No matter how clever and how technologically impressive the encoding, it fails if the decoding process fails" (Cleveland and McGill 1985, p. 828).

The science of behavior fundamentally relies on simple line graphs for decoding information (Cooper et al. 2007). Not all simple line graphs, however, have equal merit. Line graphs can vary along critical dimensions such as scaling, length of axes, and labeling. Widely divergent construction practices distort the interpretative function of graphs. The use of *essential structures* and *quality features* promotes order and guides graphical construction, subsequently enhancing visual clarity and a clear explanation of the data (Cleveland 1984a). At the time of the present review, however, no comprehensive evaluation of simple line graphs in behavior analysis has occurred. The specific research questions asked how well do selected visual graphics follow the *essential structure* and *quality features* of line graph construction.

Essential Structure

Quantity and time form a line graph's *essential structure* (Few 2009; Kriglstein et al. 2014). Behavior analysis has long valued the graphic display of quantitative rather than qualitative representations of behavior (Baer et al. 1968; Ferster and Skinner 1957; Parsonson and Baer 1978; Poling et al. 1995). Quantification leads to precision provided by numbers. Numerical representations of behavior, or quantitative measurement, serve as the medium through which all analysis occurs. The data show that, out of 4313 reviewed graphs, 3560 had a quantitatively scaled vertical axis. In other words, 83 % of the reviewed graphs prominently displayed behavior as a quantity.

The quantitatively displayed behavior can change across time. Time series line graphs must have a unit of time on the horizontal axis (Robbins 2005). Examples of units of time range from seconds (e.g., Preston 1994) and minutes (e.g., Norborg et al. 1983) to hours (e.g., Ramirez 1997) and days (e.g., Gutentag and Hammer 2000). Twenty-eight percent of line graphs in behavioral journals maintained a time unit label. The remaining 72 % scaled the horizontal axis with a non-time unit (e.g., sessions, trials, no label). By labeling the horizontal axis with sessions or trials, the line graph technically no longer qualifies as a time series graphic and markedly influences visual analysis.

All graph readers use visual analysis to uncover functional relations and experimental effects (Cooper et al. 2007; Kazdin 2011). In point of fact, visual analysis of line graphs serves as the cornerstone of studies using a single case experimental design. "Data are graphed for each participant during a study with trend, level, and stability of data assessed within and between conditions (Lane and Gast 2014, p 445)." Trend refers to the slope or angle of a data series; level applies to the median score of a data set, and stability captures the degree of variability in a set of data (Gast and Spriggs 2010). Table 4 explains and Fig. 4 illustrates how session usage produces three categories of errors affecting visual analysis: labeling error, false equality, and non-representative data.

Labeling error involves assigning a designation other than a unit of time to the horizontal axis. As an example, graph 1 in Fig. 4 shows data plotted along a horizontal axis labeled with sessions and a vertical axis with quantified behaviors per session. Baseline data show a gradually increasing, moderately variable trend with a level of 25 behaviors per session. Data in intervention show the level rising to 40 behaviors per session with a slightly increasing and

Table 4 Labeling error and visual analytic distortions

Session practice		Problem when horizontal axis has "sessions" as the label and data graphed consecutively and contiguously.
Session duration reported as a time unit in text and…	Held consistent for the duration of the study	Labeling error[a]
	Not held consistent (i.e., presented as a possible range of time)	Labeling error and false equality[b]
Sessions graphed with respect to time and…	Occur consecutively in time (i.e., one session per day)	Labeling error
	Occur consecutively in time (i.e., one session per day) but do not occur every day (i.e., sick days and/or weekends)	Labeling error and non-representative data[c]
	Do not occur consecutively in time (e.g., multiple sessions 1 day and then additional sessions later in the week)	Labeling error and non-representative data
Session duration not reported as a time unit and…	The experiment lasts for an unknown duration of time	Labeling error, false equality, and non-representative data

[a] Time series line graph must have a unit of time depicted on the horizontal axis

[b] Produces incorrect trend, level, and variability

[c] Graphed data visually distorts the actual data, trend, and variability

moderately variable trend. Visual analysis would suggest an experimental effect due to the level change.

Graphs must contain all construction elements necessary to describe the data so the reader can draw accurate conclusions (Cleveland 1994). Or, as the famous data scientist Tukey stated, "A strongly good graph tells us everything we need to know just by looking at it" (as cited in Wainer 2005). Using sessions, however, fails to provide a full account of temporal information. The graph reader does not know how long each individual session lasted, if sessions occurred contiguously in time, or the duration of the experiment (e.g., 1 day or 1 month). Labeling session on the horizontal axis (i.e., the absence of time), as in graph 1 (Fig. 4), inhibits a chronologically accurate visual analysis of time series data.

Variable length sessions appearing equal on a time series graph represent false equality shown in the contrast between graph 1 and graph 2 (Fig. 4). Individual session duration may differ within and between experimental conditions. For instance, baseline sessions may have lasted 3 min and intervention sessions 9 min. Graph 1 would then establish a false equality of quantified behaviors per session. Graph 2's time-scaled behavior per 3 min rather than per session offers a different interpretation. Conclusions from baseline in both graph 1 and graph 2 remain the same. However, intervention data noticeably drop in level from 40 to 13.3 quantified behaviors per 3 min. In addition, trend and variability decrease. Visual analysis on graph 2 would suggest an opposite experimental effect to graph 1.

The use of sessions promotes non-representative data by removing data from the cycle of time (e.g., minutes, hours, days). Graph 3 appears visually distinct from graph 1, yet both contain the same data (Fig. 4). Graph 3 presents a precise time account of the data collection. Graph 1 compresses 80 days into 20 sessions. After properly placing the data in time, the trend

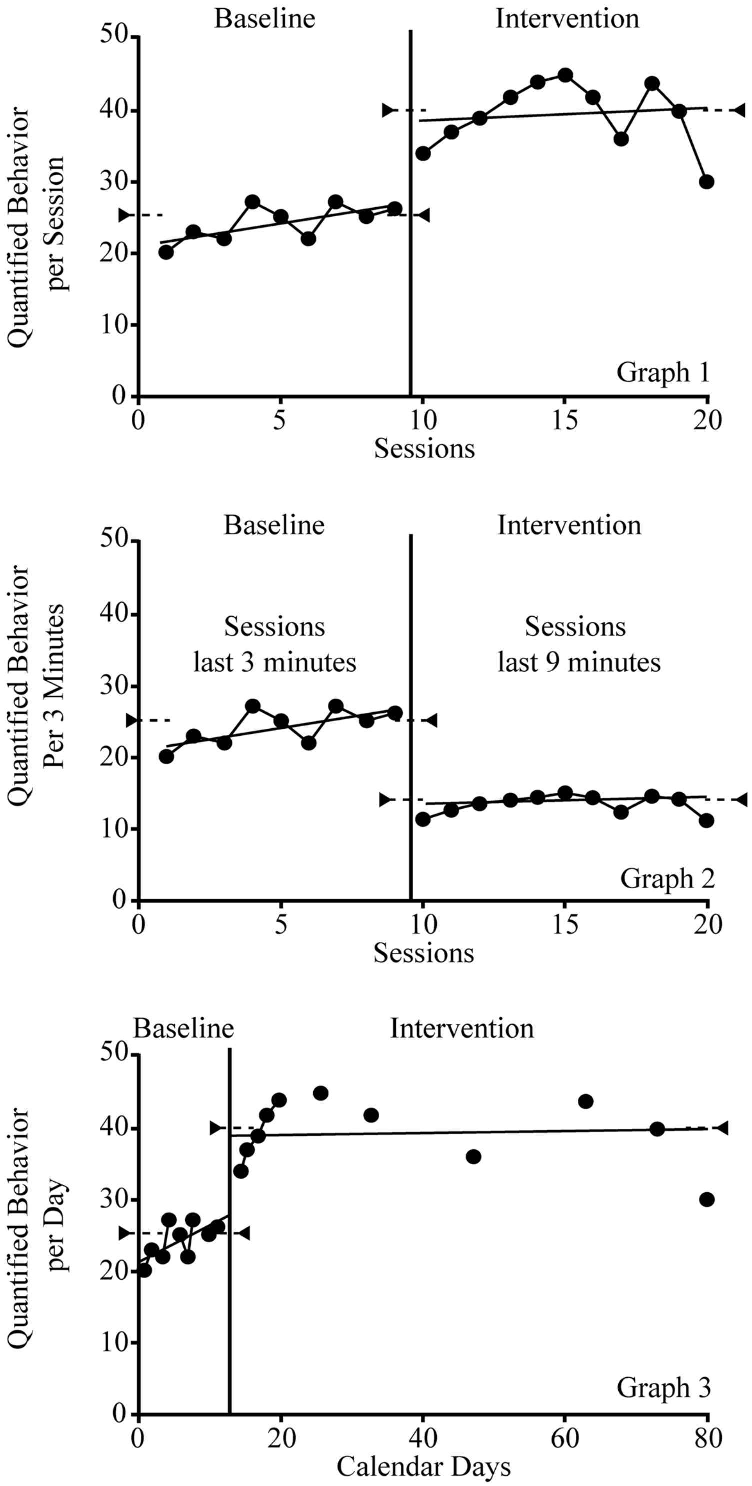

Fig. 4 Three line graphs illustrating the effects of labeling the horizontal axis with sessions

for baseline in graph 3 becomes steeper and looks more variable. Data in the intervention appear slightly less variable with a similar trend. Graph 1 hides important information obvious in graph 3. The steady application of the intervention for the first five consecutive days and the infrequent implementation across the remaining 60 days tell a revealing story. The initial five data points show a sharp increase while the remaining six gradually decline. Graphical integrity of time series such as the line graph necessitates respecting the passage of time. Additional examples of session usage error appear in Table 4.

Three possible reasons exist for the proliferation of labeling the horizontal axis with sessions: instruction, relativism, and substitution. “Instruction” refers to the guidance provided by behavior analytic (e.g., Cooper et al. 2007; Mayer et al. 2014) and single case design textbooks (e.g., Gast 2010; Kazdin 2011; Kennedy 2005). Both applied and basic researchers derive a substantial part of their graphing knowledge from textbooks. The widespread usage of sessions in journals as a labeling/graphing convention demonstrates “relativism” (e.g., 90 % of reviewed graphs in the *Journal of Behavioral Education* used sessions). The common use of sessions leads to the false impression of an acceptable practice. And, “substitution” occurs when behavior analysts measure behavior temporally but switch to sessions as a time placeholder. In other words, properly measuring behavior with time but improperly labeling the graph with a non-time unit sessions. Regardless of the sources of control or reasons why non-time units appear on the horizontal axis, *time* remains the major factor in a time series graphic (Cleveland 1994; Tufte 1983). The science of behavior should discard the practice of using sessions in line graphs; sessions offer no advantages for visually analyzing behavior and instead impede the accurate and efficient portrayal of behavioral data.

Quality Features

Quality features of line graphs convey representativeness and continuity of graphically displayed, time-oriented data. Graph readers decode quantitative information based on the visual characteristics of a graph’s construction parameters. The physical proportion of the vertical to the horizontal axis of a graph serves as one important interpretation variable. Published recommendations suggest a ratio of vertical to horizontal axis ranges of 5:8 to 2:3, with a maximum of 3:4 (American National Standards Institute and American Society of Mechanical Engineers 1960, 1979; Bowen 1992; Cooper et al. 2007; Department of the Army 2010; Johnston and Pennypacker 1980; Katzenberg 1975; Parsonson and Baer 1978; Poling et al. 1995; Schmid 1992). The respective 63, 66, and 75 % size differential of height to width presents a standardizing and protective effect. Namely, the proportional construction ratio enhances graphical legibility (Cooper et al. 2007) and safeguards against falsely decoding meaningful trends and variability that occur based on a stretched or compressed vertical axis (see the bottom two graphs in Fig. 1). Only 15 % of the 4313 reviewed graphs had a proportional construction ratio falling between 63 and 75 %. A discrepancy of the previously mentioned magnitude may adversely affect visual analysis. As stressed by Schmid (1992), “… grid proportions are of pronounced significance as the determinants of the visual impression conveyed” (p. 28).

Other factors affecting conclusions drawn from graphs include physical length and scaling of graphs within the same article that share the same axis labels. Of the reviewed graphs, 44 % of vertical axes and 39 % of horizontal axes had different physical lengths. Horizontal or vertical axes with variable lengths present a false comparison. Visual analysis requires uniformly sized data displays for proper pattern recognition.

Accurate, legible, and the fine-grained detail of data further degrade when scaling differs. Time series graphics such as the line graph must preserve the continuity of time for useful comparisons (Robbins 2005). Differently scaled horizontal axes alter the temporal relationship between data sets meant for comparison. Scaling of the vertical axes demands the same drafting precision to produce visual parity. Sixty-nine percent of vertical and horizontal axes meant for comparison contained dissimilar scaling. Improper scaling for corresponding line graphs demonstrates poor design, violates graphic principles and standards, and makes it difficult to distinguish between the multiple curves (Schmid 1992; Schmid and Schmid 1979).

The errors of *quality features* discovered in the present study may have occurred due to a lack of standardization. Standardization refers the implementation of world-class specifications governing the fabrication and delivery of products, systems, and services (International Organization for Standardization 2014). Or more simply stated, standardization involves "… an agreed-upon way of doing something" (Spivak and Brenner 2001, p.1). The *essential structure* and *quality features* of line graphs used as the evaluation standard in the present article derived from a number of authoritative sources on graphic construction. The present error rate data may reflect a general unawareness of, or misinformation about, line graph standards. Disagreement among journal editors, reviewers, and authors as to what constitutes a standard may also account for labeling errors and ill-formed line graphs.

Cleveland (1984a) found that 30 % of all graphs reviewed in a volume of the prestigious journal *Science* contained an unacceptable amount of at least one error. In the current sample, 85 % of reviewed graphs had, at minimum, one error (i.e., violation of the proportional construction rule). The results from the present usage survey suggest that behavior analysis must address why so many errors occur with line graphs when compared to long-standing accepted construction rules both within and outside of the scientific discipline (American National Standards Institute and American Society of Mechanical Engineers 1960, 1979; American Standards Association 1938; American Statistical Association 1915; Cleveland 1993, 1994; Cooper et al. 2007; Department of the Army 2010; Giesecke et al. 2012; Harris 1999; Parsonson and Baer 1978; Poling et al. 1995; Scientific Illustration Committee 1988; Schmid 1992; Schmid and Schmid 1979).

The present survey did not examine the conditions for the excessive errors found in graph construction and labeling. Nevertheless, behavior analysis must determine how to remedy the situation. Three major sources disseminate standards: education, textbooks, and journals. Education covers colleges and universities (i.e., undergraduate and graduate programs), continuing education courses (e.g., conferences, workshops), and coursework approved for certification (e.g., Behavior Analysis Certification Board) and/or licensure. Textbooks provide information ranging from foundational concepts and principles, practitioner-driven topics and implementations, and experimental applications. And, journals distill current practices and codify history. The entities responsible for the principles of line graph construction must work in synchrony to publicize, regulate, and promote graphing standards. Next to measuring behavioral data with integrity, line graphs represent the single most important factor for analyzing, interpreting, and communicating basic and applied experimental results.

Limitations

Researchers reviewed the majority of articles through online services. The quality of the copies varied (e.g., spacing, alignment, clarity). Copy quality may have influenced some scores such as *Number on Tick Marks* and *Alignment of Axes*. With 49 and 21 % error, respectively, copy quality alone seems unlikely to have driven the majority of error. Another limitation involves journal selection. The present review followed previously established guidelines for identifying behavioral journals. Regardless, some may argue that the current sample excludes additional representative examples of behavior analytic journals.

Future Directions

Future researchers could expand upon the present research by examining a larger scope of behavioral journals and move to other psychology journals. Researchers may also wish to focus in on individual journals and determine if graphing errors changed across time. Positive findings may indicate that certain journals provide additional information regarding graph construction (e.g., editorial guidance).

Research spurred by the present study extends to decisions made through visual analysis. Would graphical decisions and analyses differ if simple line graphs contained fewer errors? For instance, sessions may have varied length in time and may not occur consecutively. The graph, nevertheless, may improperly show sessions contiguously along the horizontal axis producing false equality, labeling error, and non-representative data (i.e., three possible errors associated with session usage). How would behavior analysts and practitioners of psychological disciplines respond to the same data regraphed in time? Scaling vertical and horizontal axes to an equivalent minimum and maximum values for graphs meant for comparison may also produce different decisions.

Conclusions

The present study conducted a detailed survey of the *essential structure* and *quality features* of line graphs in behavioral journals. The data demonstrated a number of areas where errors occurred with the construction and labeling of graphs. The results also indicate *quality features* that behavior analysts can improve the visual representation of data with line graphs: following the proportional construction ratio, when comparing data scaling the vertical and horizontal axes to the same physical length, enhancing tick mark usage, and using time units instead of sessions for time series data. The science of behavior may evolve more rapidly with adherence to principles and standards of graphical excellence.

References

Alberto, P. A., & Troutman, A. C. (2013). *Applied behavior analysis for teachers* (9th ed.). Upper Saddle River: Pearson.

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

American National Standards Institute & American Society of Mechanical Engineers. (1979). *American national standard: time-series charts*. New York: ASME.

American Standards Association. (1938). *Time-series charts: a manual of design and construction*. New York: The American Society of Mechanical Engineers.

American Statistical Association. (1915). Joint committee on standards for graphic presentation. *Publications of the American Statistical Association, 14*(112), 790–797.

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. doi:10.1901/jaba.1968.1-91.

Barlow, D. H., Nock, M. K., & Hersen, M. (2009). *Single case experimental designs: strategies for studying behavior change* (3rd ed.). Upper Saddle River: Pearson.
Bowen, R. (1992). *Graph it! How to make, read, and interpret graphs*. Englewood Cliffs: Prentice-Hall.
Carr, J. E., & Britton, L. N. (2003). Citation trends of applied journals in behavioral psychology: 1981–2000. *Journal of Applied Behavior Analysis, 36*, 113–117. doi:10.1901/jaba.2003.36-113.
Catania, A. C. (1998). *Learning* (4th ed.). Englewood Cliffs: Prentice-Hall.
Cleveland, W. S. (1984a). Graphs in scientific publications. *The American Statistician, 38*, 261–269. doi:10.1080/00031305.1984.10483223.
Cleveland, W. S. (1984b). Graphical methods for data presentation: full scale breaks, dot charts, and multibased logging. *The American Statistician, 38*, 270–280. doi:10.1080/00031305.1984.10483224.
Cleveland, W. S. (1993). *Visualizing data*. Summit: Hobart Press.
Cleveland, W. S. (1994). *The elements of graphing data*. Summit: Hobart Press.
Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science, 229*(4716), 828–833. doi:10.1126/science.229.4716.828.
Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River: Pearson Prentice Hall.
Critchfield, T. S. (2002). Evaluating the function of applied behavior analysis: a bibliometric analysis. *Journal of Applied Behavior Analysis, 35*, 423–426. doi:10.1901/jaba.2002.35-423.
Department of the Army. (2010). *Standards of statistical presentation*. Washington DC: Department of the Army.
Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. New York: Appleton.
Few, S. (2009). *Now you see it: simple visualization techniques for quantitative analysis*. Oakland: Analytics Press.
Gast, D. L. (2010). *Single subject research methodology in behavioral sciences*. New York: 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. 166–233). New York: Routledge.
Giesecke, F. E., Hill, I. L., Spencer, H. C., Mitchell, A. E., Dygdon, J. T., Novak, J. E., Lockhart, S. E., & Goodman, M. (2012). *Technical drawing with engineering graphics* (14th ed.). Upper Saddle River: Pearson.
Gutentag, S., & Hammer, D. (2000). Shaping oral feeding in a gastronomy tube-dependent child in natural settings. *Behavior Modification, 24*, 395–410. doi:10.1177/0145445500243006.
Harris, R. L. (1999). *Information graphics: a comprehensive illustrated reference*. New York: Oxford University Press.
International Organization for Standardization (2014, June 4). What are standards? Retrieved from http://www.iso.org/iso/about/discover-iso_meet-iso/about.htm
Johnston, J. M., & Pennypacker, H. S. (1980). *Strategies and tactics of human behavioral research*. Hillsdale: Erlbaum.
Johnston, J. M., & Pennypacker, H. S. (2009). *Strategies and tactics of behavioral research* (3rd ed.). New York: Routledge.
Kangas, B. D., & Cassidy, R. N. (2010). Requiem for my lovely. *Journal of the Experimental Analysis of Behavior, 95*, 269. doi:10.1901/jeab.2011.95-269.
Katzenberg, A. C. (1975). *How to draw graphs*. Chicago: R.R. Donnelley & Sons Company.
Kazdin, A. E. (2011). *Single-case research designs: methods for clinical and applied settings* (2nd ed.). New York: Oxford University Press.
Kennedy, C. H. (2005). *Single-case designs for educational research*. Boston: Allyn & Bacon.
Kriglstein, S., Pohl, M., & Smuc, M. (2014). Pep up your time machine: recommendations for the design of information visualizations of time-dependent data. In W. Huang (Ed.), *Handbook of human centric visualization* (pp. 203–225). New York: Springer.
Kubina, R. M., Kostewicz, D. E., & Datchuk, S. M. (2008). An initial survey of fractional graph and table area in behavioral journals. *Behavior Analyst, 31*, 61–66. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2395347/.
Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: brief review and guidelines. *Neuropsychological Rehabilitation: An International Journal, 24*(3–4), 445–463. doi:10.1080/09602011.2013.815636.
Lattal, K. A. (2004). Steps and pips in the history of the cumulative recorder. *Journal of Experimental Analysis of Behavior, 82*, 329–355. doi:10.1901/jeab.2004.82-329.
Malott, R. W., & Shane, J. T. (2014). *Principles of behavior* (7th ed.). Upper Saddle River: Pearson.
Mayer, G. R., Sulzer-Azaroff, B., & Wallace, M. (2014). *Behavior analysis for lasting change* (3rd ed.). Cornwall-on-Hudson: Sloan.

National Institute of Standards and Technology (2014). Unit of time (second). Retrieved from http://physics.nist.gov/cuu/Units/second.html
Norborg, J., Osbourne, S., & Fantino, E. (1983). Duration of components and response rates on multiple fixed-ratio schedules. *Animal Learning & Behavior, 11*, 51–59. doi:10.3758/bf03212306.
Parsonson, B., & Baer, D. (1978). The analysis and presentation of graphic data. In T. Kratochwill (Ed.), *Single subject research* (pp. 101–166). New York: Academic.
Pierce, W. D., & Cheney, C. D. (2013). *Behavior analysis and learning* (5th ed.). New York: Psychology Press.
Poling, A. D., Methot, L. L., & LeSage, M. G. (1995). *Fundamentals of behavior analytic research*. New York: Plenum Press.
Preston, R. (1994). Choice in the time-left procedure and in concurrent chains with a time-left terminal link. *Journal of the Experimental Analysis of Behavior, 61*, 349–373. doi:10.1901/jeab.1994.61-349.
Ramirez, I. (1997). Carbohydrate-induced stimulation of saccharin intake: yoked controls. *Animal Learning & Behavior, 25*(3), 347–356. doi:10.3758/bf03199092.
Robbins, N. B. (2005). *Creating more effective graphs*. New York: Wiley.
Schmid, C. F. (1992). *Statistical graphics: design principles and practices*. New York: Wiley.
Schmid, C. F., & Schmid, S. E. (1979). *Handbook of graphic presentation* (2nd ed.). New York: Wiley.
Scientific Illustration Committee. (1988). *Illustrating science: standards for publication*. Bethesda: Council of Biology Editors.
Skinner, B. F. (1938). *The behavior of organisms: an experimental analysis*. Cambridge: B.F. Skinner Foundation.
Spivak, S. M., & Brenner, F. C. (2001). *Standardization essentials: principles and practice*. New York: Marcel Dekker.
Spriggs, A. D., & Gast, D. L. (2010). Visual representation of data. In D. L. Gast (Ed.), *Single subject research methodology in behavioral sciences* (pp. 166–233). New York: Routledge.
Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire: Graphics Press.
Tufte, E. R. (1990). *Envisioning information*. Cheshire: Graphics Press.
Vargas, J. S. (2013). *Behavior analysis for effective teaching* (2nd ed.). New York: Routledge.
Wainer, H. (2005). *Graphic discovery: a trout in the milk and other visual adventures*. Princeton: Princeton University Press.