CLUMPS OR CHUNKS? - CONTEXTUAL RELEVANCE OF STUDENTS' FEATURES OF THE DATA

Christian Büscher

TU Dortmund University, Germany

For reasoning on data, learners make use of features of the data in an intuitive and informal way. More insights however are needed into learners' processes of reasoning on data to identify conditions and reasons for learners to focus on particular features of the data. This study reports on results of a design research project on German 7th grade students' reasoning on data. The analysis shows how students' focus on features of the data follows their perceived contextual relevance induced by the context of a teaching-learning arrangement.

STUDENTS' FOCUS ON FEATURES OF THE DATA

The development of adequate reasoning on data is one of the main goals of statistics education and a central interest in statistics education research (Biehler, Frischemeier, Reading, & Shaughnessy, 2018). For doing so, learners need to be able to adopt an 'aggregate view' on data, perceiving data distributions not as unstructured collections of individual cases, but as holistic entities on their own with their own emergent properties like centre and spread (Konold, Higgins, Russell, & Khalil, 2015). Developing such an aggregate view on data however seems to be challenging for learners (Bakker & Gravemeijer, 2004).

In order to find approaches to develop learners' reasoning on data, statistics education research has identified a number of 'intuitive' or 'informal' *features of the data* that students seem to focus on. In a study by Konold et al. (2002), 7th and 9th grade students use *modal clumps* to summarize data of daily roadkill: small central ranges surrounding the mode of the data to represent the amount of 'typical' roadkill. Makar and Confrey (2003) find that preservice teachers use modal clumps to summarize data of student achievement and that they partition these data into *chunks* to represent groups of low, middle, or high achievement. In a newer study, Schnell and Büscher (2015) find that 8th grade students who compare daily temperature data also partition the data into chunks, but do so in a creative way according to their individual concepts instead of following a simple 'low-middle-high' partition.

Statistics education research commonly calls for building on learners' informal use of features of the data to develop their reasoning on data (e.g. Konold et al., 2002). However, although research has shown that learners commonly focus on features of the data such as clumps or chunks, little is known about whether such focus is 'natural' for students, or if it is influenced by other factors. More insight is needed into the

2018. In NNN (Eds.). *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education*, Vol. 1, pp. XX-YY. Umeå, Sweden: PME.

conditions and reasons for learners to focus on specific features of the data. This paper presents a contribution to close this gap of research.

THE SITUATIVE NATURE OF STUDENTS' ACTIONS

To provide a framework that can contribute to explaining students' use of specific features of the data, this study draws on the epistemological Theory of Conceptual Fields (Vergnaud, 1996). According to this theory, learners' actions follow possibly unconscious organizational invariants called *concepts-in-action* and *theorems-in-action*. Concepts-in-action are "categories (objects, properties, relationships, transformations, processes, etc.) that enable the subject to cut the real world into distinct elements and aspects […] according to the situation and scheme involved" (ibid., p. 225), whereas theorems-in-action are "proposition[s] that [are] held to be true by the individual subject for a certain range of situation variables" (ibid., p. 225). Learning consists of expanding and connecting one's concepts- and theorems-in-action into increasingly complex *conceptual fields*.

Identifying relevant features of the data can provide an example action by learners in which they draw on concepts- and theorems-in-action. When investigating data, the learners from the study of Konold et al. (2002) use modal clumps to represent typical roadkill. Using the Theory of Conceptual Fields, this can be interpreted as drawing on the concept-in-action (indicated by ||...||) of ||*modal clumps*|| to cut the data into relevant and irrelevant parts. Their theorem-in-action (indicated by <...>) <*modal clumps represent the typical roadkill*> then describes their use of the feature of ||*modal clumps*||. Thus, students' concepts- and theorems-in-action influence the features of the data held relevant by them.

Central to both, concepts- and theorems-in-action, is their situative nature. Concepts-in-action depend on the "situation and scheme involved", and theorems-in-action hold true for "a certain range of situation variables" (Vergnaud, 1996, p. 225). Thus, the specific situation at hand plays an important role in determining which features of the data are focused on by learners. Regarding the learning of statistics, such a situation can be introduced through the context of a statistics teaching-learning arrangement.

RESEARCH QUESTION

In order to develop learners' reasoning on data, instruction should make use of their intuitive use of features of the data, such as their use of modal clumps or chunks of data. However, more insights into the conditions and reasons for learners' focus on features of the data are needed in order to specifically support students' learning processes. The learning-theoretical background suggests a strong influence of the context of a teaching-learning arrangement for students' use of features of the data. Thus, this study concerns the following research question: *how does the context of a teaching-learning arrangement influence students' focus on features of the data?*

RESEARCH DESIGN

Design Research provides a research methodology suitable to evaluate the effects of the context of a teaching-learning arrangement. This section outlines the methodological considerations of this study.

Topic-specific Didactical Design Research as framework

This study is part of a larger research project in the framework of Topic-specific Didactical Design Research (Prediger & Zwetzschler, 2013; for the whole project see Büscher, 2018). Research conducted in this framework consists of iterative cycles of four interrelated working areas: (1) specifying and structuring learning goals and content; (2) developing the design; (3) conducting and analysing design experiments; and (4) developing local theories on teaching and learning processes. The methodological heart of Design Research consists of conducting design experiments (Gravemeijer & Cobb, 2006). Influenced by the design of a teaching-learning arrangement and controlled by a design experiment leader, design experiments do not simply aim at observing, but at actively initiating learning processes in order to investigate the effects of the design and to understand students' reasoning.

Participants, data collection, and data analysis

The larger research project consisted of five cycles of design experiments from 2014 to 2015, in total 34 participants in 32 design experiments. This study reports on results of the fifth cycle of design experiments in December 2015, consisting of 3x3 focus design experiments with 3x2 participants. Participants were assigned into pairs, and each pair took part in a design experiment series consisting of three consecutive design experiments few days apart. In each design experiment, one pair of 7th grade students from a German secondary school worked on a teaching-learning arrangement. The students volunteered and were sampled from one class by the mathematics teacher based on high communicative ability, but not based on high mathematics achievement.

The focus design experiments were videotaped and fully transcribed, resulting in 405 minutes of transcribed video data. The data corpus was reduced by partitioning the data into episodes corresponding to the phases of the design experiment manual (see below) and by choosing episodes that allowed insights into the students' use of features of the data. For these episodes, the students' concepts- and theorems-in-action were reconstructed in an interpretative step of basic analysis adopting Vergnaud's (1996) constructs and then following open coding and data-led category development (cf. Corbin & Strauss, 1990).

Design of a teaching-learning arrangement

Each design experiment of the series, introduces one problem to the students. Due to space restrictions, only the first two problems are sketched here, focusing on elements concerning features of the data (for a more thorough description see Büscher, 2018).

In the first design experiment, the students work on the *Antarctic Temperatures Problem*. In this problem, the students take the roles of advisors to researchers in

Antarctica. The students are given temperature data from a research station in Antarctica (Figure 1). The data concern the daily temperatures for the three months of July 2013, 2014, and 2015 and asked to give a prediction for ten days of July 2016. This task was chosen as to allow insights into which features of the data the students intuitively focus on when giving such predictions.

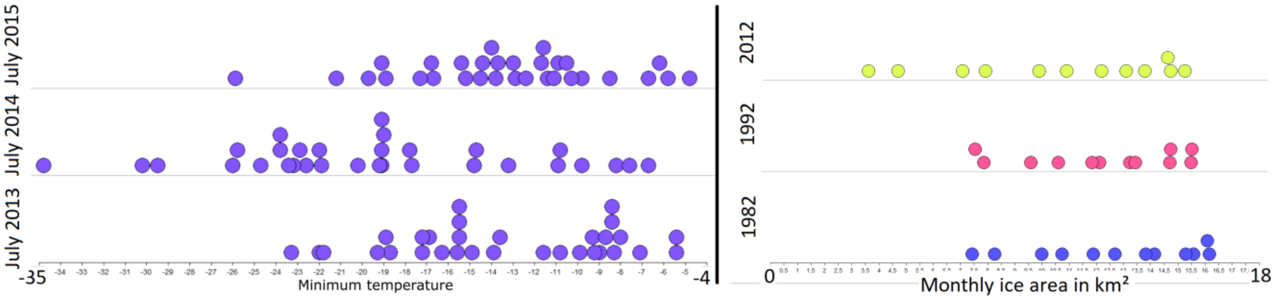

Figure 1: Data for the Antarctic Temperatures Problem (left side) and Arctic Sea Ice Problem (right side) (Translated from German)

After developing on ideas and measures, the students are given so-called filled-in *report sheets* (Figure 2). These report sheets are introduced to the students as short summaries produced by other students. The 'sketch'-field of each filled-in report sheet illustrates a different feature of the data: the 'Typical Report Sheet' focuses on a chunk in the centre, the 'Value Report Sheet' on a modal clump, and the 'MinMax Report Sheet' on extreme values and overall spread.

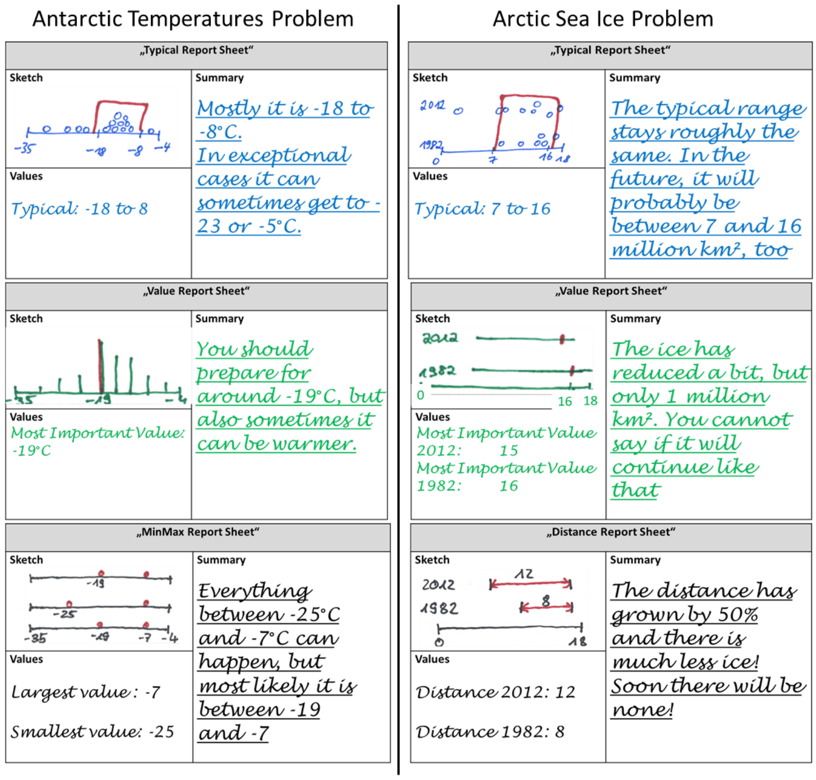

Figure 2: Filled-in report sheets for the Antarctic Temperatures Problem (left side) and Arctic Sea Ice Problem (right side) (Translated from German)

After the students are asked to evaluate the different filled-in report sheets, they are given an empty report sheet and prompted to create their own report sheet. This task allows to identify the features of the data preferred by the students at this point.

The second design experiment concerns the *Arctic Sea Ice Problem*. This problem follows a similar structure: this time, the students receive monthly Arctic sea ice data from 1982, 1992, and 2012 (Fig. 1, right side). They also again receive filled-in report sheets on Arctic sea ice (Fig. 2, right side), are asked to evaluate the report sheets, and create their own report sheet. This task allows this study to investigate whether a different context changes the students' focus on features of the data.

EMPIRICAL INSIGHTS INTO STUDENTS' FOCUS ON FEATURES OF THE DATA

This case study follows the students Jana and Mara during the first two design experiments. Empirical insights are provided in three steps of the development: students' initial predictions of Antarctic temperatures and their self-created report sheets for each of the problems.

Antarctic Temperatures Problem

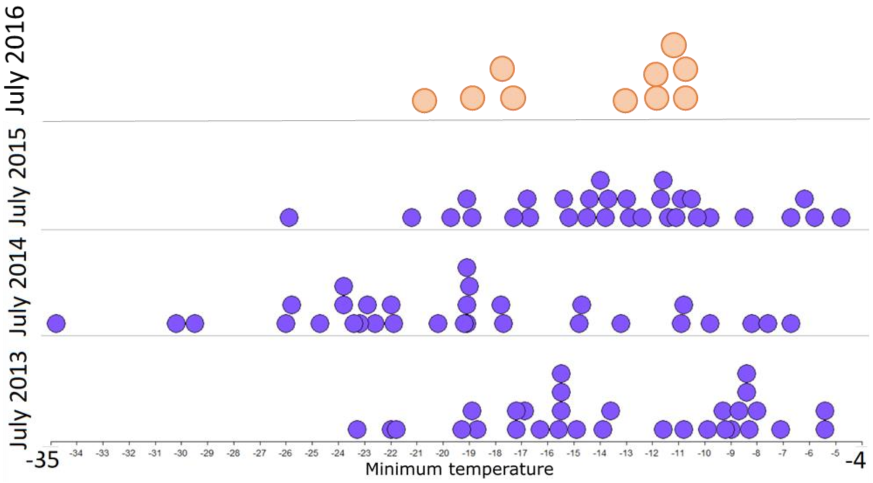

Figure 3: Jana and Mara's prediction for 2016

After giving a prediction for next year based on temperature data from three years mostly based on data from 2013 (Fig. 3), Jana (J) and Mara (M) explain their reasoning to the design experiment leader (DL).

63	J:	Well, we took those where there are as many days as possible, because…
64	M:	Because it will… it will most probably repeat itself, like…
65	J:	Because that's, like, the normality.
[…]		
70	M:	[…] there it's like, that in the core there are really many, ehm, like, those days were really often, and those [extreme values] are really far away, so that they, like, maybe only were exceptional temperatures.

The students identify the features of the data of a central ||*chunk of most data*|| ("as many days as possible", #63; the "core", #70) and of the ||*extreme values*|| (which are "far away" from the "core", #70). However, for predicting temperatures, the feature of the ||*chunk of most days*|| is more important, because <*the chunk of most data represents the normal temperatures*> (#65).

Later in the design experiment, the students create their own report sheet (Fig. 4). From this report sheet, it can be extrapolated that Jana and Mara did not focus on the features of the data emphasised by the Value Report Sheet and MinMax Report Sheet (Fig. 2), but instead continued to focus on the feature of the data of the ||*chunk of most data*|| (the "very frequent" days from -19 to -9, Fig. 3). Additionally, a short mention of the ||*extreme values*|| can be found.

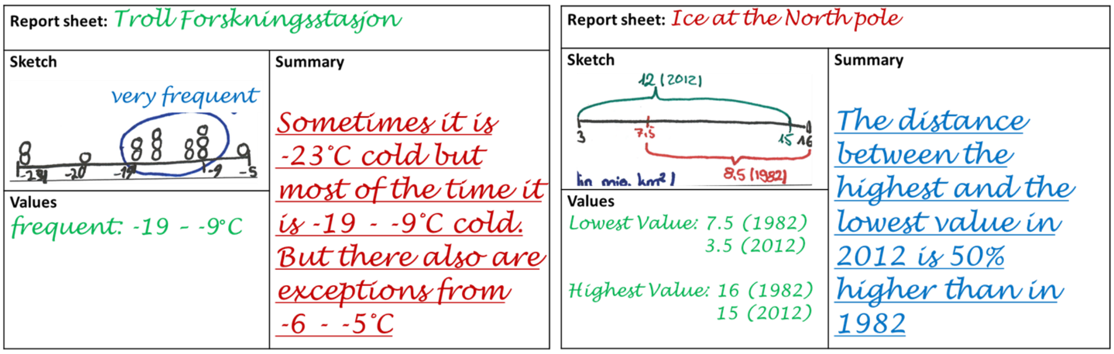

Figure 4: Jana and Mara's report sheets for the Antarctic Temperatures Problem (left side) and Arctic Sea Ice Problem (right side) (Translated from German)

Arctic Sea Ice Problem

Towards the end of the Arctic Sea Ice Problem, Jana and Mara create their own report sheet (Fig. 4). This report sheet seems to focus on other features of the data: instead of focusing on a central chunk of data, the students adopt the Distance Report Sheet (Fig. 2) and show the features of the data of ||*extreme values*|| as well as the ||*range*||. The design experiment leader asks the students to explain their reasoning.

232	J:	Because we wanted to – like, with the sketch we wanted to accurately explain the distance.
233	M:	And because the distance is larger, there also is less ice.
[…]		
240	DL	Mhm, I understand. Last time you gave a range. You didn't call it typical, but I don't remember what exactly you called it. When you talked about temperatures. I think you used an area, like where there were many dots. Now you dropped that. Why wasn't this as important to you now?
241	M:	Ehm – because – because for the temperature we were supposed to look at the temperatures for which they should prepare the most – and uhm here it is more important to see, how the ice melted and not where the most – so when, in which period – so – uhm – how much ice at most – there is.

This excerpt shows how the students purposefully focus on the feature of the data of ||*extreme values*|| and the "distance" (#233) between them. The Arctic sea ice melted more in 2012 than before, and <*the distance between extreme values represents the shrinking ice*> (#233). Challenged by the design experiment leader why they did not focus on the ||*chunk of most data*||, they argue by means of the context that <*for Arctic sea ice, the extreme values are more important than the chunk of most data*> (#241).

Summary

The two excerpts show how the students identify the different features of the data of the ||*chunk of most data*|| and ||*extreme values*||. Their focus however changes: whereas for Antarctic temperatures they focus on the ||*chunk of most data*||, they focus on the ||*extreme values*|| for Arctic sea ice. As Mara explains, this is directly influenced by the context of the teaching-learning arrangement. The different features of the data show a different *contextual relevance* depending on the context in question: For Arctic sea ice, the ||*extreme values*|| are more relevant than the ||*chunk of most data*||, and therefore, the students focus on this feature of the data. These phenomena of context-specific choice of features have also been found for the other focus students.

CONCLUSION

Research in statistics education has identified several features of the data 'intuitively' used by students (e.g. Konold et al., 2002). However, more insights into *why* students focus on specific features of the data are needed. This study shows how students' focus on features of the data cannot be simply understood as a 'natural' or 'intuitive' form of reasoning on data, but is instead actively pursued by the students themselves according to the perceived *contextual relevance* of particular features of the data. This has important theoretical as well as practical implications: learners' reasoning cannot be properly understood without paying attention to the context under investigation, and theoretical approaches should be chosen that are able to accommodate that fact – such as the Theory of Conceptual Fields (Vergnaud, 1996). Regarding the practical implications, this shows how the design of teaching-learning arrangements should explicitly acknowledge the possible contextual relevance of features of the data for the context at hand.

Context is already commonly held as especially important for the learning of statistics (e.g. Pfannkuch, 2011). This study adds another facet to the importance of context by illustrating possible influences of the context of a teaching-learning arrangement on students' reasoning on data. For this study, the in-depth analysis however only allowed to closely examine parts of the learning processes of three pairs of students. Further research is necessary to compare the results to learning processes of other students, and to find ways the influence of the context can be deliberately utilized in the design of a teaching-learning arrangement in order to support students' developing reasoning on data. This outlook is addressed in the overarching design research project (Büscher, 2018.)

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). Dordrecht: Springer Netherlands.

Biehler, R., Frischemeier, D., Reading, C., & Shaughnessy, J. M. (2018). Reasoning About Data. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 138–192). Cham: Springer International Publishing.

Büscher, C. (2018, in prep.). *Designing for Mathematical Literacy in Statistics [working title]* (Dissertation). TU Dortmund University, Dortmund.

Corbin, J. M., & Strauss, A. (1990). Grounded Theory Research: Procedures, Canons, and Evaluative Criteria. *Qualitative Sociology*, *13*(1), 3–21.

Gravemeijer, K., & Cobb, P. (2006). Design Research from the Learning Design Perspective. In J. van den Akker, K. Gravemeijer, S. McKenney, & N. M. Nieveen (Eds.), *Educational Design Research: The Design, Development and Evaluation of Programs, Processes and Products* (pp. 45–85). London: Routledge.

Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R., & Mayr, S. (2002). Students' Use of Modal Clumps to Summarize Data. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a Statistically Literate Society. [CD-ROM]*. Voorburg, The Netherlands: International Statistical Institute.

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data Seen Through Different Lenses. *Educational Studies in Mathematics*, *88*(3), 305–325.

Makar, K., & Confrey, J. (2003). Clumps, Chunks, and Spread out: Secondary Preservice Teachers' Reasoning about Variation. In C. Lee (Ed.), *Proceedings of the Third International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-3). [CD-ROM]*. Mount Pleasant, Michigan: East Michigan University.

Pfannkuch, M. (2011). The Role of Context in Developing Informal Statistical Inferential Reasoning: A Classroom Study. *Mathematical Thinking and Learning*, *13*(1-2), 27–46.

Prediger, S., & Zwetzschler, L. (2013). Topic-specific Design Research with a Focus on Learning Processes: The Case of Understanding Algebraic Equivalence in Grade 8. In T. Plomp & N. Nieveen (Eds.), *Educational Design Research - Part A: An Introduction* (pp. 409–423). Enschede, the Netherlands: SLO.

Schnell, S., & Büscher, C. (2015). Individual Concepts of Students Comparing Distribution. In K. Krainer & N. Vondrová (Eds.), *Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education* (pp. 754–760). Prague, Czech Republic: Charles University in Prague, Faculty of Education and ERME.

Vergnaud, G. (1996). The Theory of Conceptual Fields. In L. P. Steffe (Ed.), *Theories of Mathematical Learning* (pp. 219–239). Mahwah, N.J.: L. Erlbaum Associates.