STUDENTS' INITIAL COGNITIVE PROCESSES WHILE COMPARING TWO DATA SETS: AN APPROACH TO FOSTER CONCEPTUAL KNOWLEDGE ABOUT BOXPLOTS

Martin Abt, Katharina Loibl, Timo Leuders, and Frank Reinhold
University of Education Freiburg, Germany
martin.abt@ph-freiburg.de

Boxplots can be used to compare distributions. However, many students do not acquire conceptual knowledge relevant for such comparisons because instructional approaches often focus on procedural knowledge. In a digital learning environment, seventeen eighth-graders (unfamiliar with boxplots) faced the problem of estimating two long jumpers' performance based on dot plots. We investigated students' naïve preconceptions. Findings indicate that meaningful comparisons were made if students evaluated both distributions using the same criteria. Often, performance was not evaluated with just one criterion, but several were contrasted in the responses. Two intuitive ways to quantify data variability could be identified. Although many students determined which part of the distribution lies in a defined interval, only a few students intuitively considered the spread of a fixed proportion of the distribution.

INTRODUCTION

The boxplot is a representation used in descriptive statistics to display distributions. It is characterized by a compact presentation of a considerable amount of information: the boxplot provides a visualization of the median, the extremes, and—with the range and the interquartile range (IQR)—two elementary measures of variability. To summarize these characteristic values, statements about the position of rather dense areas of a distribution can be made. Although on the one hand, this compact form of representation is particularly suitable for a quick and overview-like comparison of different distributions, on the other hand, the large number of characteristic values presented makes the boxplot a complex learning object (Bakker et al., 2005). When boxplots are introduced procedurally, which is a typical instructional approach in school mathematics (Ben-Zvi & Garfield, 2004), students first create an ordered list of the data. They then obtain the 'five-point summary' by counting data values to determine the quartiles and then 'draw' the boxplot with the help of these five characteristic values. However, research has shown that the conceptual knowledge required for comparison and interpretation of data sets based on boxplots does not necessarily transfer from this procedural approach (Edwards et al., 2017). We argue that a conclusive reason for this is that (a) the conceptual idea of position parameters and range can be well understood from the procedure, but (b) the procedure for finding, for example, the IQR, inadequately characterizes the concept of this measure of variability. Although determination of the IQR via successive halving (or by counting data values to determine quartiles) is not cognitively challenging, understanding the concept of this parameter as a measure of dispersion seems to be more demanding. Indeed, to prevent the IQR from being understood only as the range of the middle half of the data rather than as a measure of dispersion, and thus from being severely truncated in its meaning (Bakker et al., 2005), we believe that students should be enabled to answer questions such as the following:

- What is the benefit of considering a second measure of dispersion in addition to the range?
- Why do we consider only a part of the distribution for the calculation of such a second measure of dispersion and why one half and not one third?
- Why does it make sense to consider the middle and not, e.g., the upper half of the data?

If students are not able to answer such questions, we suspect this might indicate insufficiently acquired conceptual knowledge resulting from a procedural approach, which may be one reason for the occurrence of typical misconceptions. In fact, studies have been able to show that the width (respectively the 'area') of the box was not even associated with IQR in the shortened sense mentioned above. Rather, as in other already known representations, students tend to interpret this area according to the principle: 'The larger the area, the larger the sample proportion represented' (Lem et al., 2013). Following this line of reasoning about potential shortcomings when introducing the boxplot with a mere procedural approach, we present an approach for teaching the boxplot with the aim of fostering a sustainable conceptual understanding of the representation. Here, we are most interested in the informal

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.* International Association for Statistical Education. iase-web.org

concepts that students develop along the way to developing conceptual knowledge about the conventional measures of center and variation (Bakker, 2004b). For this purpose, we developed a digital learning environment that confronts students with a problem situation for which they cannot rely on an already-taught mathematical concept. Instead, a problem-solving process is initiated in which the students themselves develop, e.g., a measure of data variability and, in perspective, elaborate it in a future extended learning environment into a mathematical concept with a formalized definition. This design of the learning environment follows the principles of *genetic learning* (Wagenschein, 1968) or *guided reinvention* (Freudenthal, 1973). These principles emphasize the value of not merely applying a (ready-made) mathematical concept to a problem but of using the problem as an opportunity to develop this concept precisely on one's own. To deal with mathematical ideas already in the process of creation and to work on them oneself may allow a deeper and more differentiated understanding of mathematical concepts and their contexts of origin (Leuders et al., 2011). The principle of guided reinvention has also been investigated in earlier work with the introduction of statistical representations such as the boxplot and has been mentioned as pivotal to the formation of stable concepts (Bakker, 2004a). The digital learning environment records students' activities while learning (i.e., processing data) and allows us to investigate students' learning prerequisites and steps of preliminary understanding. This analysis is fundamental for development and optimization of a fully-fledged learning environment.

Research Question: Which different naive preconceptions of measure of central tendency and of variation do eight-graders utilize when comparing two data sets in a genetic learning environment?

METHOD

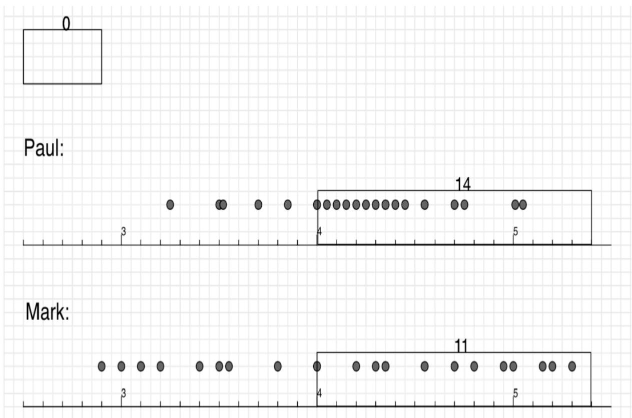
The purpose of the learning environment is to enable a genetic, authentic approach for the comparison of two data sets via a suitable contextualization, in which looking at typical boxplot elements serves as a solution to a problem and motivates the learning process from the very beginning. For this purpose, the students were confronted with the problem of comparing two long jumpers, Paul and Mark, based on distances from their twenty most recent training jumps. Students are asked to predict which of the two athletes has the best chance of winning an upcoming competition. To do this, students slipped into the role of a reporter for the school newspaper at Paul's school. We deliberately abstained from the step of first ordering an unsorted list of the jump measures and provided two dot plots right from the start. The data sets are presented in the Results section. The medians, extremes, and variability of long jump distances differed for two athletes, Paul and Mark. Mark's jumps had a higher median (4.25 m vs. 4.225 m) and a higher maximum (5.3 m vs. 5.05 m) distance. Paul jumps had a higher arithmetic mean (4.21 m vs. 4.15 m) due to a higher minimum (3.25 m vs. 2.9 m) and overall a less variable jumping distance. Both the range (1.8 m vs. 2.4 m) and the IQR (0.58 m vs. 1.43 m) are substantially lower for Paul's jumps than for Mark's jumps. Furthermore, Paul shows a prominent accumulation of jumps between 4.0 m and 4.5 m, whereas Mark's jumps are distributed over the entire range without such accumulation areas. Students were asked to mark areas in both dot plots that they would like to use to describe the different performances of the athletes. For this, they could drag any number of rectangles from the upper left corner onto the plots and then adjust the width so that any groups of points could be highlighted. The number of points contained in the rectangle was displayed. Unlike previous studies (e.g., Bakker, 2004b) we avoided stacking in the dot plots. Such structures with vertical extension have a strong signaling effect, which in our questions could promote argumentation in the sense of a mode instead of variation and spread. Afterwards, students were asked to describe the athletes' performances (at least three sentences) and make a reasoned prediction for the competition results. The study involved 17 eighth graders from a secondary school in Germany. The students had not acquired knowledge about the 'boxplot' at school before the study. The study was approved by local authorities, and students and their parents consented to anonymous use of the data for research purposes. We conducted the survey in two successive mathematics lessons during regular class time. The students worked individually and without time restrictions. The learning environment was presented on iPads using the Moodle learning platform with interactive content developed in CindyJS (Richter-Gebert & Kortenkamp, 2012).

For the data analysis, the drawn rectangles and students' responses entered in the answer field were recorded anonymously for each student according to the time of submission. Except for related timestamps, no further data were collected. We used an inductive qualitative approach to find and describe categories of students' preconceptions—based on the combination of students' drawn rectangles and their written texts (for an impression of such rectangles drawn by students see Figures 1–4). Concretely, we based the category building on the following guiding questions:

- Which areas of the datasets were highlighted with rectangles?
- Is it possible to identify a consistent pattern in the markings that would enable comparison of the two data sets based on the drawn rectangles?
- What (descriptive) characteristics are used to argue who has the better chance of winning the title? Is there a weighing visible between the best or worst jumps on the one hand and the variability of the performance on the other?
- Does this argumentation refer to the previously drawn rectangles in a meaningful way? Are (descriptive) characteristics of the data sets derived from the markings?

RESULTS

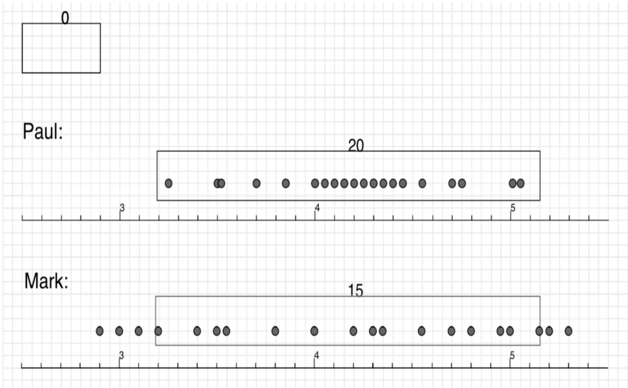
We investigated students' naive preconceptions of measures of central tendency and of variation utilized when comparing two data sets in our contextualized digital learning environment. We found categories, which mainly map whether, and if so, in which ways, students quantified the variability of the data. We were able to identify a first, largest group (*n* = 7) where students quantified variability by comparing which of the two athletes had performed more jumps in a previously defined interval. Rectangles of equal width were positioned over the defined intervals in both distributions. By comparing the points contained in each rectangle, differences in the concentration of jumps in these areas could be made visible. To predict who will win, students weighed the lower variability in Paul's performance against Mark's better highest performance. In the example of work from this group (Figure 1), we can see how in both distributions the jumps with distances between 4 m and 5.4 m were highlighted. To predict who would win, the student referred to these highlighted jumps over 4 m.



Students' prediction of the contest winner:
"It will be exciting at the state championships because Paul has the routine at over 4 meters, but Mark did the longest jump."

Figure 1. Example of a student's work from group 1

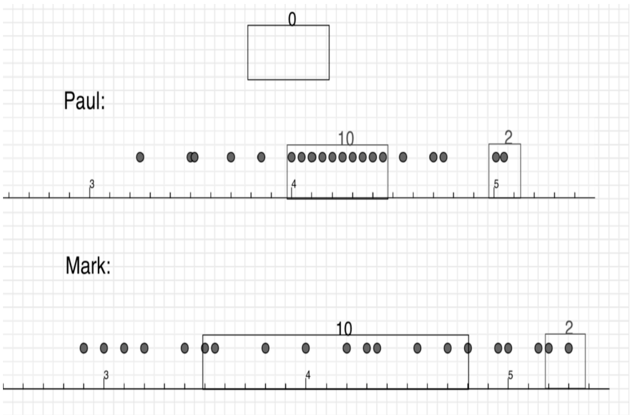
A second group (*n* = 3) showed the same strategy in their markings, comparing the number of jumps in a defined interval, but argued unilaterally either with a comparison of the best and respectively worst jumps or with a comparison of the variability of the jumps. In the example of work from this group (Figure 2), the student marks Paul's range in both data sets, but then argues only with the best and worst jumps of the two athletes.



Students' prediction of the contest winner:
"I think Paul has a better chance than Mark because Paul almost doesn't get under the three meters at all."

Figure 2. Example of a student's work from group 2

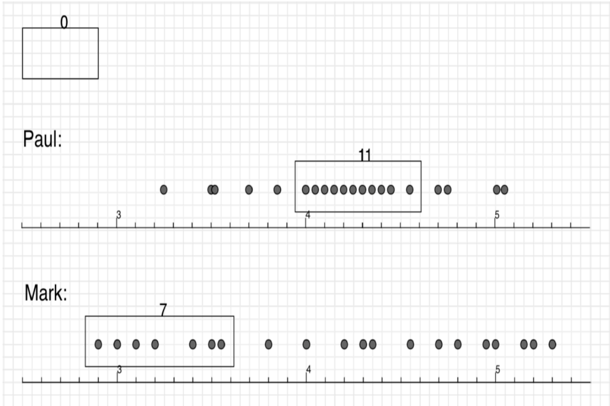
As in the previous two groups, students in the third group ($n = 2$) used a rather standardized approach in their marking and set rectangles according to consistency in both data sets. Also, as in group 1, there is a weighing of higher best and lower variability. The crucial difference is the way in which the variability of the performance pattern is quantified and made visible. Unlike in the previous two groups, it is not the number of jumps in a certain predefined interval that is compared, but conversely how broadly a certain proportion of the distribution is spread. In the example of work from this group (Figure 3), the middle ten jumps are marked beside the best two jumps, and then the differences in the spread of the middle half of the data are used as an argument.



Students' prediction of the contest winner:
"Even though Mark's best jumps were longer, I think Paul has a better chance to win because Paul's jumps are closer together when you take a closer look. Most of them are close to each other in the 4 m–4.5 m range. Mark's jumps, though, are often farther apart and very irregular. Most of his jumps are in the 3.5 m–4.8 m range. Both are spaced 10 jumps apart, so overall Paul jumps more accurately, which is why I think he has better chances."

Figure 3. Example of a student's work from group 3

The fourth and last group ($n = 5$) must be clearly distinguished from the previous three. In answers of this group, we find no link between the markings and the text and, except for one answer, no standardized procedure for marking. In fact, in this group, only two responses argued meaningfully even without reference to the markings. The student, whose work to exemplify this group is displayed in Figure 4, marks the salient clustering in Paul's distribution and the more frequent lower jump ranges in Mark's data set. Even though this student did not argue meaningfully, the rectangles drawn by the student show that characteristic features of the distributions were registered and that the markings were intended to emphasize areas of high concentration and particularly low attribute values.



Students' prediction of the contest winner: "I think Paul has a better chance to win first place because I notice that Mark has always jumped about the same and Paul has done different jumps."

Figure 4. Example of a student's work from group 4

DISCUSSION

This study addressed the question of what naïve student preconceptions an instructional approach to the learning object of boxplots might build upon. In all students' solution processes to the given genetic problem, the markings indicated that characteristic, descriptive properties of the distributions were perceived. Both the markings and the argumentations mainly focused on variability and the extrema of the distributions. In this regard, it can be stated that the contextualization of the problem was able to motivate students to apply possibly existing intuitive and intermediary knowledge. It seems noteworthy that variability was not only frequently included in students' initial decision making but was actually perceived as crucial in the majority of cases. If the marking was not standardized (group 4), no link between marking and text could be found, and, as could be expected, in most of these cases, no meaningful argumentation took place. Only a few students succeeded in arguing meaningfully despite the non-standardized markings. The study thus provides indications that a meaningful comparison of data sets is particularly successful if standardized markings were created beforehand and reference was made to these markings in the argumentation. All students in our sample who fulfilled these two conditions also succeeded in reasoning meaningfully, although their reasoning varied qualitatively in terms of whether a process of weighing became visible in the justifications (groups 1 & 3) or not (group 2). We assume that an instructional approach for teaching 'boxplot' could more easily build on the preconceptions observed in group 3. This is because preconceptions of dispersion were already apparent in group responses and, conversely, the idea—seen in groups 1 and 2—that rectangles of the same width can contain different proportions of the entire distribution can be rather obstructive in the sense of typical misconceptions (e.g., Lem et al., 2013), as mentioned previously. After all, the key property of the boxplot is the invariability of the represented proportion, regardless of the width of the boxplot portions. This study gives first indications that more conceptual knowledge-oriented teaching on the learning object of 'boxplot' can build on existing, but diverse, preconceptions. However, the small sample size does not allow a generalization of the systematics presented yet. Further research would need to investigate whether the categories can be confirmed or need to be further supplemented. A central question to which this study cannot provide an answer is to what extent the identified preconceptions can be further developed. This question needs to be addressed because this study suggests that certain intermediate knowledge is better suited for this purpose than others. Furthermore, it should be examined how dependent the intermediate knowledge is on the chosen contextualization. For example, it might be interesting to investigate, if a context that focuses less on the extremes and more on dispersion itself as a target variable maybe requires less salient 'geometric' structures in the dot plot.

REFERENCES

Bakker, A. (2004a). *Design research in statistics education: On symbolizing and computer tools* [Doctoral dissertation, Utrecht University]. CD-Bèta Press. https://dspace.library.uu.nl/bitstream/handle/1874/893/full.pdf?sequence=2

Bakker, A. (2004b). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, *3*(2), 64–83. https://doi.org/10.52041/serj.v3i2.552

Bakker, A., Biehler, R., & Konold, C. (2005). Should young students learn about box plots? In G. Burrill & M. Camden (Eds.), *Curricular development in statistics education. IASE Roundtable Conference* (pp. 163–173). ISI/IASE. https://doi.org/10.52041/SRAP.04302

Ben-Zvi, D., & Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 3–15). Springer. https://doi.org/10.1007/1-4020-2278-6_1

Edwards, T. G., Özgün-Koca, A., & Barr, J. (2017). Interpretations of boxplots: Helping middle school students to think outside the box. *Journal of Statistics Education*, *25*(1), 21–28. https://doi.org/10.1080/10691898.2017.1288556

Freudenthal, H. (1973). *Mathematik als pädagogische Aufgabe* [Mathematics as a pedagogical task] (*Bd. 1)*. Klett.

Lem, S., Onghena, P., Verschaffel, L., & Van Dooren, W. (2013). The heuristic interpretation of box plots. *Learning and Instruction*, *26*, 22–35. https://doi.org/10.1016/j.learninstruc.2013.01.001

Leuders, T., Hußmann, S., Barzel, B., & Prediger, S. (2011). „Das macht Sinn!" Sinnstiftung mit Kontexten und Kernideen [“That makes sense!” Creating of meaning with contexts and core ideas]. *Praxis der Mathematik in der Schule*, *37*, 2–9.

Richter-Gebert, J., & Kortenkamp, U. H. (2012). *The Cinderella.2 manual*. Springer. https://doi.org/10.1007/978-3-540-34926-6

Wagenschein, M. (1968). *Verstehen lehren: Genetisch, sokratisch, exemplarisch* [Teaching understanding: Genetic, Socratic, exemplary]. Beltz.