PME 43

CONFERENCE OF THE INTERNATIONAL GROUP FOR
THE PSYCHOLOGY OF MATHEMATICS EDUCATION

07-12 July 2019 **Pretoria South Africa**

Proceedings

43rd Annual Meeting of the International Group for the Psychology of Mathematics Education

VOLUME 2

Research Reports (A-K)

Editors:
Mellony Graven, Hamsa Venkat, Anthony A Essien and Pamela Vale

Proceedings of the 43rd Conference of the International Group for the Psychology of Mathematics Education

Pretoria, South Africa

7 – 12 July 2019

Editors:

Mellony Graven

Hamsa Venkat

Anthony A Essien

Pamela Vale

Volume 2

Research Reports (A-K)

EYE-TRACKING SECONDARY SCHOOL STUDENTS' STRATEGIES WHEN INTERPRETING STATISTICAL GRAPHS

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

[1]Utrecht University, [2]Christelijk Lyceum Delft

Students persistently misinterpret histograms. Based on a literature review we conjectured that the confusion of histograms with case-value plots was leading to inappropriate interpretation strategies. Hence, the question for this research is: what are the most common strategies for secondary school students when estimating the mean in histograms and case-value plots? In a task with twelve graphs (histograms or case-value plots) we measured students' eye movements (N=10, grade 10–11. The most common strategies students use are: a case-value plot interpretation strategy and a computational strategy both applied to histograms. Furthermore, some students reported strategies that are not in line with their gaze data nor their estimated mean.

BACKGROUND

For more than two decades misinterpretations with histograms have been reported in the literature (e.g., Friel & Bright, 1995; Whitaker & Jacobbe, 2017) as well as attempts to improve students' understandings of histograms through interventions (e.g., Meletiou-Mavrotheris & Lee, 2005). An extensive review of the literature on histograms showed numerous misinterpretations that are widespread amongst students from every school level as well as countries (Boels, Bakker, Van Dooren, & Drijvers, 2019). In addition, many people are not clear on what a histogram is. In the context of this paper we use the following criteria for a histogram (Boels, et al., 2019): (1) it is a graph with bars; (2) the data of only one statistical variable are presented; (3) these data are measured at ratio measurement level; (4) the vertical axis displays frequency.

The persistence of misinterpreting histograms was our rationale for searching for underlying difficulties. In a literature review we found two difficulties: the lack of understanding the big statistical ideas data and distribution (Boels et al., 2019). The big idea data refers amongst others to the measurement level of the data. Distribution refers amongst others to how the data are distributed (e.g., variability or spread) and how this is depicted in a graphical representation (shape). The literature suggests that a histogram is confused with a case-value plot (e.g., Cooper & Shore, 2010; Garfield & Ben-Zvi, 2007). A case-value plot is a kind of look-a-like of the histogram as it shares the same salient features with a histogram. These salient features are: two axes, numbers along the vertical axis and—in our construction—the same number of bars and the variable weight in kilograms. Nevertheless, on the level of the big ideas data and distribution a histogram is very distinct from a case-value plot. For example, a case-value plot depicts two variables (here: name and weight, see Figure 1) whereas a histogram represents only one (here: weight). Furthermore, in a case-value plot each bar is one measurement while in a histogram each bar usually stands for several

2019. In M. Graven, H. Venkat, A. Essien & P. Vale (Eds.). *Proceedings of the 43rd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 113-120). Pretoria, South Africa: PME.

measurements. As a result, in a case-value plot the heights of the bars are the measured value whereas in a histogram the positions of the bars on the horizontal axis are the measured values. In addition, the variability of the data in a histogram is the horizontal spread weighted by the frequency whereas the variability of the data in a case-value plot is given by the vertical difference in the heights of the bars.

a. Mean in histogram (here: 3.3 kg) b. Mean in case-value plot (here: 3.7 kg)

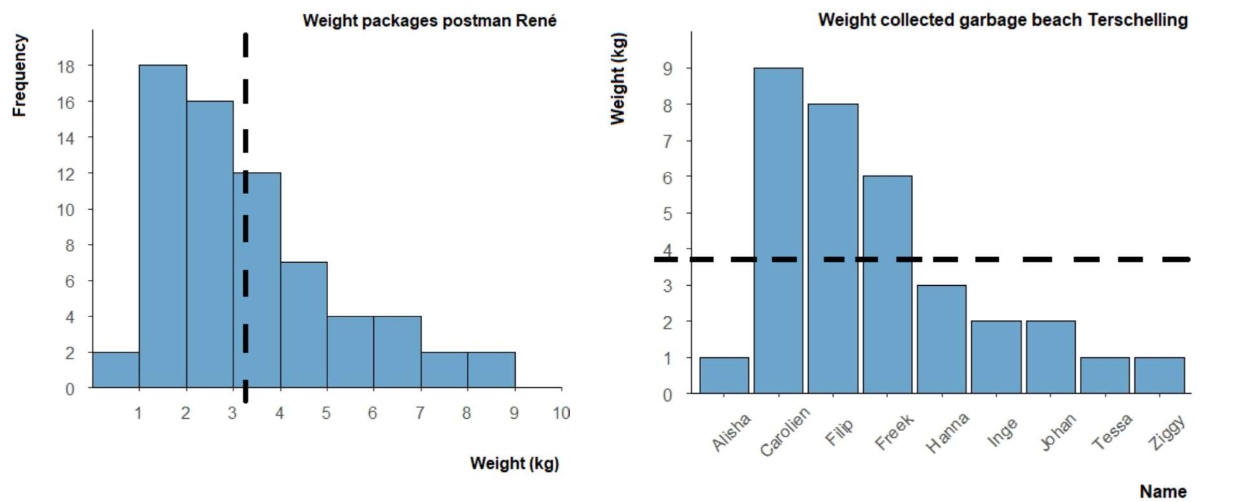

Figure 1: Measures of centre are depicted differently in a histogram (a, Item 1) from a case-value plot (b, Item 4).

Although the confusion of histograms with case-value plots can explain a substantial part of the misinterpretations, this does not shed any light on how this confusion arises. In a first exploratory study, we therefore studied the interpretations strategies that students used to answer questions about histograms and case-value plots (Boels, Ebbes, Bakker, van Dooren, & Drijvers, 2018). In our first study we were searching for persistent misinterpretations after learning about histograms. We therefore included university students only. In preparing that study, we were concerned that we would not find any misinterpretations at all as we thought that we had made the differences between the case-value plots and the histograms too obvious, see Figure 1. For example, we had clearly placed labels next to the axes showing that the requested information (average weight) could be found on either the horizontal axis (histogram) or the vertical axis (case-value plot). We nevertheless found several students applying a case-value plot interpretation strategy onto a histogram (e.g., a horizontal looking pattern and reading of the numbers on the vertical axis thus the frequency instead of the measured values).

In an ideal world, students would enter the university without misinterpretations, as interpreting histograms is part of the Dutch secondary school curriculum. In order to track down where these misinterpretations come from, we need to study secondary school students' interpretations of histograms. These students have learned about histograms in grade 9 or 10 (usually at age 14–16). Hence, the aim of our current study is to identify the most common strategies of secondary school students when interpreting histograms and case-value plots.

In searching for a suitable task for interpreting histograms, we noted that many researchers ask for the variability depicted in a histogram (e.g., Garfield & Ben-Zvi, 2007). Nevertheless, we decided to ask for estimating the mean from a histogram. The first reason for asking for the mean is that finding the mean in a histogram can be regarded as a prerequisite for using measures of variability. As measures of centre are depicted differently in histograms from case-value plots, the same holds true for measures of variability. Moreover, variability can be considered as variation around the mean. The second reason for using the mean instead of measures of variability is that secondary school students are very familiar with the arithmetic mean (as this is, for example, used for grading almost all summative assessment tests in secondary schools in the Netherlands) and not so much with measures of variability. Hence, the question for this research is: what are the most common strategies for secondary school students when estimating the mean in histograms and case-value plots?

METHOD

To answer our research question we constructed twelve items with histograms or case-value plots. Six of these items were either single histograms (three) or case-value plots (three). The question for these items was to estimate the mean weight of the packages of a postman (histograms) or garbage collected on a beach (case-value plots). These questions were open ended. Three of the other six items concerned two histograms each and the other three two case-value plots each. These six items held the question in which graph the estimated mean was bigger. The three answer options were: the graph on the left, the graph on the right or both graphs approximately the same. All histograms had a look-a-like case-value plot that shared the same salient features with the histograms such as number of bars and range of the weight scale, see for example Figure 1 (Item 1 and Item 4). Two items with the same salient features never followed one another and there were no more than two items with the same graph type in succession. The graph construction was in line with the recommendations to use stimuli that differ systematically on relevant features but are similar for irrelevant features (Orquin & Holmqvist, 2017). To avoid priming (Lashley, 1951) we started with a single histogram in the first two items as we expected that students misinterpreting histograms would apply a case-value plot interpretation strategy onto a histogram (see data analysis for an explanation on the strategies) and not so much the other way around. The twelve items were followed by another thirteen items that were part of an exploratory study not reported here.

Two questionnaires were used to gather data on the students' background (e.g., their grades for mathematics) as well as their precognition on—for example—the meaning of frequency and arithmetic mean. By asking first to estimate the mean of five assessment scores (each between one and ten) and then to calculate this mean, the difference between an estimation and a calculation was subtly stressed as we did not want students to make precise calculations.

Figure 2: Setup of the equipment (the person in the photo was not a participant)

Eye-tracking in combination with retrospective thinking aloud with students' eye movements as a cue was used. We could have collected only the answers for these twelve items (e.g., Whitaker & Jacobbe, 2017). Although this method is fruitful for discovering how widespread the misinterpretations are as well as for getting indications of the type of misinterpretations students have, it does not tell us in detail which strategies students use to answer these items. In addition, we considered other methods, such as combining the assessment items with a thinking aloud protocols or with retrospective interviews, both without the use of eye-tracking. We rejected all these options for several reasons. Firstly, when people need to explain their strategy during thinking aloud, this not only slows down their work on the task, but also changes their cognitive processes (Ericsson, 2006). Secondly, retrospective interviews are often unreliable but can be improved when participants' eye movements are used as a cue (van Gog & Jarodzka, 2013). As the eye-mind hypothesis does not always hold (Anderson, Bothell, & Douglass, 2004; Schindler & Lilienthal, 2019) a retrospective interview is needed to link the eye movements data to the students' strategy.

Ten secondary school students of one pre-university level school (grades 10–11, mean age: 16 years; range 15–18 year) participated in the study reported here which is part of a larger study with secondary school students and their STEM teachers. Students participated voluntarily and consent was signed by the students and in most cases also by their parents (required for students aged 15 or less). All participants received a small gift (mathematical puzzle) for their participation. The first author recruited the students, carried out the research and conducted the interviews at the participants' school. To stay within the time frame of one lesson of this school the time per student was limited to 45 minutes. In this time frame all work with the participant was done, which included filling in two questionnaires, answering the tasks described above and being interviewed with participants' own gaze data as a cue. The eye-tracking part took approximately 10–15 minutes per student, leaving about 20–25 minutes for the interview.

A Tobii Pro X2-60 eye-tracker with a 60Hz sampling rate was used, mounted with adhesive mounting brackets on a laptop with a 13-inch display. The Tobii Pro Studio 3.4.5 software recorded in real time where people were looking on the screen using harmless infrared light to detect the gaze. We also used a chin rest to gain a better quality of the eye-tracking data, see Figure 2. For recording the retrospective interviews a Rode NT-USB studio microphone was used. Students were tested in a one-on-one setting in a room in their own school. Before starting the task, a calibration procedure was used with nine points on the screen. The calibration was followed by a validation procedure with four points on the screen. Between all items a fixation cross was used for validation as well as making sure that all eye movements start at the right-hand side of the screen which was a blank area in the next screen with the item.

a. Case-value plot strategy applied onto a histogram (Item 1)	b. Case-value plot strategy applied onto a case-value plot (Item 4)

The colours indicate where students' gaze was less (green), medium (yellow) and most (red). This student stated that the mean weight was approximately ten (a, histogram) and five (b, case-value plot). The arrow points at fixations around frequency ten in the histogram. Note that there are no fixations on the horizontal weight axis in the histogram although the label weight on the horizontal axis was read by this student.

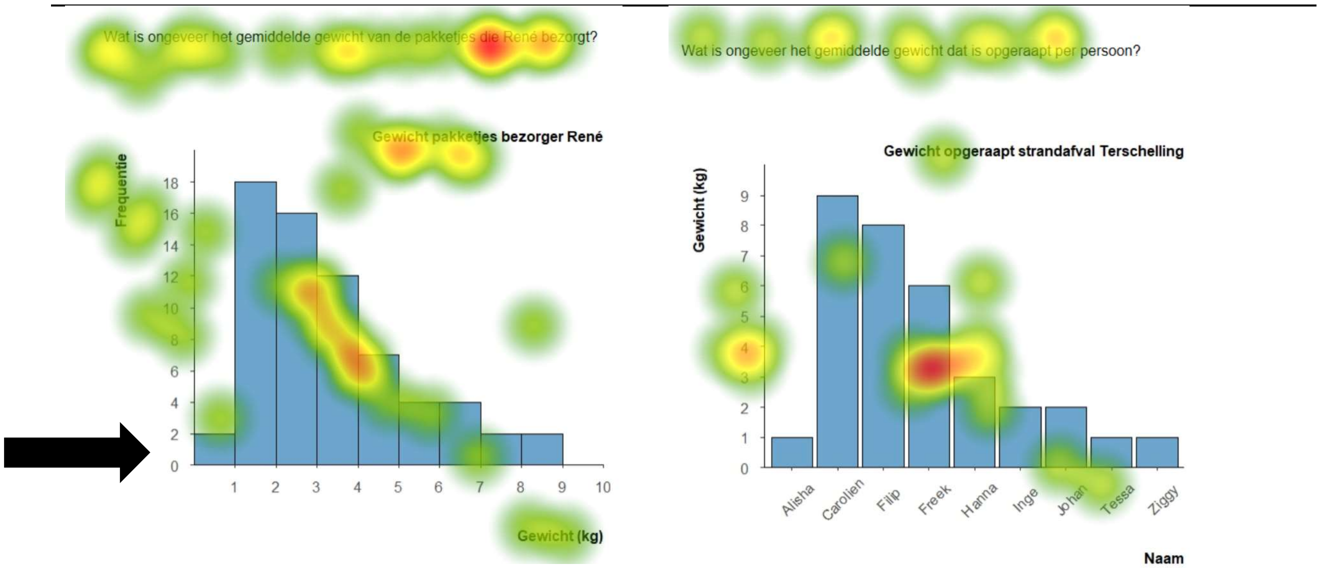

Figure 3: Heatmap of a student applying a case-value plot interpretation strategy onto a histogram (a) and a case-value plot (b).

DATA ANALYSIS

We analysed the answers given by the students for each item. Furthermore, from our previous studies we conjectured that students would use two interpretation strategies: a case-value plot interpretation strategy and a histogram interpretation strategy (Boels, Bakker, & Drijvers, 2019; Boels et al., 2018). We define a case-value plot interpretation strategy as a strategy in which students display a horizontal looking pattern and read the requested mean at the vertical axis, see Figure 3. During the interview, these students report that they take the middle of the bars or make the bars "the same" [height]. Sometimes students move their hands in a horizontal line when they explain what they did. If that happened, the interviewer reported this aloud in the interview. We defined a histogram interpretation strategy is as a strategy in which students display a vertical looking pattern and read the requested mean at the horizontal axis. During the interview, these students—for example—reported that they balance the graph. Sometimes they might also point to a balancing point at the horizontal axis. The interviews and gaze data were qualitatively coded with open, axial and selective coding (Corbin & Strauss, 1990). From these data we inferred the strategies that the students used. We also used heat maps, showing students' gaze data in gradients from green via yellow to red (see Figure 3) and gaze plots showing the order of students' gaze data.

RESULTS

For the single histogram items students correctly answered on average 0.5 item out of 3 items which was much less than the average number of correct answers for single case-value plots, see Table 1.

3 items; single histograms	3 items; double histograms	3 items; single case-value plots	3 items; double case-value plots
0.5 [0–2]	0.7 [0–3]	1.5 [1–2]	2.3 [2–3]

Table 1 Average number and [range] of correct answers per student.

a. Computational strategy applied onto a histogram (Item 1)	b. Computation strategy applied onto a case-value plot (Item 4)
This student stated that the mean weight was approximately six (a, histogram) and three plus one third (b, case-value plot). Note the many fixations on both axis and the going back and forth of the fixations from axis to the graph area indicating the computational strategy.	

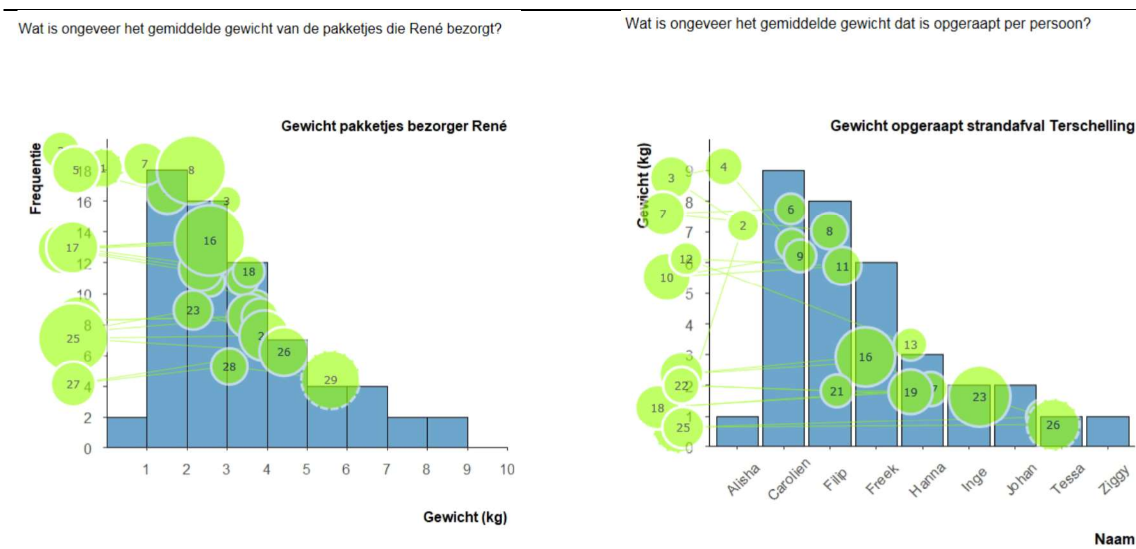

Figure 4: Plot of a student's gaze. This student uses a computational strategy.

In the qualitative analysis of our data we found the two interpretation strategies described in the data analysis as well as a new strategy that we call a computational strategy. In this strategy a student computes a total (either a total of the frequency or a total of the frequency times measured value) and divides this by either the number of the bars or sometimes the total frequency. In the gaze data the looking pattern appears as jumping from bar to bar, for example along the horizontal axis (counting the number of bars, see Figure 4).

Furthermore, some participants report strategies and answers that did not match the answer given earlier during the online eye-tracking session. For example, the participant whose strategy is shown in Figure 3 reported a lower answer than ten in his retrospective interview and said to have looked on the horizontal axis to find this answer. This is not in line with the answer that was given during the online eye-tracking session (ten) nor with the gaze data as there are no gazes on the horizontal axis. The validation data suggest that gaze data are accurate so this is not due to any off-set nor drift, indicating that this participant did not look at the horizontal axis at all. Although the axis could have been seen in a peripheral view, this is unlikely for reading off numbers.

CONCLUSIONS AND DISCUSSION

The first conclusion is that these ten students used two strategies most frequently: a case-value plot interpretation strategy applied onto a histogram and a computational strategy. The case-value plot interpretation strategy is in line with many findings in the literature (e.g., Cooper & Shore, 2010). The computational strategy was found in our literature review but is rarely reported (Ismail & Chan, 2015).

The second conclusion is that several students reported strategies as well as answers in their retrospective interview that did not match the answers given during the online measurement and nor their gaze data. This indicates that a retrospective interview (a thinking aloud protocol) might not always be reliable for secondary school students. Gaze data are measured online (meaning: during the performance of the task) and are therefore a useful complement in finding students' strategies.

The third conclusion is that even students who read titles and axis labels misinterpret histograms. We therefore speculate that it will not suffice to learn students to (better) read the labels of a graph. We conjecture that underlying students inappropriate use of strategies is that they do not have understood the big ideas of data and distribution and specific how these effect measures of centre and shape in different types of graphs.

Acknowledgement

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

Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye movements do not reflect retrieval processes: Limits of the eye-mind hypothesis. *Psychological Science, 15*(4), 225-231.

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

Boels, L., Bakker, A., & Drijvers, P. (2019). Unravelling teachers' strategies when interpreting histograms: an eye-tracking study. *CERME 11,* Utrecht. Accepted manuscript.

Boels, L., Ebbes, R., Bakker, A., van Dooren, W., & Drijvers, P. (2018). Revealing conceptual difficulties when interpreting histograms: An eye-tracking study. Paper presented at the *Tenth International Conference on Teaching Statistics,* Kyoto, Japan.

Cooper, L. L., & Shore, F. S. (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.

Ericsson, K. A. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance,* (pp. 223-241). Cambridge, UK: Cambridge University Press.

Friel, S. N., & Bright, G. W. (1995). Graph knowledge: Understanding how students interpret data using graphs. Paper presented at the *Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education,* Columbus, OH.

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*(3), 372-396.

Ismail, Z., & Chan, S. W. (2015). Malaysian students' misconceptions about measures of central tendency: An error analysis. *Proceedings of the 2nd ISM International Statistics Conference 2014 (ISM-II), 1643,* 93-100.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112-136). New York: Wiley.

Meletiou-Mavrotheris, M., & Lee, C. (2005). Exploring introductory statistics students' understanding of variation in histograms. *Proceedings of the 4th Congress of the European Society for Research in Mathematics Education*. Sant Feliu De Guíxols, Spain.

Orquin, J. L., & Holmqvist, K. (2017). Threats to the validity of eye-movement research in psychology. *Behavior Research Methods*, *50*(4), 1645-1656.

Schindler, M., & Lilienthal, A. J. (in press). Domain-specific interpretation of eye-tracking data: Towards a refined use of the eye-mind hypothesis for the field of geometry. *Educational Studies in Mathematics.*

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: Springer.

Whitaker, D., & Jacobbe, T. (2017). Students' understanding of bar graphs and histograms: Results from the LOCUS assessments. *Journal of Statistics Education, 25*(2), 90-102.