Unravelling teachers' strategies when interpreting histograms: an eye-tracking study

Lonneke Boels[1,2], Arthur Bakker[1] and Paul Drijvers[1]

[1]Utrecht University, Freudenthal Institute, Utrecht; l.b.m.m.boels@uu.nl;

a.bakker4@uu.nl; p.drijvers@uu.nl

[2]Christelijk Lyceum Delft, The Netherlands

The scarce research on teachers indicate that they often misinterpret histograms. We conjecture that the confusion of histograms with case-value plots is the source of many misinterpretations. Therefore, the question for this study is: what are the most common strategies of secondary school teachers when interpreting histograms and case-value plots? To answer this question, we use a method that allows for a more in-depth analysis of twelve teachers' interpretations of graphical representations than was ever possible before: eye-tracking combined with retrospective verbal reports. Preliminary results show that several teachers apply a case-value plot interpretation strategy on a histogram. Furthermore, some participants use an area interpretation strategy or histogram interpretation strategy applied to a case-value plot. In addition, gaze data suggest that teachers use strategies that did not reach awareness and therefore will not be reported during thinking aloud protocols.

Keywords: Secondary school mathematics, statistics, verbal report, graphical representations, bar graphs.

Background

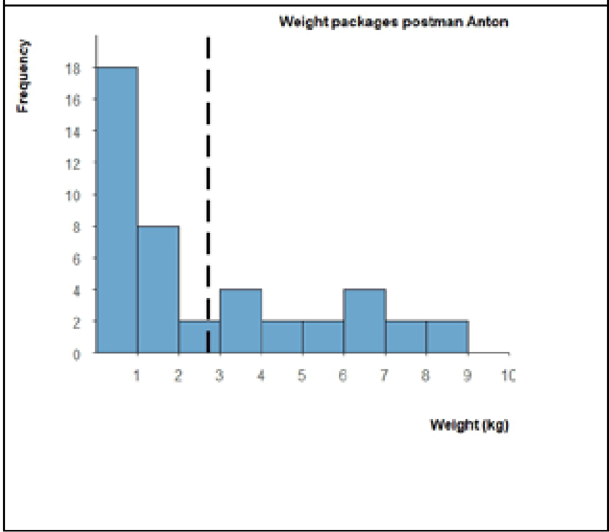
Histograms are very difficult to interpret (e.g., delMas & Liu, 2005; Lem, Onghena, Verschaffel, & Van Dooren, 2013). Misinterpretations of histograms persist despite many interventions aiming to tackle these misinterpretations. For example, Kaplan, Gabrosek, Curtiss and Malone stated in 2014: "the fact remains that the data indicate not only that students entering a statistics course have certain misconceptions about histograms, but also that these misconceptions persist after instruction" (p. 17). In an extensive review of more than 80 publications (Boels, Bakker, Van Dooren & Drijvers, 2019) we identified that most misinterpretations regarding histograms relate to the misunderstanding of two statistical big ideas: *data* and *distribution*. One of the important insights that goes with understanding the big idea of *data* is knowing how many variables are at stake. A misinterpretation existing amongst students, teachers and even researchers is that histograms could display two variables instead of what is correct: only one statistical variable[1] (the one given on the horizontal axis, see Figure 1a; e.g., Boels, Bakker, et al., 2018). The big idea of distribution encompasses centre, shape and variability. One misinterpretation is that a histogram has more variability when it is "bumpier," meaning more variation in the frequency (the heights of the bars) instead of the variation in the data (e.g., Boels, Bakker, et al., 2018; Dabos, 2014; delMas & Liu, 2005; Lem, et al., 2013).

[1] In line with other researchers (e.g., Garfield & Ben-Zvi, 2007), we prefer the term variable over attribute because some people may use attributes for categorical data only.

In line with other researchers (e.g., Cooper & Shore, 2010; Garfield & Ben-Zvi, 2007; Kaplan et al., 2014) we conjecture that the confusion of histograms with case-value plots is the source of many misinterpretations. A histogram and a case-value plot share several salient features (e.g., vertical bars, numbers along the vertical axis). The differences between those two graphical representations—regarding the number of depicted statistical variables as well as the measurement level of the data—are less apparent for most people.

As the number of variables differ in a case-value plot compared to a histogram, so does the interpretation of measures of centre and variability. In Figure 1 the difference for a measure of centre—the mean—is shown for a case-value plot and a histogram. This depicted difference in assessing the mean, also influences the assessment of both the variability and the shape of the distribution. In a histogram, teachers have to look at the horizontal positions of the bars in combination with the bars' heights to assess centre and variability. In contrast, in a case-value plot, teachers have to look at the variation in the heights of the bars only.

a. Mean in histogram (here: 2.7 kg)	b. Mean in case-value plot (here: 2.4 kg)



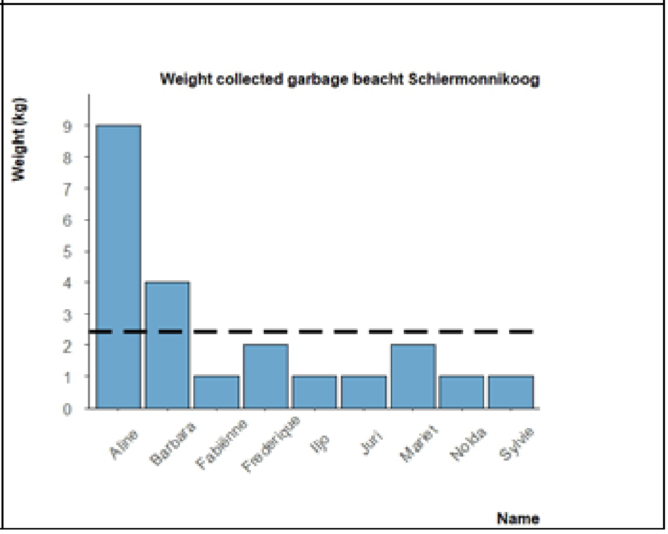


Figure 1: The mean weight (dashed line) is assessed differently in a histogram (left; Item 2) than in a case-value plot (right; Item 8)

The confusion of histograms and case-value plots can explain many of the misinterpretations with histograms reported in the literature but does not answer the question what exactly goes wrong when people interpret histograms. To tackle the persisting problem of misinterpreting histograms, we need to deepen our understanding of the kind of strategies that people use when interpreting histograms. In an exploratory case study we found that students are often unaware that histograms and case-value plots ask for different interpretation strategies (Boels, Ebbes, Bakker, Van Dooren, & Drijvers, 2018). From our personal experience, we conjectured that the same holds true for some of their teachers. The aim of our current study is therefore to identify the most common strategies of teachers when interpreting histograms and case-value plots. Hence, the question for this research is: what are the most common strategies for secondary school STEM teachers when interpreting histograms and case-value plots?

To answer this research question, we used eye-tracking combined with a retrospective interview. The eye-tracking makes it possible to literally see teachers' strategies when solving a task. The advantage of eye-tracking over other techniques, is that it makes an in-depth study possible of the strategies with more detail than was ever possible before. In our exploratory case study, for example, we found that several participants had an initial preferred strategy that they were not aware of and were therefore not reporting during the retrospective interview (Boels, Ebbes, et al., 2018). Eye-tracking has several advantages over other research techniques. For example, using assessments items only, such as CAOS (e.g., delMas, 2005) has already been done extensively and cannot easily answer the question about strategies. Using a thinking aloud protocol (TA) to discover strategies works well if people verbalize their strategy during TA—although TA might slow down the task solving process. But when people need to explain their strategy—why they do what they do, this alters their cognitive processes (Ericsson, 2006) and is therefore unsuitable for our goal. The study of Van Gog and Jarodzka (2013) showed that a retrospective interview improves when a cue is added: the replay of participants' eye movements during the recall. Considering all these arguments we decided to use eye-tracking combined with retrospective reports to unravel teachers' strategies.

From a previous exploratory study with six university students (Boels, Ebbes, et al., 2018) we obtained the following two interpretation strategies that are relevant for the current study: a histogram interpretation strategy and a case-value plot interpretation strategy. A histogram interpretation strategy is associated with a vertical looking pattern and reading of the numbers on the horizontal axis for locating the mean. Furthermore participants using this strategy may use statements as for example "balancing" the graph. A case-value plot interpretation strategy is associated with a horizontal looking pattern and reading of the numbers on the vertical axis. On top of that, these participants may use words like "redistributing" or "make all bars even" (e.g., same height). The exploratory study leads us to the following conjectures for the study reported here:

1) The most common strategy for interpreting histograms is a case-value plot interpretation strategy, followed by the histogram interpretation strategy.
2) The most common strategy for case-value plots is a case-value plot interpretation strategy. Only a few teachers will apply a histogram interpretation strategy onto a case-value plot.
3) Several teachers will have an initial preferred strategy independent of the type of graph at stake.

Method

In total twelve items were either constructed (two) or re-used (ten) from the exploratory study (Boels, Ebbes, et al., 2018). The teachers were asked to either estimate the arithmetic mean of the data in the graph or to compare the arithmetic means of two graphs as estimating the mean can be seen as a necessary prerequisite for assessing the variability. A second reason for choosing the mean was that the target audience of the larger project are secondary school students who are more familiar with measures of centre than measures of variation and the same holds true for their teachers. From the exploratory study we learned that the multiple-choice answers might work as an anchor for the participants (e.g., Tversky & Kahneman, 1974) in items with only one graph. Hence, for the six items with a single graph we used open questions.

We constructed (or re-used from the previous study) a case-value plot for every histogram, with the same salient features such as number of bars, “shape,” range and variable (weight), resulting in a total of twelve questions. Figure 1 gives an example of Item 2 and 8. Items were constructed that differed systematically on the relevant features but that were the same for irrelevant though sometimes salient features as recommended by Orquin and Holmqvist (2017). These features were carefully chosen so that a teacher applying the same interpretation strategy on both the histogram and the case-value plot could be expected to answer the question for the case-value plot correctly and for the histogram incorrectly, and vice versa.

Similar to the previous study we arranged the items in such a way that there were never more than two of the same graph types (histogram or case-value plot) in succession. As we expect that most teachers who confuse case-value plots with histograms will apply a case-value plot strategy onto a histogram, we started with two single left-skewed histograms. This was done to avoid priming (e.g., Lashley, 1951). The first two single histograms were followed by an item in which participants had to compare two case-value plots. Graphs with the same salient features (e.g., Item 2 and Item 8) never directly followed one another. The question for the histogram in Figure 1a was: What is approximately the average weight of the parcels delivered by Anton? The question for the single case-value plot in Figure 1b was: What is approximately the average weight that has been collected per person? In Figure 2, an example of an item with two histograms is given (Item 5). The multiple choice answer options for comparing two histograms were for example: a) Willem delivers on average the heaviest parcels, b) Julia delivers on average the heaviest parcels, c) the average weight of the parcels is approximately the same for both.

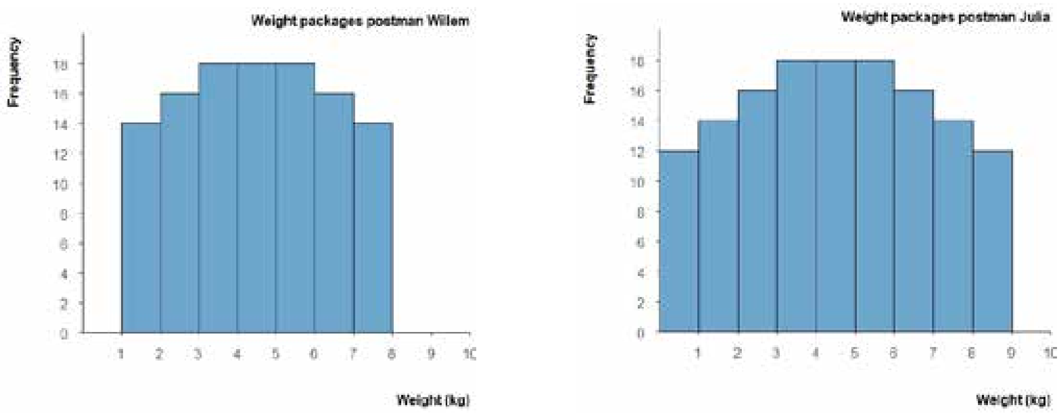


Figure 2: Example of Item 5 in which teachers were asked to compare two histograms and indicate whether the mean weight was higher for one of the two postman or roughly the same

The sample included twelve secondary school teachers of three Dutch secondary schools preparing for college or University grade 7 to 12 (ten teachers were from the same school). Participation was voluntary and science and mathematics teachers were asked to participate. Consent was signed before starting the task. To minimize distractions, the teachers were tested in a one-on-one setting. An explanation was given on the aim of the study as well as how to operate the equipment and software. After completing the task—that included the twelve items mentioned in this research as well as some other items that are not discussed here—participants were asked to report what they thought as well as their strategy during the specific tasks. To improve the quality of these retrospective verbal reports we used their own gazes as a cue (Van Gog & Jarodzka, 2013).

An eye-tracker measures where a person looks on a computer screen and uses infrared light to detect the position of the eyes. The Tobii Pro X2-60 eye-tracker was used, mounted on a laptop with a 13-inch screen by using magnetic mounting brackets. The sampling rate was 60 Hz. The fixations (where people look) and saccades (going from one fixation on the screen to another) were recorded real time by the software Tobii Pro Studio 3.4.5. The first author qualitative coded the gaze data and verbal reports using open, axial and selective coding (Corbin & Strauss, 1990). Furthermore, the coding was checked with gaze data. The gaze data were displayed in heat maps that show where participants looked. The colouring is from green (few fixations) via yellow to red (many fixations), see Figure 3. In Table 1 an example is given of the analysis of gaze and verbal data of one participant. In the next part of the study, not reported here, an analytical analysis of the pattern will be done by using so called areas of interest (AOIs) such as the graph area, vertical axis, horizontal axis and so on and also—if technically possible—by using specialised artificial intelligence software.

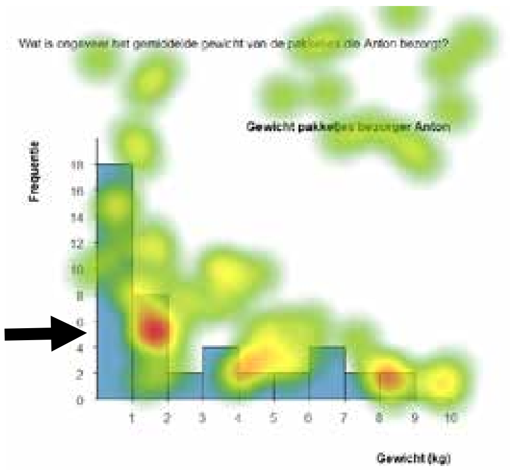
Item number	Gaze data – open codes	Verbal data – open codes	Selective code for strategy
2 (single histogram)	Reads question Reads title graph Looks at top highest bar Looks in middle of graph area (white space) Looks at low bar (7-th) Looks in middle of graph area (white space) Looks at low bar (9-th) …	Look at title Looked at highest and/or lowest number on *y*-axis Look how often lowest bars occur Low bars do not contribute much to mean Higher bars contribute more to mean …	Case-value plot interpretation

Table 1: Example of qualitative analysis (open coding) of gaze and verbal data of a participant as well as the selective code of the combined data

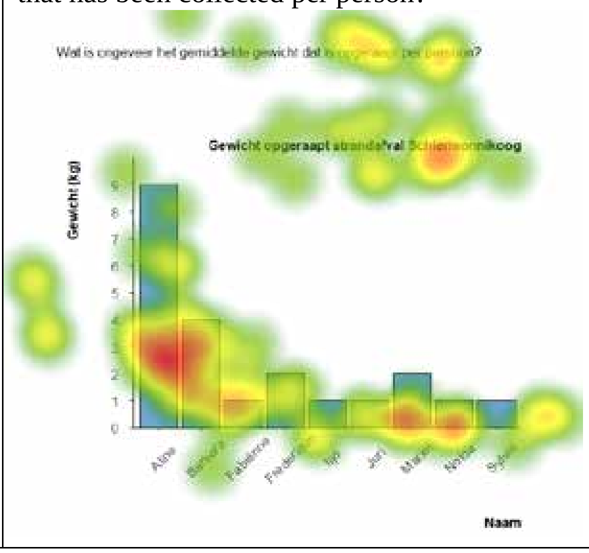
Results

The most common strategies of these teachers were: a histogram interpretation strategy, a case-value plot interpretation strategy and an area interpretation strategy. Several participants seemed to use the same interpretation strategy for both the histograms and case-value plots. An example of a case-value plot interpretation strategy applied to a histogram is shown in Figure 3a. The heat map shows many fixations around frequency 5. The teacher said during the eye-tracking: "Well, uh, about 4 or 5 or something like that." In the retrospective report this teacher said: "I looked at the highest number on the *y*-axis […]. I looked at the lowest frequency. If this occurs less than you have to count that less and […] the highest beams you take more of these, so to speak, from their weight. […] and I have taken the frequency as kilos." This teacher realised at the end of the tasks that s/he systematically used the frequency axis as the weight axis and refers to that in the explanation. The correct answer for this item would be any number between 2 and 3.4 with the actual mean for this specific dataset being 2.7.

a. What is approximately the average weight of the parcels delivered by Anton?



b. What is approximately the average weight that has been collected per person?



This teacher applied a case-value plot strategy on both the histogram (a, Item 2) as well as the case-value plot (b, Item 8). The arrow (a) points at fixations in the histogram at frequency 5 leading to the incorrect answer for the histogram: “about 4-5” kg.

Figure 3: Histogram (a) and case-value plot (b) from Figure 1 with heat map overlay

An example of an area interpretation strategy is where a teacher looked at a histogram using a vertical looking pattern and in the verbal data mentioned that the area of the bars left and right of the mean have to be of the same size. An example of a histogram interpretation strategy applied on a case-value plot is when a teacher looked at Item 10—the case-value plot variant of Item 5 in Figure 2—and said:

Teacher: Only the spread with uh that that right one [the case-value plot on the right with two bars less but the same number of names] is less than that [pause] left one.

Researcher: And why is the spread less?

Teacher: Well, because there are two numbers added thus are, it goes further on.

Researcher: Yes, so you are pointing now [on the screen]: they are wider.

Teacher: Yes, he is wider.

This teacher concluded that the mean was the same and that only the spread differed. In a case-value plot, a higher difference in heights of the bars indicates more spread. The case-value plot with the two students who collected nothing had the lowest mean and most spread. Another finding is that many teachers do not read axis labels and/or graph titles. But even if they do, this does not guarantee that they interpret the histogram or case-value plot correctly. The gaze data indicate that the teacher mentioned above did read the graph title and the labels on the vertical axis—weight (kg)—but still applied the wrong strategy.

Conclusion and discussion

The first conclusion is that the most common strategy for interpreting means in histograms is a case-value plot interpretation strategy. Using a case-value plot interpretation strategy for a histogram implies not understanding that a histogram is for one statistical variable only and that histograms therefore differ from a case-value plot with two depicted statistical variables. As stated in the background section, understanding the big idea of data implies understanding how many variables are depicted in a histogram. This conclusion is therefore in line with the finding in our review study on histograms (Boels, Bakker et al., 2019). Furthermore it is in line with misinterpretations found by others (e.g., Cooper, 2008; Lem et al. 2013). In addition, we speculate that—although several teachers did not read labels on the axes—teaching teachers to read labels will not suffice to eliminate the confusion of histograms with case-value plots, as—for example—the teacher mentioned earlier who applied a histogram interpretations strategy onto a case-value plot, did read the labels.

The second conclusion is that teachers sometimes used an area interpretation strategy for finding the mean in a histogram. In this interpretation strategy the data in the histogram are correctly interpreted (e.g., the statistical variable is on the horizontal axis) but the misinterpretation is related to the big idea of *centre*. Instead of the mean, the median is found with this strategy of equal areas. This finding is in line with findings from others (e.g., Cooper, 2008). We speculate that this interpretation strategy is due to excessive exposure to symmetric—specifically normal—distributions in histograms in schoolbooks and statistics courses where mean and median are indeed the same.

In addition, the gaze data can be used to identify strategies that participants may not be aware of and therefore will not be reported during thinking aloud protocols or retrospective interviews. Eye-tracking therefore adds a new research tool to researchers' toolkit, making it possible to analyse participants' strategies in more detail than was ever possible before. Furthermore, researchers and teacher educators can use these results to better design curriculum materials. Finally, in line with the wish list from CERME10 in 2017 (Bakker, Hahn, Kazak, & Pratt, 2018), this research concentrate on teachers as well as contributes to teachers' Statistical Knowledge for Teaching (Groth, 2007).

This research is funded with a Doctoral Grant for Teachers from The Netherlands Organisation for Scientific Research, number 023.007.023 awarded to Lonneke Boels.

References

Bakker, A., Hahn, C., Kazak, S., & Pratt, D. (2018). Research on probability and statistics education: Trends and directions. In T. Dreyfus, M. Artigue, D. Potari, S. Prediger, & K. Ruthven (Eds.), *Developing research in mathematics education - twenty years of communication, cooperation and collaboration in Europe* (pp. 46–59). Abingdon, UK: Routledge

Boels, L., Bakker, A., Van Dooren, W., Drijvers, P. (2019). *Conceptual difficulties when interpreting histograms: A review.* Manuscript submitted for publication.

Boels, L., Ebbes, R., Bakker, A., Van Dooren, W., & Drijvers, P. (2018). Revealing conceptual difficulties when interpreting histograms: an eye-tracking study. In M. A. Sorto, A. White, & L. Guyot (Eds.), Looking back, looking forward. *Proceedings of the Tenth International Conference on Teaching Statistics* (pp. 1–4), Kyoto, Japan. Voorburg, The Netherlands: International Statistical Institute.

Cooper, L. L., & Shore, F. S. (2008). Students' misconceptions in interpreting center and variability of data represented via histograms and stem-and-leaf plots. *Journal of Statistics Education, 16*(2), 1–13

Cooper, L., & Shore, F. (2010). The effects of data and graph type on concepts and visualizations of variability. *Journal of Statistics Education, 18*(2), 1–16.

Corbin, J., & Strauss, A. (1990). Grounded theory research: Procedures, canons and evaluative criteria. *Zeitschrift Für Soziologie, 19*(6), 418–427. doi.org/10.1007/BF00988593

Dabos, M., (2014). A glimpse of two year college instructors' understanding of variation in histograms. In K. Makar, B. de Sousa, & R. Gould (Eds.), Sustainability in statistics education. *Proceedings of the Ninth International Conference on Teaching Statistics* (pp. 1–4), Flagstaff, Arizona, USA. Voorburg, The Netherlands: International Statistical Institute.

delMas, R., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal, 4*(1), 55–82.

Ericsson, K. A. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. *The Cambridge handbook of expertise and expert performance*, 223-241.

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review, 75*, 372–396. doi.org/10.1111/j.1751-5823.2007.00029.x

Groth, R. E. (2007). Toward a conceptualization of statistical knowledge for teaching. *Journal for research in Mathematics Education*, 427–437. http://www.jstor.org/stable/30034960

Kaplan, J.J., Gabrosek, J.G., Curtiss P. & Malone, C. (2014). Investigating student understanding of histograms. *Journal of Statistics Education, 22(2),* 1-30.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–131). New York: Wiley.

Lem, S., Onghena, P., Verschaffel, L., & Van Dooren, W. (2013). On the misinterpretation of histograms and box plots. *Educational Psychology, 33*(2), 155–174.

Orquin, J. L., & Holmqvist, K. (2017). Threats to the validity of eye-movement research in psychology. *Behavior Research Methods*. doi.10.3758/s13428-017-0998-z

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. doi.org/10.1017/CBO9780511809477.002

Van Gog, T., & Jarodzka, H. (2013). Eye tracking as a tool to study and enhance cognitive and metacognitive processes in computer-based learning environments. In: R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 143–156). New York, NY: Springer. doi.org/10.1007/978-1-4419-5546-3_10