2017

Thinking Critically About Data Displays

Luke Duesbery

Jenelle Braun-Monegan Stone

Kimy Liu

Jan McCoy

Thinking critically about data displays

Luke Duesbery[a], Jenelle Braun-Monegan[b], Kimy Liu[c] and Jan McCoy[d]

[a]School of Teacher Education, San Diego State University, San Diego, CA, USA; [b]School Teaching and Learning, Pacific University, Eugene, OR, USA; [c]College of Education, Kinesiology, and Social Work, Teacher Education, Education Specialist Credential Program, California State University Stanislaus, Turlock, CA, USA; [d]Oregon Department of Education, Salem, OR, USA

ABSTRACT

The quality of a data display can have an impact on the interpretation of those data. A survey of the literature indicates that data displays can vary in quality of accuracy, clarity, and efficacy. In this study we develop and apply an evaluative rubric to graphs in a sample of six education journals: three research and three practitioner. Results indicate that graph quality is typically high in educational journals, however, in practitioner oriented journals issues around graph clarity and efficacy should be addressed. Common error patterns are pinpointed, and four recommendations are made to authors and editors: focus on meaningful labels, increase amount of data displayed, portray multiple relationships, and elaborate with supporting text.

KEYWORDS

Data graphics; graphs; visual literacy; data displays; specifiers; research use; research to practice; information science

Thinking critically about data displays

The power of using graphs to facilitate decision-making has long been valued by those charged with making critical decisions. King Louis XVI described graphs as speaking *all* languages (Playfair, 2005). To this day, visual displays are thought to be an efficient and beneficial method of presenting data for decision-making. Data graphics are common in newspapers, periodicals, textbooks, and even research literature. The axiom 'A picture is worth a thousand words' illustrates the power of the efficiency of displaying information visually and even captures the nuanced idea that graphs, or more generally information graphics, transcend the written word. Graphs communicate information that cannot be fully described by words, just like a picture has details that cannot be captured fully by a verbal description.

The Association of College and Research Libraries (ACRL), a division of the American Library Association (ALA), describes visual literacy as a '…set of abilities that enables an individual to effectively find, interpret, evaluate, use, and create images and visual media' (Association of College and Research Libraries, 2011). Specifically, as it relates to the study described here, the ACRL Standards for Higher Education (2011) posits that a visually literate person can be taught to evaluate graphs, charts and data models to

determine their accuracy and reliability (see Standard Four). Creating meaning from a data graphic requires specific skill (Duesbery, Werblow, and Yovanoff, 2011), and can be taught. In the seminal work, Visual Literacy argues that while visual literacy is old concept, the meaning continues to evolve. Where once it may have focussed on, for example, high school students learning to interpret a painting, its modern meaning is more broad. The International Visual Literacy Association (IVLA) remarked that because of the diversity of underlying disciplines in the area of study, that each scholar might produce his or her own definition (Avgerinou, 2012). In that spirit, for the purposes of this study we define visual literacy as a body of skill that one can learn, build, and use to better interpret images of all kinds to find deeper meaning. One key subfield of visual literacy is the method by which we should interpret and value images rooted in the presentation of data, designed to elicit deeper contemplation.

Seminal works in the field of interpreting data graphics by scholars such as Tufte (1983), Tukey (1972) and Wainer (2015, 2011, 2005, 1997) tend to rest on the premise that there may be a single set of rules by which we can create *excellent* data displays. Indeed, this is an appealing perspective. The purpose of this study is to first determine what leading scholars tell us constitutes a good graph, and to then develop a commensurate and meaningful rubric to use for evaluating the quality of graphs. We then apply this rubric in a pilot study to examine journals in the field of education.

Relevance

A well-designed graph can significantly improve the communication of content, and a poorly designed graph, in contrast, can lead to confusion. Understanding the worth of graphs in research journals is important. Specifically, in the field of education, good decision-making in schools hinges on using data well, and extracting useful information from data is facilitated by interacting with data graphics.

Data displays are of concern to educational decision-makers for several reasons. There is mounting pressure in the field of education to emphasize a scientific data-driven decision-making model. Cizek (2005, p. 38) noted, ‘Not only is more information about student performance available, but it is increasingly used as a part of decision-making’. These high-stakes data-based educational decisions can take many forms and be based on different data sources. For some, high school graduation is based on student performance data. For others, the very future of the school remaining open rests on interpreting school data. Making appropriate decisions is more important than ever. This emphasis on data interpretation leads directly to the increased relevance of data graphics that support that endeavour. Indeed, graphs are common in educational research now and historically (Smith, Best, Stubbs, Archibald, & Roberson-Nay, 2002). A more recent U.S. Department of Education report concluded that when presented with data, teachers often respond based on prior knowledge rather that the data, and when presented with a histogram roughly a third of teachers could not find clear errors (Means, Chen, DeBarger, & Padilla, 2011, p. 27).

There is also evidence to suggest that educators have difficulty interpreting data displays (Duesbery, Werblow, Yovanoff, 2011; Goodman & Hambleton, 2004; Wainer, Hambleton, & Meara, 1999; Hambleton & Slater, 1996). If educators are expected to make high-stakes decisions with data, they need to understand the data presented to

them. In a 2004 survey conducted for CoSN by Grunwald & Associates, among barriers to data driven decision-making cited by school decision makers were: lack of training in the use of data (50%), lack of understanding of what to do with the data (39%), and displays too complicated to understand (22%).

Finally, given the abundance and relative ease of modern data-display development tools, virtually anyone can display data graphically, but not necessarily well. It is often true that default formats available in graphing programs are not the best choices for data display and that modifying graphs beyond their default settings can be challenging because very few educational researchers receive instruction on displaying data. Understanding what constitutes a high quality graph is the first step in creating higher quality graphs.

The practical implications of this study are clear. Educators need to use data to inform decision-making and often use graphical data-displays to do so. Gaining a better understanding of what constitutes a good graph, and understanding how well authors in field journals adhere to good graphing precepts, gives the reader insight into how we can improve the quality of the displays and consequently the quality of the decision-making.

It should be noted that this study is not unique. Similar studies in the medical sciences (Cooper, Schriger, Close, 2002; Cooper, Schriger, Tashman, 2001) and psychology (Smith et al., 2002) have explored the quality of data graphics in journals, however, not with the same attention to precision of measurement. Studies thus far have closely examined the more superficial elements of graph construction, rather than the deeper more meaningful attributes examined here.

Precepts of good graphing

Determining the precepts of good graphing is no small undertaking, and there are certainly myriad methods one might to employ to arrive at such a list. In this study, we choose to examine the work considered seminal – texts by Edward Tufte, Howard Wainer and John Tukey. Each of these scholars is prolific in the area of data graphics, and are generally accepted as authorities. Thus, our search for what constitutes *excellent* graphing comes not from any empirical method, but rather from a synthesis of popular and respected literature. This synthesis reveals three overriding precepts that drive excellent graphics – graphical accuracy, graphical clarity and graphical efficacy – each of which is discussed in turn.

Graphical accuracy

In a data graphic, there is a tacit assumption relationships are accurately displayed and convey only accurate, directly comparable reflections of the data relationships (Cleveland, Harris, & McGill, 1983). Along with precise placing of data points, an accurate data display also features clear and unambiguous titles, labels and descriptions when appropriate. The choice of scale should display the full range of data, and facilitate comparison between and among data points and trends. In contrast, a data display that lacks accuracy may have errors in data placement or labels, or might carry ambiguous information without supporting text. An inaccurate data display may lead

to inaccurate inferences about relationships, and may even purposely mislead the graph reader. An often cited inaccuracy is the use of inconsistent units on axes, or the manipulation of scale to alter the appearance of trends in data (Wainer, 2005).

Graphical clarity

In keeping with maintaining clarity in data graphics, Tufte (1983, p.107) suggests avoiding the use of legends. He instead calls for the direct labelling of data on the graph. Further, he supports the removal of redundant axes, guiding tick marks, and what he calls *Chart Junk*. To measure graph clarity, he employs a data to ink ratio (Tufte, 1983, p. 93). This ratio speaks to the proportion of a graphic's ink devoted to the (non-redundant) display of data:

$$\text{Data Ink Ratio} = \frac{\text{Total ink used to display the data}}{\text{Total ink used to display the graphic}}$$

Self admittedly borrowing from Tufte, Wainer (1997) tells us that the goal of displaying information graphically to communicate is threefold: Reduction of text, clarity of focus and highlighting importance of a particular aspect. These tenets are mirrored in the relevant chapter of the Publication Manual of the American Psychological Association (2010, p. 152) co-authored by Wainer. In achieving these communication goals, the good data display becomes efficient; it conveys a message of complex and multivariate data, and does so in the simplest manner possible. In this respect, perhaps Wainer (1997, p. 3) stays closer to the Tukey ideal of graphical simplicity. Evidenced by the title of his work, *Visual Revelations*, his central point is that 'revelation accompanies simplicity'. Meaningful data display must be represented clearly to add value.

Tufte (1983, p. 56) also directs creators of data graphics to avoid multi-dimensional portrayals of data to minimize dimensional exaggeration. In contrast to a two-dimensional graph that uses bars to represent a value, a three-dimensional graph uses volume, thus amplifying our interpretation of the value. Despite this recommendation to avoid the multi-dimensional, creators of graphs often turn inappropriately to a three-dimensional portrayal of data. For example, with bar graphs imagine the tall rectangular prisms often exported from Microsoft Excel. While these graphs may certainly be visually pleasing, they nonetheless carry with them an exaggeration of the data. While not intentionally deceptive, these graphs are misleading in that the three-dimensional object remains the same height as its two-dimensional counterpart but adds depth. Thus, metaphorically speaking, the volume of the prism is conveyed as the data message, not simply the height. In truth, even the two-dimensional object in a chart might be seen as an exaggeration, since it carries both height and width. In this case the area of the object is portrayed as the data metaphor, not simply the height. As a measure of graphical precision, Tufte (1983, p. 57) employs the Lie Factor:

$$\text{Lie factor} = \frac{\text{Portrayed size of effect shown in graphic}}{\text{Actual size of effect in data}}$$

Consider Figure 1 in which two graphs each representing the same data are displayed. In the upper panel, data are displayed in two dimensions, and in the lower panel in three dimensions. The lower panel, thus, exaggerates the data. To read more

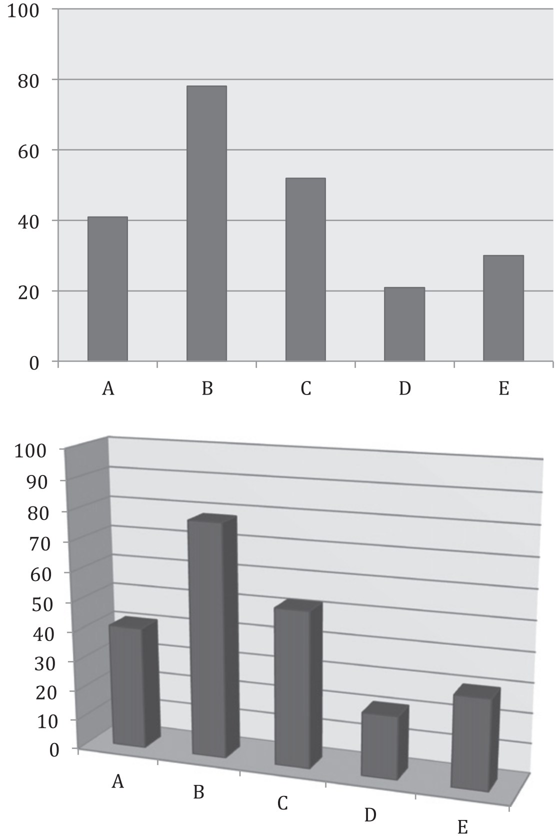

Figure 1. Example of high (top) and low (bottom) clarity in graphs.

about the lie factor, look for Tufte's 1983 seminal work, *The Visual Display of Quantitative Information*.

Graphical efficacy

A graph has the potential to be far more than simple representation of tabular data. In an early manuscript, Tukey (1972) differentiated three kinds of data graphics. The first and the simplest form is a graph that takes the place of a table. This *simple graph* does nothing more than portray data simply, and promotes nothing more than reading unadorned values. The second kind of graph, what Tukey calls the *propaganda graph*, is used to convey a particular message to the graph reader. In this case, the graph directs the reader towards a pre-determined conclusion. The third and the most valuable graph is the *analytic graph*, designed to elicit exploration, contemplation and comparison.

This third kind of graph, one that involves visual data analysis, lies at the root of modern exploratory data analysis (Friedman & Stuetzle, 2002), and has been attributed to a trend that began in the 1960s called Direct Manipulation of Graphics (Cleveland, 1985). Exploring data visually seems a viable method of analyzing data, and stands in direct contrast with the more traditional confirmatory analyses (Wainer, 2005). Bertin (1983) similarly isolated three potential functions of a graph. At its most base

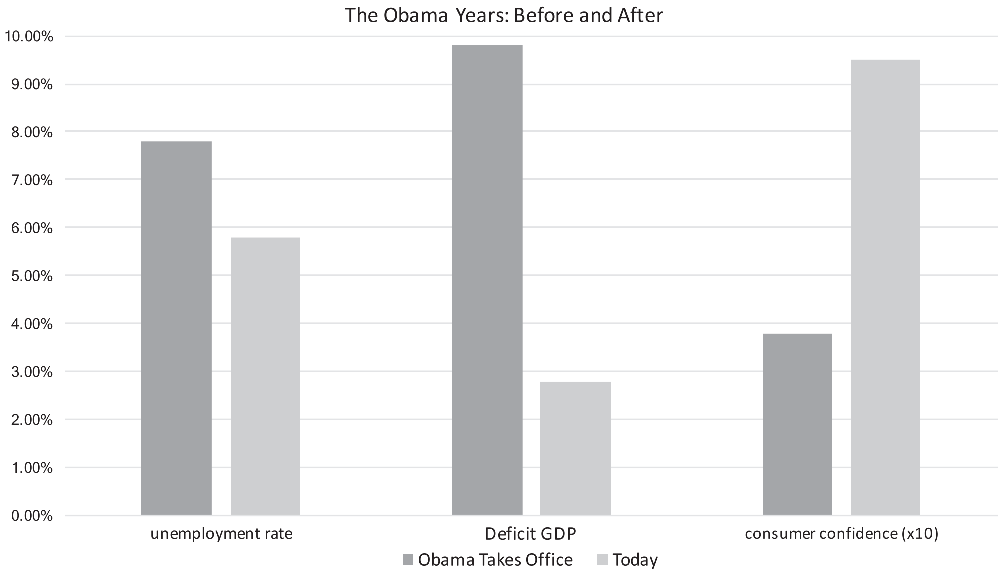

Figure 2. Example of a propaganda graph. A hand selected set of data are presented to make the point that President Obama was successful over his eight-year term.

level a graph stores information. Beyond this, it can be used as a vehicle for communication. Lastly, the graph can be used for processing – again – the *analytic* graph. As Tufte puts it, graphics *reveal* the data; they do not simply present it (p. 51). In a particularly effective graph, relationships that may not have been transparent come to the surface, and surrounding text is used to complement the data portrayed.

Like Tukey, Wainer cautions us against creating graphs that merely portray data to a single end – the propaganda graph. Propaganda is information, real or not, used to mislead. The propaganda graph similarly misleads; the message from the data graphic is predetermined, and data is populated to carry forward the message as if it was discovered with data, when in reality the graphic was fashioned with a point of view in mind. The question is which comes first? The data and then the idea (natural), or the idea and then the data (deceptive). For example, let's assume I am a Democrat and support Barack Obama. I want to make a graphic to display his success (the idea). Based on that notion, I hand pick data that supports my contention and create the display (see Figure 2).

It is not our mission to manipulate the display of data to suit a particular purpose, although this is clearly possible. With good data display, Wainer (1997, p. 57) writes, 'We can be forced to discover things from a graph without knowing in advance what we were looking for'. Our mission, instead, is to be honest in our representation of data, to allow the user to draw inferences that might not have been arrived at without the display, and thereby critically inform decision-making.

Like in Tukey's *analytic graph*, Tufte posits that in excellent data graphics we should strive to present as much information as the user can manage, which invariably leads to the creation of multivariate data graphics. Unlike Tukey, both Tufte (1983) and Wainer (1997) envision graphs replete with data. The more entries in the data matrix per square inch of graph space, the higher are the density of data, and the more

effective is the graph. Tukey can be seen to have a more basic approach to the utility of data graphics. 'The great usefulness of graphs is their portrayal of the gross and easily visible. They should not be used for detail' (Tukey, 1972, p. 297). Tukey believed that fine detail belonged in tables, for fast viewing. Gross trends visible in data belonged in his *analytic graphs*. Tukey did believe it was possible to combine the two in a single display, but clearly his emphasis lay in a different direction than that of Tufte. 'The idea that if 20 or perhaps 50 points are good, then 200 or 500 are better is almost always wrong' (Tukey, 1972, p. 301). Tukey felt that too much data reduces the utility of a graph. Critical to this discussion is the choice of data graphic. A line chart, for example, is ideal for displaying high numbers of data points, across time. In contrast, a bar chart is not suited well for copious amounts of data, but is well suited to disaggregating data by a small number of groups. The appropriate choice of graph type can also lead to an effective display; one that is suited to the nature of the data relationship, and the quantity of data portrayed.

Summary

To determine data display excellence within educational journals, this study explicitly tests for the *presence* and *quality* of three graphical quality indicators: (1) Graphical accuracy, (2) Graphical clarity and (3) Graphical efficacy.

Method

Research questions

In this descriptive study, we posed two research questions as preliminary to further research in the area. First, what do leading scholars tell us constitutes a good graph, and second, perhaps more importantly, to what degree, in education journals, do graphical representations of data adhere to these precepts?

Sample

In this preliminary study, we selected a relatively small sample of articles from the leading professional organizations in general education, special education and educational administration. We purposefully selected three leading journal that were research oriented, and three that were rooted in practice. We began with journals published by the American Educational Research Association (AERA), but were forced to branch out to others because of some difficulties encountered. For example, *Education Researcher* was considered, but rejected because of its focus on presenting comments, book reviews and organizational notes. The journal stood out as something quite different than the others, as it presented very little new research. In addition, *Educational Leadership*, published by the Association for Supervision and Curriculum Development (ASCD) had no graphs in the last 64 issues, and was similarly dismissed from the analysis. The final selection of journal titles included is presented in Table 1.

Because journals varied in the number of pages in each issue, we chose to analyze all graphs contained within the last 200,000 words of each journal. This meant that we

Table 1. Sample of journals analyzed.

	Title	Publisher	Articles	Graphs
Research	American Educational Research Journal	American Educational Research Association	16	19
	Exceptional Children	Council of Exceptional Children	12	12
	Educational Evaluation and Policy Analysis	American Educational Research Association	10	18
Practice	Teaching Exceptional Children	Council of Exceptional Children	24	3
	The School Administrator	American Association of School Administrators	45	2
	Action in Teacher Education	Association of Teacher Educators	26	12

could more directly compare simple frequencies, in addition to the more relevant issues of quality. For some journals, this meant we needed to look at as many as nine issues, but for others as few as two. When counting words and pages, only submitted articles were considered. Association news, advertisements, book reviews, commentaries, regular editorial columns and the like were not included in word counts as they typically do not contain any graphs.

Instrument

The three graphical quality measures that emerged from our literature review were graphical accuracy, clarity and efficacy. We further subdivided each into three sub categories. *Accuracy* into data points, labels and axis/scale, *clarity* into chart junk, font/spacing and data-ink use, and *efficacy* into depth of relationships, relationship with surrounding text, and choice of display type. Each of these nine subcategories was evaluated on a three point scale, thus a maximum score of 27 was possible for each and every graph. More detailed operationalization of these categories is provided in the scoring rubric in Table 2. The 27 point score was scaled to an index score out of 100 for easier interpretation.

In applying the rubric, the lead author first analyzed all graphs independently, and then trained a second coder to establish reliability. A 30% sample of those was given to the second coder for repeated independent analysis. Inter-rater reliability moderate ($r = .78$), leading to a high degree of certainty that the rubric produced moderately objective data.

Analyses

Simple descriptive statistics and frequencies form the foundation of analyses used to answer our research questions. With such a small sample of journals it seems unwise to conduct tests of significance for differences in means. Hierarchical linear modelling would likely be the most appropriate methodological approach, however, it is not possible with the small sample. Effect sizes are made interpretable with the use of a scaled index score out of 100. First author evaluations were used in all analyses.

Results

In all, 142 articles were reviewed from the six journals, 38 from research and 104 from practitioner journals. The research journals had an average of about 13 articles in the

Table 2. Graph quality scoring rubric.

		1 point	2 points	3 points
Accuracy	Data points	(a) The representation of data is misleading or confusing and *will* lead to inaccurate conclusions.	(a) The representation of data *might* lead to inaccurate conclusions.	(a) The display accurately depicts the data and leads to the conclusions inherent in the data.
	Labels	(b) Labels are misleading or missing, or the data displayed is inaccurate.	(b) Labels may be somewhat ambiguous or some data *may* be inaccurately placed	(b) Labels are clear and meaningful and the numbers displayed appear correct
	Axes/scale	(c) Axis scale distorts the data; axes use inconsistent scaling or unit choice detracts from accuracy.	(c) Axis scale or unit *may* distort data to support only one side of an argument; axes may not use consistent scaling.	(c) Axis scale supports the range of data.
Clarity	Chart junk	(a) Superfluous graphics obfuscate the data (i.e. *chart junk*).	(a) The graph employs non-data graphics that may confuse the reader; e.g. rendering of three dimensions	(a) Data is plainly visible, and displayed with precision.
	Font/spacing	(b) Font and spacing choices make reading labels difficult. Legends obscure data interpretation.	(b) Choice of font/spacing *may* make the graph more difficult to read. Legends may create difficulty.	(b) Font/spacing leads to a more readable display. Labels are applied directly to data when appropriate.
	Data-ink	(c) A single variable is displayed when multiple variables *could* have been displayed; excessive non data-ink is present.	(c) More data might be displayed without loss of quality; non-data ink reduces the clarity of the graph.	(c) A high data-to-ink ratio maximizes the display of information.
Efficacy	Depth	(a) Comparisons inherent in the data are not apparent in the data display.	(a) Additional variables might be added to add depth.	(a) Encourages a deeper understanding of the data; the display facilitates comparison within and across data
	Text	(b) The data displayed is not related to surrounding text.	(b) The data may only be tangentially supported by surrounding text.	(b) The display is supported by surrounding text and complements text.
	Chart	(c) An inappropriate chart type is used to display the data; data displayed would be better suited for a simple table.	(c) Data might be better displayed in an alternate graph type.	(c) Graph choice is appropriate for the data, e.g. line for time series or trend, bar for comparison or trend, etc.

last 200,000 words. In contrast, practitioner journal articles were much shorter, on average having about 35 articles for the same 200,000 words. This trend was mirrored in the frequency of graphs and tables. Research journals had, on average, about 59 tables and 16 graphs per 200,000 words, while practitioner journals had on average about 18 tables and 6 graphs in the same 200,000 words.

Within journal types, the number of graphs was also highly variable. Research journals had between 12 and 19 graphs over the last 200,000 words, and practitioner journals had between 2 and 12. The number of tables in the research journals was fairly stable, ranging between 57 and 63, but was remarkably variant within the practitioner journals, ranging from 0 tables in *The School Administrator (TSA)* to 31 tables in *Action in Teacher Education (ATE)*. On average, the research journals had longer articles, and more tables and graphs per article.

Despite differences in frequencies, it is still possible to report the results of our quality indicators. The reader will remember that we used a 27 point rubric, nine points

Table 3. Summary of quality indicators (0–100 Index Score).

	Journal	Articles	Tables	Graphs	Accuracy	Clarity	Efficacy	Total
Research	AERJ	16	58	19	75.4	78.4	79.5	77.8
	EEPA	10	63	18	87.8	90.0	88.9	88.9
	EC	12	57	12	88.9	84.0	87.7	86.8
	Mean	12.7	59.3	16.3	84.0	84.1	85.4	**84.5**
Practice	ATE	26	31	12	81.5	70.4	77.8	76.5
	TEC	33	22	3	81.5	70.4	77.8	76.5
	TSA	45	0	2	88.9	88.9	66.7	81.5
	Mean	34.7	17.7	5.7	84.0	76.6	74.1	**78.2**

Table 4. Distribution of graph type by journal orientation.

	Line	Bar	Box	Scatter	Picture	Pie	Total
Research	26	12	7	2	1	1	49
Practice	0	4	0	3	0	10	17

devoted to evaluating each of accuracy, clarity and efficacy. The journal with the highest average total graph quality index score (0–100) was 88.9 for *Education Evaluation and Policy Analysis (EEPA)*, followed closely by *Exceptional Children (EC)*, with 86.8. *Teaching Exceptional Children(TEC)* and *Action in Teacher Education (ATE)* together shared the lowest overall graph quality index score of 76.5. A summary of quality indicators for each journal is provided in Table 3.

In general, research-oriented journals tended to have higher overall graph quality. Research journals had an average index score of 84.5 (shown bold in Table 3), and practitioner journals had an average score of just 78.2 (shown bold in Table 3). In terms of graph type, differences were also found based on the journal orientation. The sample of research journals had 26 line graphs, whereas the practitioner journals had none. The trend was also evident in the frequency of box-and-whisker plots used. The research journals had seven box-plots, whereas the practitioner journals, again, had none. Practice-oriented journals relied on more pie graphs with a count of 10 whereas in research-oriented journals only one pie graph was found in this sample. Frequencies of graph type are summarized in Table 4.

Discussion

In terms of simple frequencies, research-oriented journals were far more likely to include graphs and tables. Compared to practitioner journals, there were three times more graphs in the same size sample of research journals, and nearly four times as many tables. This itself, speaks volumes. Authors of articles in research journals place more emphasis on the graphical display of quantitative information, whether by graph or table. The relative infrequency of graphs and tables in practitioner journals came as a surprise to us. Two of the three practitioner journals selected had just two or three graphs in the last 200,000 words.

Even though the quality of graphs varied, in general they were all quite high. On our 100 point scaled index score, graph scores averaged 77 ($SD = 9$), and ranged from 63 to 100. A moderate score for a graph, where each of the nine sub-categories was scored the middle value, would be 67. Very few graphs were given the lowest rating

in any of the nine sub-categories. This means, in general, graphs in this sample of educational journals were of moderate to high quality.

Given our three graphical quality indicators – accuracy, clarity and efficacy, differences were found between journal types. There were few differences in the area of graphical accuracy: appropriately placing data points, selecting an appropriate scale range, and including appropriate labels. Graphs in practitioner and research journals were of nearly the same accuracy. This trend, however, was not evident in graph clarity or graph efficacy.

Graphs in research journals tended to be clearer. Clarity, here, refers to minimizing irrelevant graphics (chart junk), employing clear font and spacing use, and portraying an optimal amount of data given the space used. On average, graphs in research journals were about ten scale points higher in clarity than those found in practitioner journals (research = 84.1, practitioner = 76.6). Graphs in research journals were also more effective in communicating trends in data. Efficacy, here, refers to displaying trends in data beyond simple data points, supporting data with meaningful complementary contextual information, and choosing an appropriate graph type. On average, graphs in research journals were about ten scale points higher in efficacy than those found in practitioner journals (research = 85.4, practitioner = 74.1). Differences in the areas of clarity and efficacy together contributed to an overall difference of about six scale points in average graph quality (research = 84.5, practitioner = 78.2).

Smith et al. (2002) found that 'harder' sciences, such as physics and chemistry, tended to produce simpler graphs. They hypothesized that the increased complexity of the data analysis in 'softer' sciences, such as psychology and education, led to this difference. This idea of a simpler graph might translate into the *clarity* variable examined here. Our study pointed to higher clarity in the research-oriented journals. However, these routinely used more complex statistical analyses than did the practitioner-oriented journals. Thus, while we found similar results to Best et al., we could not attribute this difference to complexity of the data analysis.

Recommendations

As aforementioned, graph quality tended to be quite high on our 27-point rubric that emerged from the literature. Almost all graphs were highly rated in terms of accuracy. This is good news, since accuracy is probably the starting point of good graphing. Beyond this, however, issues of clarity and efficacy rise to the surface. The most common shortcoming we found in the area of clarity was the use of inappropriate labels. Specifically, many authors used distracting legends, or failed to label data with meaningful labels.

> *Recommendation 1: Focus on using meaningful text labels*. We found that many graphs were fairly simple in nature. They tended to display only a few data points over a large space, where many more could have been displayed. As distinct from issues of efficacy, clarity of data points involves maximizing the data-to-space ratio.

> *Recommendation 2: Increase the amount of data portrayed*. The most common shortcoming we found in the area of efficacy was the display of unidimensional data where multidimensional data might have been displayed. Often this could have been achieved by a simple grouping or reordering of data points. Trends within trends might have been better displayed.

Recommendation 3: Consider portraying multiple relationships. We also found that a high number of graphs failed to fully employ the use of supporting text. For example, titles tended to be abbreviated, where they could have easily supported data better by alerting the graph reader to relationships.

Recommendation 4: Consider elaborating supporting text elements. Illustrating data relationships through graphs can provide insight into trends that might not otherwise be seen. In our opinion, practitioner journals need more, and better, graphs. In most areas, the graphs in practitioner journals tended to come up short when compared to those in research journals. But, often little or no formal training in the area of statistics is required for interpreting graphs, making them ideal for a practitioner audience. We feel this disconnect deserves attention.

Recommendation 5: Build visual literacy skills. Commensurate with building higher quality graphs (the aforementioned four recommendations), we should also consider the consumer of the data graphic. One might say the quality of the graph is in the eye of the beholder, so we should devote attention to building graph reading skills, indeed visual literacy, in our teachers and school administrators.

Limitations, future directions and conclusions

This preliminary study in the area of data graphics quality helps us to understand where we might go next. Several weaknesses prevent us from making the overarching inferences we desire, towards improving data displays in journals. We are keenly aware, for example, that the results of this study may be a product of editorial or publisher decision-making. Powers outside the authors influence have the ability to dictate graphical specifications that may not be in the best interest of the author. This may be especially prevalent in the frequency of graphs found – as printing graphics is more expensive than printing text. Results are couched in editorial decision-making, but still give us insight.

In terms of instrument quality, the rubric developed and used in this study had a moderate degree of inter-rater reliability. The validity of the instrument is evidenced by the research-derived rating categories. Could it be used elsewhere or again? We think yes. As we move to further refine this instrument for subsequent studies, we have plans to focus on generating more refined sub-category definitions of graph quality, which will lead to higher reliability. Unfortunately, obtaining a measure of concurrent validity is not possible, because we know of no other similar instrument. We are considering paralleling the qualitative instrument devised here with a measure of graph reading performance, which will yield results that might be used towards this end.

The field of perception in psychology provides a body of research addressing what Kosslyn refers to as *specifiers* (1994). Specifiers are dimensions of a display used to represent data, i.e. axes, legends, tick marks, lines, dots and bars. In general, researchers in perception address these specifiers in relative isolation. The interested reader should consult the work of Kosslyn (1994, 1985), Cleveland (1985), Carswell (1992) and Cleveland & McGill (1985, 1984) for more detail. While certainly a worthy endeavour, it was not the purpose of this study to examine these specifiers in detail, but rather to examine more broadly defined qualitative aspects of graphs. Future research might

concentrate on the prevalence and impact of specifiers found in data graphics, and the degree to which each might lead to *excellence* in data display.

Finally, we admit that the age of print graphics may be coming to an end, and with the movement to digital media the nature of the static graph may also change. An era when the user can select graphical elements and renderings on demand is upon us. But, perhaps this stresses the importance of understanding what a quality graph looks like, and how one might read (or construct) one.

We set out to determine to prevalence and quality of data graphics in education journals. Our instrument emerged from the work of leaders in the field. Results here point to a call for increased vigilance in the areas of data clarity and efficacy, especially for authors who publish in practitioner-oriented journals. Four straightforward recommendations – focus on meaningful labels, increase amount of data displayed, portray multiple relationships and elaborate with supporting text – provide graph creators with the means to enhance graph quality.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Jenelle Braun-Monegan http://orcid.org/0000-0003-0143-1649
Kimy Liu http://orcid.org/0000-0002-2016-8033
Jan McCoy http://orcid.org/0000-0001-6519-7148

References

American Psychological Association (2010). *Publication manual of the American Psychological Association*. Washington D.C.: author.

Association of College and Research Libraries (2011, October). *Association of College and Research Libraries visual literacy competency standards for higher education.* Retrieved from http://www.ala.org/acrl/standards/visualliteracy.

Avgerinou, M.D. (2012, August). *What is "visual literacy?"*. Retrieved from http://ivla.org/new/what-is-visual-literacy-2/.

Bertin, J. (1983). *Semiology of graphics*. Madison, WI: University of Wisconsin Press.

Carswell, C. M. (1992). Choosing specifiers: An evaluation of the basic tasks model of graphical perception. *Human Factors*, *34*, 535–554.

Cleveland, W. (1985). *The elements of graphing data*. Monterey, CA: Wadsworth.

Cleveland, W. Harris, C., & McGill, R. (1983). Experiments on graphical perception. *Bell System Technical Journal*, *62*, 1659–1674. doi: 10.1002/j.1538-7305.1983.tb03504.x

Cleveland, W. & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, *79*, 531–554. doi: 10.1080/01621459.1984.10478080

Cleveland, W. & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science*, *229*, 828–833. doi: 10.1126/science.229.4716.828

Cizek, G. (2005). High stakes testing: Contexts, characteristics, critiques, and consequences. In R. P. Phelps (ed.), *Defending standardized testing* (pp. 23–54). Mahwah (NJ): Lawrence Erlbaum Associates.

Cooper, R. J., Schriger, D. L., & Close, R. J. (2002). Graphical literacy: The quality of graphs in a large-circulation journal. *Annals of Emergency Medicine*, *40*, 317–322. doi: 10.1067/mem.2002.127327

Cooper, R.J., Schriger, D.L., Tashman, D.A. (2001). An evaluation of the graphical literacy of Annals of Emergency Medicine. *Annals of Emergency Medicine*, *37*, 13–19. doi: 10.1067/mem.2001.111569

Duesbery, L., Werblow, J., & Yovanoff, P. (2011). Graphical literacy mediates the interaction of decorative dimensionality and cognitive demand in graph comprehension. *Journal of Educational Computing Research*, *45*, 75–93. doi: 10.2190/EC.45.1.d

Friedman, J., & Stuetzle, W. (2002). John Tukey's work on interactive graphics. *The Annals of Statistics*, *30*, 1629–1639. doi: 10.1214/aos/1043351250

Goodman, D.R., & Hambleton, R.K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education*, *17*, 145–220. doi: 10.1207/s15324818ame1702_3

Hambleton, R. K., & Slater, S.C. (1996, April). Are NAEP summary reports understandable to policymakers and educators? Presented at the Annual Meeting of National Council for Measurement in Education, New York (NY): National Council of Measurement in Education.

Kosslyn, S. M. (1994). *Elements of graph design*. New York (NY): W.H. Freeman.

Kosslyn, S. M. (1985). Graphics and human information processing: A review of five books. *Journal of the American Statistical Association*, *80*, 499–512. doi: 10.1080/01621459.1985.10478147

Means, B., Chen, E., DeBarger, A., & Padilla C. (2011). *Teachers' ability to use data to inform instruction: Challenges and supports*. Washington, D.C.: U.S. Department of Education.

Playfair, W. (2005). *The commercial and political atlas and statistical breviary*. New York, NY: Cambridge.

Smith, L. D., Best, L. A., Stubbs, D. A., Archibald, A. B., & Roberson-Nay, R. (2002). The role of graphs and tables in hard and soft psychology. *American Psychologist*, *57*, 749–761. doi: 10.1037//0003-066X.57.10.749

Tufte, E. (1983). *The visual display of quantitative information*. Andover, MA: Cheshire Press.

Tukey, J. (1972). Some graphic and semigraphic displays. In T.A. Bancroft, (ed.), *Statistical Papers in Honor of Geroge W. Snedecor* (pp. 293–316). Ames (IA): Iowa State University Press.

Wainer, H. (1997). *Visual revelations: Graphical tales of fate and deception from Napoleon Bonaparte to Ross Perot*. New York (NY): Copernicus.

Wainer, H. (2005). *Graphic discovery: A trout in the milk and other visual adventures*. Princeton, NJ: Princeton University Press.

Wainer, H. (2011). *Picturing the uncertain world: How to understand, communicate, and control uncertainty through graphical display*. Princeton (NJ): Princeton University Press.

Wainer H. (2015). *Truth or truthiness: Distinguishing fact from fiction by learning to think like a data scientist*. New York (NY): Cambridge University Press.

Wainer, H., Hambleton, R. K., & Meara, K. (1999). Alternative displays for communicating NAEP results: A redesign and validity study. *Journal of Educational Measurement*, *36*, 301–335. doi: 10.1111/j.1745-3984.1999.tb00559.x