SECONDARY SCHOOL STUDENTS INTERPRETING AND COMPARING DOTPLOTS: AN EYE-TRACKING STUDY

Lonneke Boels[1,2], Wim Van Dooren[3]

[1]Utrecht University of Applied Sciences, [2]Utrecht University, the Netherlands

[3]KU Leuven, Belgium

Dotplots can increase students' reasoning about variability and distribution in statistics education but literature shows mixed results. To better understand students' strategies when interpreting non-stacked dotplots, we examine how and how well upper secondary school students estimate and compare means of dotplots. We used two item types: single dotplots requiring estimation of the mean and double ones requiring comparison of means. Gaze data of students solving six items were triangulated with data from stimulated recall. Most students correctly estimated means from single dotplots; results for comparison were mixed. A possible implication is that single, non-stacked dotplots can be seen as a step towards teaching students to interpret univariate graphs but further research is needed for comparing graphs.

THEORETICAL AND EMPIRICAL BACKGROUND

The ability to interpret graphs is an important educational goal. For instance, graphs can reveal patterns in data that may not be noticed when looking purely at computational measures (such as means or correlations). In this paper, we will focus on graphs that are used to represent the distribution of a single variable. The distribution of a variable is one of the key concepts of statistics, and a prerequisite for understanding more complex distributions. Research has started to investigate what role various graphical representations (histograms, boxplots, and dotplots) have on the reasoning about the distribution of a variable (e.g., Lem et al., 2013a). More specifically, it has revealed a range of strategies and possible misinterpretations of each graphical representation.

In recent years, such strategies and misinterpretations are being investigated by means of eye-tracking data, that can yield a unique insight in strategies students use when interpreting the graphs and drawing conclusions. A recent review (Boels et al., 2019a) revealed a range of difficulties when interpreting histograms, and eye-tracking data have shown that students tend to interpret them as if these were case-value plots (Boels et al., 2022). Also for boxplots, various misinterpretations have been documented (Lem et al., 2013b), and currently, attempts are made to reveal these by eye-tracking data. The current paper focusses on strategies used on the third graph type, i.e., dotplots.

According to a local instruction theory on developing students' statistical literacy (Bakker, 2004), dotplots can increase students' understanding of variability in data (delMas & Liu, 2005), support students' reasoning about distribution (Bakker & Gravemeijer, 2004; Garfield & Ben-Zvi, 2008) and scaffold students' interpretation of

2023. In M. Ayalon, B. Koichu, R. Leikin, L. Rubel & M. Tabach (Eds.). *Proceedings of the 46th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 131-138). PME 46.

histograms (Lyford & Boels, 2022). However, the literature on students' dotplot interpretations showed mixed results. For example, Lem et al. (2013a) demonstrated that first-year university students tended to employ a local view on the distribution of a variable when interpreting dotplots, thereby focusing on individual observations rather than the distribution as a whole, more often than with the other representations. Moreover, they found that students had more difficulties comparing means, medians, and variation of data presented in dotplots when distributions were asymmetric compared to symmetric distributions. In addition, students used heights of dotplots to compare skewness of distributions, similar to what they applied to histograms. Lyford (2017) showed that in several cases students interpreted dotplots better than histograms. For example, although various students used stack heights in stacked dotplots, they did so less often than in histograms. However, when students compared 'bumpy' and 'spaced uniform' graphs, students answered correctly significantly more often for histograms than for dotplots. Therefore, also for dotplots, we want to achieve a better understanding of students' strategies when interpreting them, thereby relying in part on eye-tracking data. The current study addresses the research question: *how and how well do upper secondary school students estimate and compare arithmetic means of dotplots?*

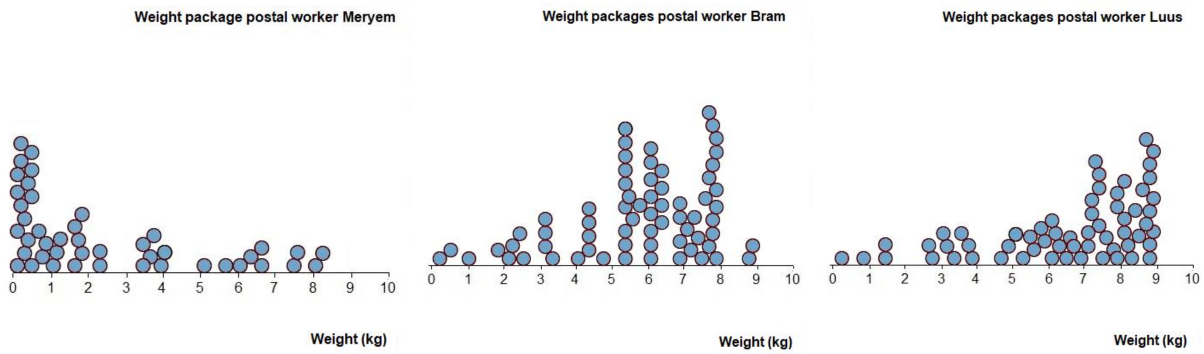

Figure 1: Item13, 15 and 18 of the original data collection for which students were asked to estimate the arithmetic mean from each dotplot. For item 13, for example, the actual mean is 2.7 (Table 1) and the range for correct answers was [1.6 – 3.8].

METHOD

We present answers and gaze data of five Grades 10–11 secondary school students. The students followed a pre-university track. They solved a total of six dotplot items. We designed two item types: open ended questions requiring estimation of the mean (Figure 1) and multiple choice items requiring comparison of means (e.g. Item 17, Figure 2). Note that our students had never seen a dotplot before in their education, but are familiar with case-value plots (where the height of each bar is the measured value) and histograms (where the position of each bar indicates the range of measured values). For each item type we designed three items. Gaze data were triangulated with verbal data from stimulated recall (cued retrospective thinking aloud) for which students' own eye movements were used as a cue (Van Gog et al., 2005). Data triangulation is needed because there is no straightforward relation between students' solution strategies and gaze patterns (Schindler & Lilienthal, 2019). The data presented in this article stem from a larger data collection with 50 upper secondary students solving 25 items with various statistical graphs (e.g., histograms, case-value plots). In line with recommendations of Orquin and Holmqvist (2017), stimuli differed systematically on relevant features (e.g., positions of dots) but were kept similar for irrelevant features (e.g., color of dots, weight scales).

A Tobii Pro X2-60 eye-tracker with a 60 Hz sampling rate was used, mounted on a HP ProBook 6360b laptop with a 13-inch display (refresh rate: 59 Hz). The Tobii Pro Studio 3.4.5 software (n.d.) recorded in real time where people were looking on the screen using harmless infrared light to detect the gaze. A chin rest was used for better gaze data quality. Mean accuracy was acceptable (1.16°) with highest accuracy on the for this research most relevant graph area (0.27°; considered good); average precision (0.58°; RMS-S2S; Holmqvist et al., 2022) is considered good (see Boels et al., 2022 for more details).

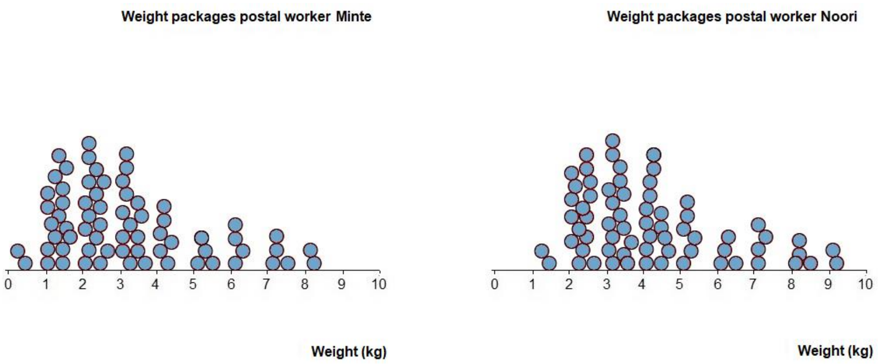

Figure 2: Example of a double dotplot item. Students were asked to compare arithmetic means, with three answer options: higher mean on the left, higher mean on the right, or approximately the same means. Here, the higher mean is on the right.

MAIN RESULTS

Regarding how well students interpret dotplots: four of the five students correctly estimated the mean from all single dotplot items (Table 1). One student (L03) overestimated the mean for the first dotplot item (Figure 1). However, for comparing means, results were more mixed, and only one of the students consistently gave a correct answer (Table 2).

For length reasons, the elaboration on how students interpreted the dotplots is restricted to the *single* dotplot items. Interpretations of double dotplot items will be presented during the PME 46 conference. We found four different strategies for single dotplots. The most common strategy is a strategy that we previously called a *histogram (interpretation) strategy* (Boels et al., 2019b): Students estimate the mean by finding the 'balance' point of the graph, or a 'clump' of dots. When students apply this strategy, a vertical scanpath pattern is visible in their gaze data.

Table 1: Characteristics of students and students' estimations of means from *single* dotplots items. Answer ranges were set to actual means [m= …] +/- 1.1. Correct answers in bold. Item numbers refer to their placement in the original item sequence.

Student	Age	Grade	Sex	Answers		
				Item13 [m=2.7]	Item15 [m=5.7]	Item18 [m=6.4]
L01	16	11	M	**3**	**5**	**7**
L02	18	11	M	**2**	**5**	**6.5**
L03	16	10	F	*4*	**6**	**7**
L04	17	11	F	**2**	**6**	**6.5**
L05	15	10	F	$\mathbf{2\frac{1}{3}}$	**6**	**5.5**

Table 2: Students' answers for comparing means from double dotplots items. Correct answers in bold.

Student	Answers		
	Item14	Item16	Item17
L01	Frans	**Same**	**Noori**
L02	**Sam**	**Same**	**Noori**
L03	Frans	Mustafa	**Noori**
L04	**Sam**	Mustafa	**Noori**
L05	**Sam**	Ilse	Same

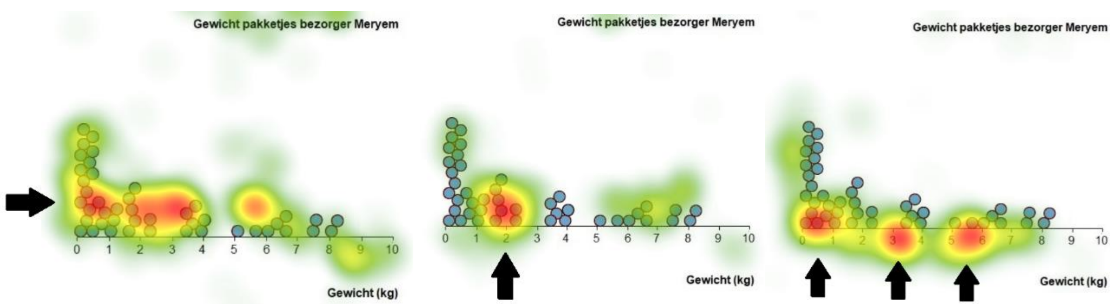

Figure 3. Heatmaps of Item13. Left: case-value plot strategy (L01). Middle: histogram strategy (L02). Right: computational strategy (L05). The colours indicate where students' gaze was less (green), medium (yellow) and most (red).

In a *computational strategy*, students add the measured values (positions of dots). An indication for this strategy is long fixations on each stack or number along the axis. Surprisingly, as shown in Figure 3, student L01 seemed to have used a strategy that incorrectly used the *heights* of the dotplots, instead of their horizontal positions. In such strategy, the dots are equally spread out along the horizontal axis. The height of the resulting stack is then estimated. We previously (Boels et al., 2019b) called this a *case-value plot (interpretation) strategy*. The difference between this strategy (Figure 3, left) and a histogram strategy (Figure 3, middle) is clearly visible in the heatmaps by the difference in horizontal spread-outness of gazes. The verbal data of student L01 do not substantiate this claim, but we think that is most likely due to this student switching to a correct strategy for later items and only reporting the latter for all items.

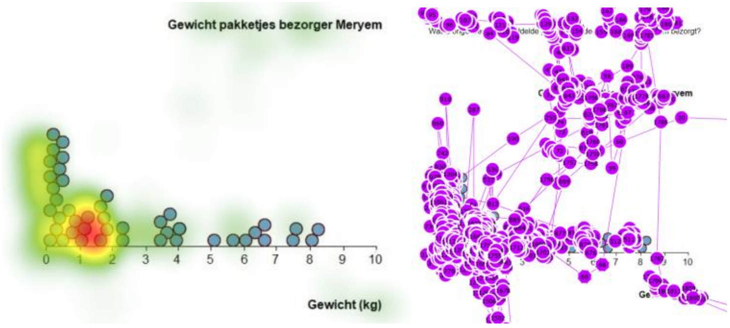

Figure 4. Gaze pattern of student L04 for Item13: heatmap and gazeplot .

Table 3: Students' strategies for single dotplot items. Correct strategies in bold.

Student	Strategy		
	Item13	Item15	Item18
L01	Case-value plot strategy	**Histogram strategy**	**Histogram strategy**
L02	**Histogram strategy**	**Histogram strategy**	**Histogram strategy**
L03	Unclear strategy	**Histogram strategy**	**Histogram strategy**
L04	**Histogram strategy**	**Histogram strategy**	**Histogram strategy**
L05	**Computational strategy**	**Histogram strategy**	**Histogram strategy**

Eye-tracking data showed that initially students did not know quite how to approach the first dotplot item. This is also visible, for example, in Table 3 where for Item13 four different strategies were found for these five students, compared to one strategy for Item15 and Item17. In addition, from the *video* of the eye movements we inferred that some students switched strategies. For example, the video of L04 for Item13 showed at the start long fixations around the numbers 0, 1 and 2 and the corresponding stacks of dots, and a longer fixation on the top half of the highest stack. Such long fixations might indicate thinking, which is necessary for a computational strategy. However, for a full computational strategy we would expect long fixations around all numbers and stacks along the horizontal scale (Figure 3, right). Instead, there are much fewer and much shorter fixations on the dots at the higher numbers. These shorter fixations seem to indicate that the shape and location are looked at and that ultimately no computations were performed. As these shorter fixations occurred toward the end of the trial, shortly before the answer 2 was given, it appeared that this student switched strategy. The verbal data confirm the computational start and strategy switch:

L04: And then I saw that a lot of them had a weight of between zero and one and because of that I could work out that [this] was a pretty low mean. And then I did an approximate estimate.

Researcher1: Yes, okay. And I had the idea that you were also going to count here [started with counting] is that possible?

L04: No[t] with this one […] With this one I first thought I'll count. So I had already started counting but then I thought that's too much counting work and then I just started making an estimate because then I saw that, I guess so much was [in the left part] relative to the right.

The computational strategy that student L04 used at the start cannot be clearly inferred from the heatmap and gazeplot (Figure 4), although the heatmap shows that this student focused on the stacks with lower numbers. However, the video of the gazes *does* show a gaze pattern—at the start—that belonged to a computational strategy.

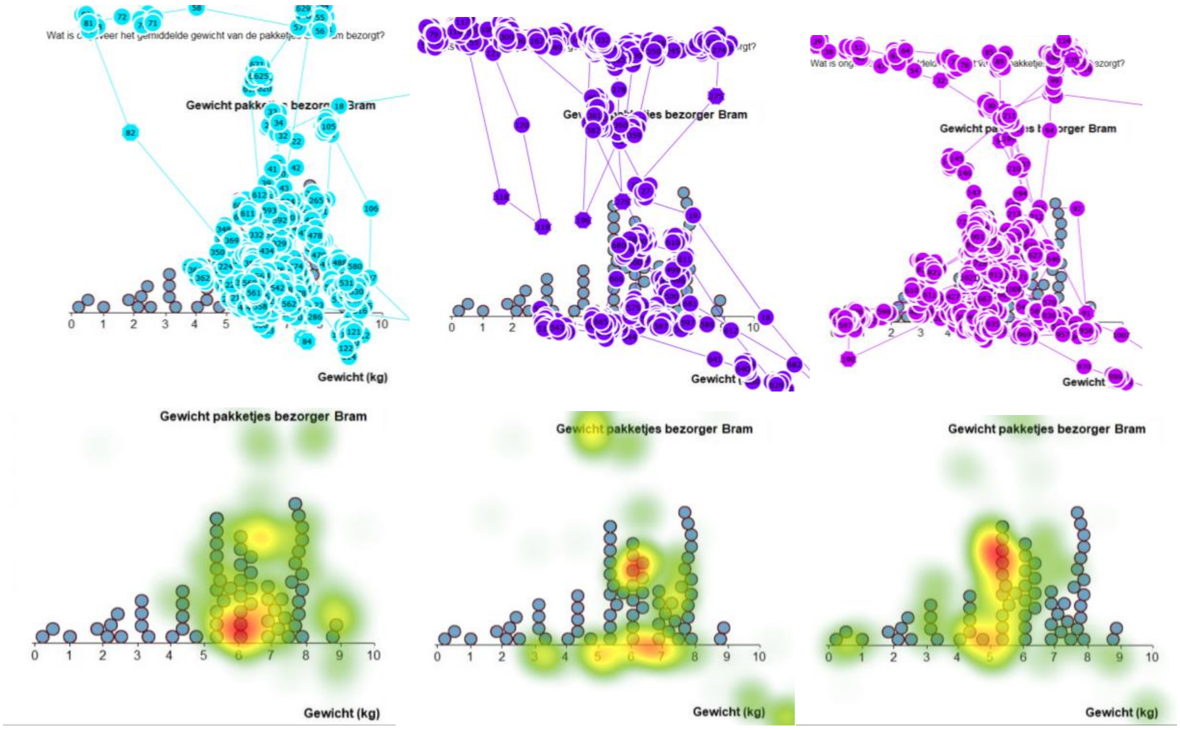

Figure 5. Gazeplots (top) and heatmaps (bottom) of correct strategies for estimating the mean, applied to Item15 by student L02 (left), L03 (middle), and L04 (right).

Both from the videos of the gaze data and Table 3 it became clear that students settled their strategy for single dotplot items (Figure 5) after Item13.

CONCLUSIONS AND DISCUSSION

From our study it appears that students are quite capable of estimating means from single dotplots, although they never learned about dotplots in school. For comparing means of dotplots, students answers suggest mixed results. Of course, we need to consider that this paper has the limitation that we involved a small number of students, and that graphs were presented in a fixed order due to technical restrictions.

Contributing to the local theory of interpreting statistical graphs, our study suggests that single non-stacked dotplots are well understood by upper secondary school students who never encountered these graphs in their curriculum. A possible implication is that single non-stacked dotplots can be seen as a step towards teaching students to interpret univariate graphs (e.g., histograms, boxplots, stem-and-leaf plots). However, for comparing distributions, students' variation in answers are in line with the mixed results Lyford (2017) found for undergraduate students. Therefore, further research is needed to investigate when and how students correctly *compare* dotplots.

This study is the first to reveal by means of eye-tracking the kind of strategies that students employ when interpreting dotplots. This is the major methodological advantage of eye-tracking data in this context: It reveals more details about students' thinking processes compared to concurrent thinking aloud (Van Gog et al., 2005). Concurrent thinking aloud may affect the actual thinking process and thereby not provide valid measures. In that sense, eye-tracking may even be seen as a—for research purposes—less obtrusive investigation method. The eye-tracking data may also reveal the entire range of strategies employed by students (both correct and incorrect strategies), and even show that students switch from one strategy to another while solving a specific problem. Such data are not only useful for research; they may also be relevant for educational practice. For instance, teachers can use a selected number of gaze patterns to draw students' attention to correct and incorrect interpretations of dotplots.

ACKNOWLEDGEMENTS

The authors thank Arthur Bakker and Paul Drijvers for their contributions to the design of items, and the data collection. The data collection for this research was funded with a Doctoral Grant for Teachers from The Netherlands Organization for Scientific Research (NWO), number 023.007.023, awarded to Lonneke Boels.

REFERENCES

Bakker, A. & Gravemeijer, K. P. E. (2004). Learning to reason about distribution. In D. Ben-Zvi, & J. Garfield (Eds.). *The challenge of developing statistical literacy, reasoning and thinking* (pp. 147–168). Springer.

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools.* Doctoral dissertation. Freudenthal Institute, Utrecht University.

Boels, L., Bakker, A., & Drijvers, P. (2019b). Eye tracking secondary school students' strategies when interpreting statistical graphs. In M. Graven, H. Venkat, A.A. Essien, &

P. Vale (Eds.). *Proceedings of the 43rd Conference of the International Group for the Psychology of Mathematics Education, 2,* (pp. 113–120). Pretoria, South-Africa: PME.

Boels**,** L., Bakker, A., Van Dooren, W., & Drijvers, P. (2019a). Conceptual difficulties when interpreting histograms: A review. *Educational Research Review, 28,* Article 100291.

Boels, L., Bakker, A., Van Dooren, W., & Drijvers, P. (2022). [submitted] *Secondary school students' strategies when interpreting histograms and case-value plots: An eye-tracking study.* Freudenthal Institute, Utrecht University.

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

Garfield, J. B., Ben-Zvi, D., Chance, B., Medina, E., Roseth, C., & Zieffler, A. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer.

Holmqvist, K., Örbom, S. L., Hooge, I. T. C., Niehorster, D. C., Alexander, R. G., Andersson, R., Benjamins, J. S., Blignaut, P., Brouwer, A-M., Chuang, L. L., Dalrymple, K. A., Drieghe, D., Dunn, M. J., Ettinger, U., Fiedler, S., Foulsham, T., Van der Geest, J. N., Witzner Hansen, D., Hutton, S., … Hessels, R. S. (2022). Eye tracking: empirical foundations for a minimal reporting guideline. *Behavior Research Methods*.

Lem, S., Onghena, P., Verschaffel, L., Van Dooren, W. (2013a). External representations for data distributions: in search of cognitive fit. *Statistics Education Research Journal, 12*(1), 4–9.

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

Lyford, A. J. (2017). Investigating Undergraduate Student Understanding of Graphical Displays of Quantitative Data Through Machine Learning Algorithms (Doctoral thesis).

Lyford, A., & Boels, L. (2022). Using machine learning to understand students' gaze patterns on graphing tasks. Invited Paper - Refereed. In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022).* Rosario, Argentina. ISI/IASE.

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. (2019). 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 101*, 123–139.

Tobii (n.d.). Tobii Studio. Users' Manual. Version 3.4.5.

Van Gog, T., Paas, F., Van Merriënboer, J. J., & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied, 11*(4), 237–244.