



HAL
open science

Interpretation of statistical information and the role of contextual knowledge – a pilot study among fourth-grade students

Lisa Birk

► **To cite this version:**

Lisa Birk. Interpretation of statistical information and the role of contextual knowledge – a pilot study among fourth-grade students. Thirteenth Congress of the European Society for Research in Mathematics Education (CERME13), Alfréd Rényi Institute of Mathematics; Eötvös Loránd University of Budapest, Jul 2023, Budapest, Hungary. hal-04410795

HAL Id: hal-04410795

<https://hal.science/hal-04410795v1>

Submitted on 22 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Interpretation of statistical information and the role of contextual knowledge – a pilot study among fourth-grade students

Lisa Birk

University of Münster, Germany; lisa.birk@uni-muenster.de

The relevance of the development of statistical literacy has become apparent within the last decade as a consequence of our data-driven society. Therefore, fostering the competencies needed to interpret statistical information has become a well-accepted focus within the elementary school mathematics classroom. The role of contextual knowledge is highlighted and poses a crucial part of statistics. The here presented pilot study hence aims at describing situations in which fourth-grade elementary school students make use of their contextual knowledge when interpreting statistical information presented in stacked dot plots to contribute to a local learning theory that explains when and how the integration of contextual knowledge might foster or hinder the interpretation of statistical information. The role of the task is highlighted as connections can be established between the integration of contextual knowledge and what the task aims at.

Keywords: Early statistical thinking, contextual knowledge, interpreting statistical information.

Early statistical thinking and the role of context – theoretical background

From a young age onwards, children are confronted with data and various forms of visual data representations. Especially the necessity to interpret these as well as a critical evaluation of claims coming from presented data are competencies linked to responsible citizenship. Statistics educators therefore agree that learners need to engage with statistics from an early age onwards (English & Watson, 2015). To be able to prepare students for our data-driven society, more and more research is conducted on early statistical thinking to develop learning theories as well as teaching materials and approaches. This also has an impact on elementary curricula worldwide (Leavy et al., 2018).

Statistical literacy has become a central goal of statistics education in the last decades (Gal, 2004), addressing the interpretation of data (representations) as one of multiple subsets of competencies. When considering the concrete teaching contents from the perspective of statistics education, one focus is set on data and its distributions. It has been broadly accepted that a more aggregate view of data should already be adapted in an elementary setting to recognize distributions as a whole with respective characteristics (Konold et al., 2015). Friel, Curcio, and Bright (2001) have created a framework of three levels of graph comprehension that describes the interpretation of graphical data representations and integrates a shift from local properties of data towards more global concepts.

The role of context is essential in statistics due to the underlying understanding of data: they are regarded as naturally in-context-embedded numbers (Ben-Zvi & Aridor-Berger, 2015). It is of no surprise that context, therefore, is also considered a crucial component in the development of statistical literacy and has been the focus of various studies in the last years. Here, we define contextual knowledge as personal experiences, beliefs and factual knowledge from the real world that concerns that context portrayed in the statistical investigation. Contextual knowledge has been

understood to potentially influence interpretations of data so that a closer consideration might help understand learning processes within the statistics classroom: When interpreting statistical information, for instance, conclusions can be based on the data, on contextual knowledge, or an integration of both approaches. Differences between the contextual knowledge and the conclusions drawn from the data presented might function as a learning opportunity or as a challenge when data is ignored in favour of contextual knowledge (Ben-Zvi & Aridor-Berger, 2015). In science education, a taxonomy has hence been proposed displaying how students and scientists react to data that does not conform to their experiences and conjectures – ranging from ignoring the data to a reinterpretation of the portrayed data to fit their conjecture to a change in their personal theory due to the data they were presented with (Chinn & Brewer, 1993). Two prior studies are of special interest concerning the potential influence of contextual knowledge: Frischmeier & Schnell (2021) investigated how elementary students cope with data that contradicts their conjectures formed based on their contextual knowledge. They found that students' expectations and conjectures influence the interpretation of the data. In a different study, Ben-Zvi & Aridor-Berger (2015) investigated informal statistical inferences and how students switch between the data world and the real world. It was highlighted that the integration of both worlds can already be established among young learners and is regarded as a step towards better statistical reasoning. Even though more research is conducted investigating the role of context, there is no study yet on how and in which situations elementary students integrate their contextual knowledge when interpreting statistical information given in form of stacked dot plots.

Integration of contextual knowledge into the interpretation of statistical information – a design-based research study

Due to the lack of prior research on situational differences integrating contextual knowledge at the elementary level, its role in the interpretation of statistical information in form of stacked dot plots is explored within this research study. It functions as a pilot study for a larger research project: with a design-based research approach (Prediger & Zwetschler, 2013), we will investigate how to develop thinking in distributions (Biehler, 2007) and a more aggregate view on data (Konold et al., 2015). The focus of the project will be on examining the role of contextual knowledge in the development of said competences. For that, three cycles are planned that generate research-based teaching learning arrangements as well as local learning theories on different aspects of the general research aim.

Research question

To investigate the role of contextual knowledge in form of personal experience, the research question for this paper is the following: In how far do elementary school students integrate their contextual knowledge when interpreting stacked dot plots on topics that are within their reach of personal experience?

Data collection and participants

To answer the research question, nine half-standardized interviews (6 interviews with an individual child, 3 interviews with a pair of students each) were conducted as part of a master's thesis (Schulz, 2022). Twelve students of a fourth grade (9 to 10 years old) of a German elementary school were

asked to work on four tasks concerning the interpretation of statistical information in form of stacked dot plots and then elaborate their answers to the interviewer. Tables and bar graphs on categorical data had been introduced but there was no explicit intervention before the interviews. Four different task contexts were chosen for the presented distributions: shoe sizes of a fifth grade, highest daily temperatures in April, books read within three months, and minutes needed for students' way to school. The presented distributions varied in sample size. The tasks differed from closed formats such as rating statements (e.g. statement "girls have larger feet than boys" with options *true*, *false*, and *I cannot tell*), to half-open tasks focusing on different levels of graph comprehension (e.g. imagining a hypothetical distribution for a different sample) (Friel, Curcio & Bright, 2001), to open tasks allowing students to freely describe the distribution (e.g. what can you tell from this graph?).

Three different methods were chosen to collect the data: three students worked on the presented task individually, three students worked on the tasks following the approach of productive irritation, presenting the students with fictitious statements made by other students eliciting a need for correction or justification (Nührenbörger & Schwarzkopf, 2016) and three pairs of students asked to work on half of the tasks and then presenting their findings to the other students present in the interview. These approaches were chosen to consider differences in students' explanations in different interview settings for the following cycles of the design-based research study. The interviews of the students that worked individually were conducted while students were working on the tasks. For the three pairs of students, interviews were conducted after the students had worked on the tasks.

Data analysis

The transcripts of the nine interviews form the basis for the here presented findings. Following the method of qualitative content analysis (Mayring, 2014), categories were formed inductively after the first sighting of the data. This approach was chosen to truly depict the situations as found within the data as no theories on the situational use and integration of contextual knowledge when interpreting statistical information were found in the research literature.

Three different categories were identified to describe the students' integration of their contextual knowledge: firstly, students integrate their contextual knowledge when asked to think beyond the data and make up a hypothetical distribution (hypothetical distribution); secondly, students refer to their contextual knowledge when asked for certain characteristics of the given distribution while not making use of the data (characteristic of given distribution); and lastly, students make use of their contextual knowledge when validating the data presented in the distribution (validation) (Table 1).

Categories are not mutually exclusive so that certain student statements can be considered as falling under more than one category (see transcript of student TA). The coding of the transcripts was undertaken in collaboration with colleagues to ensure semantic validity. To consider the impact of the respective tasks, categories are allocated to the tasks and possible influences are discussed (Table 2). The interviews of the pairs of students were treated as a unit as potential influences of having to explain the results and interpretations to a partner cannot be ruled out at this point.

Table 1: Coding manual for the different situations of integration of personal experience

	Definition	Example
Hypothetical distribution (hd)	When asked to think beyond the data and describe a hypothetical distribution in a similar context, students refer to their contextual knowledge to explain their choice.	“Would the distribution look differently if the children of a seventh grade had been asked? If so, how? Yes, I actually know that, because my brother is in seventh grade [...] and he already has very big feet.”
Characteristic of given distribution (cd)	Students are asked to elaborate on a certain characteristic of the given distribution. Instead of giving a data-based answer, they refer to their contextual knowledge in form of personal experiences.	“What temperature would you say is typical [for this distribution]? Why? So, in summer I would regard a temperature of 20 as typical, because in summer it's usually warmer. In winter I would think 6, 8 degrees or so is typical, because it is winter. And in spring 10, 12, 14, 16 degrees”
Validation (v)	Students refer to their contextual knowledge when validating the data (representation) or integrate their personal data as a form of validation.	“So, okay, uh, that's very funny now, that a fifth grader has shoe size 27. [...] Most of them have shoe size 35. I would have guessed 39 or 38.”

Results

To answer the guiding research question, excerpts of the transcripts from all three data collection methods and three of the four different tasks will be presented to explore the categories. In a last step, results from the analysis of all nine interviews will be presented with the help of a table portraying the different situations in which the students make use of their contextual knowledge, allocating the situations and regarding categories to the tasks posed.

When asked to describe a hypothetical distribution in a context similar to the one in the task, all students referred to their personal experience as a form of contextual knowledge to explain their answer. They either directly linked the distribution they made up to a person that they know that would fit the sample (see transcript of student OL) or they integrated what they knew about the situational context, for instance arguing that shoe size is connected to height and hence, older students who tend to be taller have larger feet.

Student OL: (reads aloud) *Would the graph change if the children of a seventh grade had been asked? If so, how?* Yes, I actually know that, because my brother is in seventh grade [...] and he already has very big feet. So, I think then I would say there would be nobody under 36 here.

The task might have a strong impact here as it purposefully reaches beyond the data and asks for an inference which might be impossible to come up with without integrating contextual knowledge. Yet, especially the integration of facts gained from personal experience as seen in the above-given

transcript might be regarded as descriptive of the integration of contextual knowledge. The students were not asked to make up distributions in the other tasks.

When asked for an answer that could be deduced from the presented data set, some of the students substituted the interpretation of the given data with personal experience. In the following transcript (see transcript of students TO and LE), students TO and LE negotiate whether to rely on the data presented or on their own experience and attitudes as a form of contextual knowledge. LE tries to guide TO back to the question, arguing that TO's answer does not conform to the initially posed task but TO repeatedly disregards the visually displayed data in favour of his own attitude. Nonetheless, he shows a slightly more aggregate view on data here when referring to the middle range when asked for a typical temperature for the data set.

- Student TO: And the last question was (reads aloud): *Which temperature would you regard as typical here? Why?* [...] Um, I said 12 degrees as a typical temperature [...], because that's just not too high, not too low, because that's so
- Student LE: The middle
- Student TO: In the middle range and
- Student LE: And why not 10? From 20 and 0 degrees.
- Student TO: Because 10, because 10. Wait, I have to think for a minute. Because, um, 10 is just so cold. I think 12 is comfortable.
- Student LE: But it, it doesn't ask what you [find pleasant].
- Student TO: Yeah, I know.] But I would say, I find 12 is a typical um normal temperature.

Another situation in which students integrated their contextual appeared when students tried to validate the data represented in the distribution. Students either argued that some data seemed unrealistic to them because they had personally or through close relatives experienced other values for the variable covered in the graphical representation of the data set or they commented on the validity of the data by comparing the presented distribution to their own value of the variable (see transcript of student TA). By contrasting the presented values to their personal ones, students appear to try and make sense of the data. They do not seem to see a difference between values that come from within the sample of the distribution and their own values for the variable of interest.

- Student TA: [...] I would estimate that it would take me about 15 minutes by foot and only 10 minutes by bike. Yes. (I: Mhm.) But that's not normal that children need 30 minutes or here, 55 minutes. Then I would rather say, um, recommend that if you need 55 minutes to get to your school, that you look for one close to your place (I: yes), where there is also one.

When summarizing the analysis, one can conclude that the integration of contextual knowledge, especially in form of personal experience, appears to be of high relevance when elementary school students describe and interpret distributions (Table 2). Firstly, all elementary school students integrated their contextual knowledge when asked to surpass the given data to come up with a hypothetical distribution in a to them well-known context (9 out of 9 students/student pairs). Here, an integration of contextual knowledge seems rather logical as there is no further data given which could be used to solve the task. The integration of contextual knowledge most frequently occurred when students were asked for a certain characteristic of the given distribution (22 out of 43 occurrences), even though answering the question did not rely on the integration of contextual knowledge. In these situations, students tended to rely on their personal experience and beliefs as forms of contextual knowledge rather than on the data they were presented with (see transcript of

students TO and LE). Students validating the presented data was identified as the third situation in which elementary school learners integrated their contextual knowledge. Throughout the interviews, all students at one point or the other referred to their personal experience as a form of contextual knowledge to elaborate on the validity of the data. Most students made use of this during the task on minutes that different children needed for their way to school. The task followed an open format in which the learners were asked to describe what they could tell from the graph they were presented with. This hints towards a potential connection between either the data context, the validity of the data or the open format and the perceived necessity to validate the presented data by referring to the students' contextual knowledge which should be further investigated. An integration of the contextual knowledge and the interpretation of the given data might be considered a goal during this task as it would allow for a critical consideration of both the presented data and the personal experience from which the students draw. It should be investigated whether said integration leads to a more holistic picture of the context of the data.

Due to the frequency of the integration of contextual knowledge when describing and interpreting data (43 different situational occurrences of the integration of contextual knowledge, 3 tasks without integration of contextual knowledge; Table 2), one could argue that there does not seem to exist a distance between the presented data and the students working on the task. The integration might hinder making data-based decisions at this point while the critical evaluation of portrayed data only becomes possible due to its use.

Table 2: Different forms of the integration of contextual knowledge, separated by interview and task

	BE	JO	OL	HE	JA	TA	FE & BN	LE & TO	LO & CO
Shoe sizes	hd	hd	cd/hd	v/hd	hd	cd/hd	cd/hd	cd/hd	v/cd/hd
Temperature	cd	cd	cd/v	cd/v	cd	cd	cd	cd	cd
Books read	cd	cd	cd	cd	-	cd	cd	-	cd
Minutes for way to school	v	v	v	-	v	cd/v	v	v	v

Interpreting statistical information with the help of contextual knowledge – first findings

Statistical literacy has become a crucial goal within the statistics classroom if we aim at enabling (young) learners to take an active role in our data-driven society. The expanding integration of statistics in elementary school curricula is therefore a logical consequence. Still, there is very little research on the interpretation of statistical information among elementary school students. The use of data contexts that students can relate to has repeatedly been demanded (Leavy & Hourigan, 2019) – nonetheless, the situations in which elementary school students integrate their contextual knowledge when interpreting statistical information has not been investigated yet. This is where the

pilot study proceeds: it aimed at identifying situations in which students integrate their experience when interpreting statistical information to prepare a deeper understanding of the matter.

We were able to show that for the nine interviews conducted elementary school students of a fourth grade integrated contextual knowledge in most of the tasks posed. Especially when asked about a certain characteristic of the given distribution, students frequently did not refer to the data portrayed but rather their contextual knowledge which is in accordance with the findings of Frischemeier & Schnell (2021). When considering the taxonomy of Chinn & Brewer (1993), answers could be identified as rather experience- than data-based. The presented data was rather ignored than reinterpreted, changes in theory could not be observed due to the design of the study. Other situations in which the elementary school students integrated their personal experience included when being asked to imagine a potential distribution in a similar context and when evaluating the given data as part of their interpretation. As categories and tasks show a high consistency, one can conclude that the tasks influence whether and how students integrate their personal experiences. The study highlights the role of contextual knowledge in the form of personal experience for the interpretation of statistical information which needs further research to ensure the consideration of underlying potentials for the development of statistical literacy, especially in young learners.

The here presented pilot study aims at providing first insights into when and how elementary school students of a fourth grade integrate their personal experiences when interpreting statistical information. Some limitations need to be considered: Even though first categories could be formed, the need for a deeper understanding of the potential learning opportunities and challenges of the integration of contextual knowledge became evident. An interpretative approach should be chosen next to investigate patterns in the respective answers. The data was collected in three different settings which might have had an influence on the results – to ensure comparability, possible influences should be further analysed. Another limitation is the sample size of the pilot study. Further interviews and coding by other researchers seem necessary to generalize the findings. As next steps, further interviews should be conducted to enable a more holistic picture. Also, the influence of the tasks is highlighted. Tasks should therefore be constructed that do not enforce the integration of contextual knowledge and enable all situational integrations. Within the next cycles of the design-based research study, a stronger focus will be put on thinking in distributions and the question of whether the integration of contextual knowledge might be regarded as helpful in the development of thinking in distributions or a certain distance to the distribution might be more fruitful as proposed by Frischemeier & Schnell (2021). The here presented insights guide the design of the next teaching materials and tasks (i.e. with regards to wording, openness) to be put in practice in classroom settings.

References

- Ben-Zvi, D. & Aridor-Berger, K. (2016). Children's wonder how to wander between data and context. In D. Ben-Zvi & K. Makar (Eds.), *The teaching and learning of statistics* (pp. 25–36). Springer. https://doi.org/10.1007/978-3-319-23470-0_3
- Biehler, R. (2007). Denken in Verteilungen – Vergleich von Verteilungen. [Thinking in distributions – comparing distributions.] *MU Der Mathematikunterricht*, 53(3), 3–11.

- Chinn, C. & Brewer, W. (1993). The role of anomalous data in knowledge acquisition: a theoretical framework and implications for science instruction. *Review of Educational Research*, 63(1), 1–49. <https://doi.org/10.3102/00346543063001001>
- English, L. D., & Watson, J. M. (2015). Statistical literacy in the elementary school: Opportunities for problem posing. In F. M. Singer, N. F. Ellerton, & J. Cai (Eds.), *Mathematical problem posing* (pp. 241–256). Springer. https://doi.org/10.1007/978-1-4614-6258-3_11
- Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 32(2), 124–158. <https://doi.org/10.2307/749671>
- Frischemeier, D., & Schnell, S. (2021). Statistical investigations in primary school – the role of contextual expectations for data analysis. *Mathematics Education Research Journal*, 1–26. <https://doi.org/10.1007/s13394-021-00396-5>
- Gal, I. (2004). Statistical literacy. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 47–78). Springer. https://doi.org/10.1007/1-4020-2278-6_3
- Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics*, 88(3), 305–325. <https://doi.org/10.1007/s10649-013-9529-8>
- Leavy, A., Meletiou-Mavrotheris, M., & Paparistodemou, E. (Eds.). (2018). *Statistics in Early Childhood and Primary Education – Supporting Early Statistical and Probabilistic Thinking*. Springer Nature. <https://doi.org/10.1007/978-981-13-1044-7>
- Leavy, A. M., & Hourigan, M. (2019). Expanding the Focus of Early Years Mathematics Education: Statistics and Probability. In O. N. Saracho (Ed.), *Handbook of research on the education of young children* (pp. 99–112). Routledge. <https://doi.org/10.4324/9780429442827-7>
- Mayring, P. (2014). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution*. https://doi.org/10.1007/978-94-017-9181-6_13
- Nührenböcker, M. & Schwarzkopf, R. (2016). Processes of mathematical reasoning of equations in primary mathematics lessons. In K. Krainer & N. Vondrová (Eds.), *Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education (CERME9, 4-8 February 2015)* (pp. 316–323). Prague, Czech Republic: Charles University in Prague, Faculty of Education and ERME. <https://hal.science/hal-01281852>
- Prediger, S., & Zwetschler, L. (2013). Topic-specific design research with a focus on learning processes: The case of understanding algebraic equivalence in grade 8. *Educational design research – Part B: Illustrative cases*, 407–424.
- Schulz, A. (2022). Denken in Verteilungen im Mathematikunterricht der Primarstufe – Eine explorative Interviewstudie mit Grundschüler:innen [Unpublished master's thesis]. University of Münster, Germany.