Statistical investigations in primary school – the role of contextual expectations for data analysis

Daniel Frischemeier¹ · Susanne Schnell²

Received: 1 September 2020 / Revised: 19 August 2021 / Accepted: 27 August 2021
© The Author(s) 2021

Abstract

As data are ‘numbers with context’ (Cobb & Moore, 1997), contextual knowledge plays a prominent role in dealing with statistics. While insights about a specific context can further the depth of interpreting and evaluating outcomes of data analysis, research shows how it can also hinder relying on data especially if results differ from expectations. In this article, the aim is to investigate how young students informally deal with empirical evidence, which differs from their initial expectations in a specific context. We present a case study with three pairs of students at the age of 9 to 10 who compare groups in survey datasets. The interpretative analysis shows how conjectures of varying degrees of confidence shape the students’ statistical expectations and can play different roles in interpreting results from data analysis.

Keywords Early statistical thinking · Comparing groups · Conjecturing · Context · TinkerPlots

Introduction

In the era of fake news and alternative facts, a competent understanding and handling of data are essential for the process of supporting pupils in their development to engaged citizens (Engel, 2017). Therefore a conceptual understanding of data and statistical reasoning should be developed as soon as possible, preferably already in primary school (Ben-Zvi, 2018). A recent overview of studies on the development

✉ Daniel Frischemeier
dfrische@uni-muenster.de

Susanne Schnell
schnell@math.uni-frankfurt.de

¹ Fachbereich Mathematik Und Informatik, WWU Münster, Fliednerstraße 21, 48149 Münster, Germany

² Institute of Mathematics and Computer Science Education, Goethe University Frankfurt, Robert-Mayer-Str. 6-8, 60325 Frankfurt, Germany

of early statistical thinking of young learners can be found in the book of Leavy et al. (2018). There are several approaches to introduce young students in the practice of statistics (Watson & English, 2015); one way to engage them in statistical projects is by using real and meaningful data from a context the children are familiar with. For example, a question like “Who tends to have more games on their smart-phone – children in grade 3 or grade 4?” is a typical hook to initiate meaningful statistical inquiries, collecting and exploring data, comparing groups, and making (informal) inferences. Activities like these aim at fostering statistical reasoning, i.e., making sense of statistical information and constructing and evaluating data-based arguments (Ben-Zvi & Garfield, 2004, p. 7).

However, even before beginning a statistical investigation, students often already have expectations regarding the outcome, especially if the context of the data are related to their everyday life. Similar to mathematics education (Smith et al., 1993), pre-existing knowledge and expectations can play an important role in guiding the statistical investigation and interpretation of the results. Especially the role of contextual conjectures that diverge from the results of a statistical investigation has yet to be clarified from a statistics education perspective.

The aim of this article is to investigate how young students informally deal with empirical evidence, which differs from their initial expectations in a specific context. We will argue that addressing pre-existing expectations can be crucial for fostering the development of statistical reasoning. After introducing the key ideas and activities for statistics education in primary school, we discuss empirical studies in terms of possible benefits and obstacles from context knowledge. Our case study with third graders working on a group comparison task focuses on situations, in which the students’ expectations are not met by the outcome of the data analysis. As the empirical excerpts show, these contextual expectations can be translated into more statistical conjectures and influence the way different measures are interpreted. Ultimately, a better understanding of the role of pre-existing conjectures is needed to develop adequate teaching and learning materials.

Literature review

Developing early statistical reasoning in primary school

One of the most important overall goals of statistics education in school is to promote statistical reasoning:

“Statistical reasoning may be defined as the way people reason with statistical ideas and make sense of statistical information. This involves making interpretations based on sets of data, representations of data, or statistical summaries of data” (Garfield & Gal, 1999, p. 207).

To achieve this, it is widely suggested (Leavy & Hourigan, 2018) to use real-world data that is meaningful to children, and provide them with technological tools so that learners are able to explore large and multivariate data sets and to

focus on key concepts of statistics rather than introducing singular procedures and tools (Garfield & Ben-Zvi, 2009). Important statistical concepts include distribution (as aggregated, global view on data), representations, variability, uncertainty, etc. (Burrill & Biehler, 2011; Garfield & Gal, 1999).

Especially for young learners who do not have formal statistics knowledge, conceptual foundations should be built on their informal ideas for example for measuring center and spread as conceptual before introducing formal procedures later in school (Konold et al., 2002). One of the informal ideas are ‘modal clumps’ (in German referred to as ‘hills’), which were introduced by Konold et al. (2002) to refer to “a range of data in the heart of a distribution of values. These clumps (cf. Figure 1) appear to allow students to express simultaneously what is average and how variable the data are” (p. 1) (cf. Figure 1). Several empirical studies show how modal clumps support young students in describing and interpreting distributions and there is empirical evidence that modal clumps can facilitate young learners thinking to get first ideas of center and spread at an early stage (Bakker & Gravemeijer, 2004; Fielding-Wells, 2018; Frischemeier, 2019; Konold et al., 2002). In this respect, Konold et al. (2002) showed that 7 to 9 grade students made successful use of modal clumps to get first notions of center and variability of distributions. Bakker and Gravemeijer (2004) investigate middle school students’ (grade 7) attention to center, spread and shape when reasoning about numerical data. The participants of their study tended to reason from data points and then used modal clumps to describe aspects of center and spread of the distributions.

Building upon that, a variety of studies have shown (Fielding-Wells, 2018; Frischemeier, 2019; Watson & Moritz, 1999) that one type of activity fruitful for developing pre- or informal statistical reasoning is to compare groups. At the heart of this activity is describing and comparing distributions by referring to their shapes, centers, spreads, and naturally addressing several fundamental statistical

Fig. 1 TinkerPlots representation of the distribution of the number of games on the smartphone for 3rd graders (bottom) and 4th graders (top) including medians (8 vs. 4), means (~ 9.2 vs. ~ 5.4), boxplot-precursors and drawn-in hills; means were not discussed with students

concepts (see Burrill & Biehler, 2011). Very early work on young students' reasoning when comparing distributions has been done by Watson and Moritz (1999). In their qualitative interview study they found out that "students used numerical and visual strategies, either individually, or in conjunction with each other, to make comparisons between the data sets presented in graphs" (p. 163). In addition, Watson and Moritz (1999) showed that already young learners are able to develop statistical reasoning for comparing two distributions, e.g., by taking into account visualization strategies. Fielding-Wells (2018) exemplifies how even young students (ages 10–11) used statistical representations such as dot plots as data models to compare groups. In addition to that, there are empirical studies reporting young learners using medians (cf. Schnell & Frischemeier, 2020).

Inherent in statistical exploration is the need to account for uncertainty, especially when inferring from a sample to an unknown population (Ben-Zvi et al., 2012; Makar & Rubin, 2009). In order to do so, a probabilistic language is needed (Makar & Rubin, 2009), which for instance makes use of relativizing expressions such as 'maybe' (or 'may be'), 'roughly', or the use of conjunctive. Ben-Zvi et al. (2012) observed how young students' articulations of uncertainty emerged in the processes of informally analyzing gradually increasing samples. In their study with 10- to 11-year-old students, the participants began using probabilistic language "that articulates the level of confidence or uncertainty in statistical tendency or a prediction" (p. 916). In different episodes, the young learners changed between deterministic and relativistic statements and "tended to either express extreme confidence in knowing something (certainty-only) or express that nothing could be concluded (uncertainty-only)" (p. 923).

The role of context and contextual expectations in statistical investigations

In contrast to mathematics where numbers often are represented in their abstract form (Gattuso & Ottaviani, 2011), Cobb and Moore (1997) emphasize the role of context when defining the term data: "data are not just numbers, they are numbers with a context" (p. 801). Integrating statistical and contextual information, knowledge and conceptions is a crucial part of statistical reasoning: "context determines how and what data to collect, as well as how to analyze the data and interpret the results. This results in a constant interplay between considering a statistical problem and the context of the problem (Groth, 2007; Wild & Pfannkuch, 1999)" (Weiland, 2019, p. 19). One important prerequisite for being able to even conduct data analysis is the awareness of a "need for data" (Wild & Pfannkuch, 1999), which means "the recognition of the inadequacies of personal experiences and anecdotal evidence leading to a desire to base a decision on deliberately collected data" (p. 227).

Being familiar with a certain context can be beneficial for doing statistics: Langrall et al. (2011) emphasized in a study with 18 middle school students the positive effects of contextual expertise on in-depth data exploration, especially on drawing and justifying conclusions. They stated "that students used context knowledge to (a) bring new insight or additional information to the task, (b) explain the data, (c) provide justification or qualification for claims, (d) identify useful data for the task at hand, and (e) state facts that may enhance the picture of the data but are irrelevant to the process of

analyzing the data” (p. 47). Additionally, according to Pfannkuch and Rubick (2002), contextual knowledge is also “essential for conjecturing possible relationships within the set of data. Both contextual and statistical knowledge influenced students’ understanding and interpretation of the data” (p. 16).

However, strongly held expectations can also hinder an open data exploration; for instance, the well-known confirmation bias leads to interpreting (statistical) information in a way that confirms pre-existing expectations and discharging other, contradicting evidence (Jermias, 2001, p. 146).

Chinn and Brewer (1993) proposed a theoretical framework of how science students and scientists react when their beliefs about the physical world conflict with the information in data. The two extreme positions are that students totally ignored anomalous data, or that students completely changed their theory because of their findings in the data. According to Chinn and Brewer (1993), different factors such as the assumed validity of the data can also play a crucial role in how data, pre-existing beliefs, and theory are coordinated.

Legare et al. (2016) concluded from an experiment with children age 3 to 6 that dealing with situations deviating from expectations can be a powerful tool for deepening learning as well as challenging pre-existing beliefs “by increasing awareness of uncertainty and the potential for multiple interpretations of the same information” (p. 54). In addition, Busch and Legare (2019) found out in another experimental setting with 5- to 9-year-old students that inconsistent and ambiguous evidence motivated the students to deepen their inquiries, seeking additional data, and exploring alternative explanations. In terms of age-related development, the study argues that even young children are sensitive to inconsistencies in observed data, but the ability to notice incomplete data and to seek out further information in response to these discrepancies develops over time.

In contrast, Masnick et al. (2007) investigated how primary school children (grade 2 and grade 4) identify patterns and noise in data from physical experiments concerning preexisting context knowledge. Their results showed that context-related beliefs influence the way students draw conclusions from data: When their expectations matched the data, they were able to generate adequate explanations for sources of variability in the data. When their expectations were not met, children showed difficulties integrating new data and were not likely to revise their beliefs. While the results of these studies are consistent with others, the authors themselves mention the data collection in highly coordinated and structured psychological experiments as a limitation. It stays an open question how young students deal with inconsistencies in a more open, statistically oriented approach as in an exploratory data analysis setting.

For this, in addition to contextual expectations, Dvir and Ben-Zvi (2021) also highlight the role of *statistical* expectations (which they call “conjecture models”), which are understood as a developing expectation of the look of a fitting model for the data (e.g., a linear regression or a bell shape, Dvir & Ben-Zvi, 2021). The authors discuss how these statistical models also have an impact on the data analysis process and that a strongly held [statistical] conjecture can indeed promote internal motivation to deepen inquiry, or even to ‘prove’ the conjectures’ legitimacy (Dvir & Ben-Zvi, 2021). However, they state that conjectures can also hinder open data exploration as they can lead students to seek mainly for their affirmation.

Overall, these findings imply that contextual expectations are an important component in learning statistics: context, data, and statistical tools have to be coordinated. Especially in case of a mismatch between expectations and results from data analysis, students seem to have difficulty relying on the data. Even if pre-existing conjectures might be hindering open data explorations, they should at least be made explicit. For this, Biehler (2001) suggests asking learners to articulate their expectations and conjectures (“which results do you expect?”, “which differences do you expect?”, “which shape of the distributions do you expect?”) before they engage in the statistical investigation.

While many of the reported studies look at data from physical or psychological experiments, statistics education in primary school often works with social statistics from surveys, e.g., data on students’ age, hobbies, opinions, leisure time and sporting activities, etc. Contextual expectations for this type of data can draw on the children’s own experiences and beliefs from their personal situation. Therefore, they might play an even bigger role than for experimental data.

Overall, this paper aims at investigating how young learners deal with discrepancies between their pre-existing conjectures in a specific context and the results from data analysis in survey data. Thus, the paper addresses the following research questions:

- (1) *How do young students express their initial conjecture in terms of confidence and uncertainty, and how does this relate to their conjectured distribution of data?*
- (2) *How do young students deal with data explorations in survey data that are opposed to their initial conjectures?*

Specifically, we will investigate how our young students with conjectures of varying degrees of confidence develop *data-centered insights* and how they relate their *data-centered insights* to their *initial conjectures*. Based on this literature review we define *contextual knowledge* as knowledge and beliefs about daily life and experiences. In line with Biehler (2001), we consider *conjectured distributions* as distributions of data, which are drawn and sketched by learners with regard to their expectations and conjectures. In addition to that, we regard *initial conjectures* as inferences about the context before the beginning of the data exploration. *Data-centered insights* are insights, which the learners get from the exploration of the data referring to and comparing different and specific facets of the distributions. We will define the relevant constructs *initial conjectures*, *conjectured distributions*, and *data-centered insights* from the aspects raised in this literature review after introducing the framework of the case studies, tasks, and theoretical constructs for analysis.

Methodology

Methodological framework

The presented study was situated in a larger framework of an ongoing design research project, aimed at facilitating young students’ statistical learning and reasoning. For this, a teaching unit was developed (as presented below), in which students

work with real data by using the statistics education software TinkerPlots (Konold & Miller, 2011) as a digital tool for data exploration activities. This paper focuses on the case study (Yin, 2014) of three pairs of students (grade 3, ages 9–10), which was conducted after the classroom implementation of the teaching unit in a German primary school. For this, interviews (roughly 30–40 min each) were conducted to gain deeper insights into students' reasoning and specifically to study how young learners deal with discrepancies between their expectations and the results from the analysis of survey data sets.

Teaching unit

While this article focuses exclusively on the interview data, the content of the teaching unit serves as the foundation (both in regards to technical procedures and conceptual understanding) and is thus briefly illustrated here.

The teaching unit for grade 3 consisted of 13 lessons—each lasting 45 min—and was taught by the mathematics teacher of the class with support from a prospective teacher as a research assistant. At first (lessons 1–4) the students were introduced to the basics of data analysis, e.g., what is data, and how to collect and represent data with hand-written data cards. In lessons 5 to 10, students were introduced to the software TinkerPlots (Konold & Miller, 2011) and learned to create and interpret different representations of survey data, such as stacked dot plots. Furthermore, formal measures (e.g., the median) and informal measures (e.g., modal clumps) were developed. TinkerPlots served as a tool, which enabled the students to explore large data sets, to create distributions of numerical variables (e.g., in form of stacked dot plots), and to use hills and medians to describe and compare groups.

As the final project in lessons 11–13, the students used TinkerPlots and worked in groups and were asked to generate a question leading to an exploratory group comparison activity on a provided sample dataset about leisure time activities and media use of primary school students in the form of survey data. This dataset (called “Grundschulen NRW”) is a real data set of more than 600 primary school students collected by Engels (2018). The teaching unit concluded with the students' presentations of their findings.

Participants and data collection

To gain in-depth insights into students' informal statistical reasoning and the role of contextual expectations, an interview study was conducted as part of the overall research project. A convenience sample of five pairs of grade 3 students (three girl groups, two boy groups; ages 9–10) were interviewed by the research assistant roughly two weeks after the end of the teaching unit. Pairs were formed of students who already had worked productively together before.

Out of the five pairs, two already had expectations that matched the outcome of the data analysis. Following Legare et al. (2016) and Busch and Legare (2019), this

paper focuses on the other three pairs of students, as we believe that dealing with such a conflict allows for deeper insights into the students' reasoning.

In the interviews, the children were provided with a sample (461 primary school students, 236 in grade 3 and 225 in grade 4) of the Grundschohlen NRW data set (Engels, 2018) with variables about leisure time activities and media use.

On this basis, students were asked to address the statistical question of whether 3rd graders tend to have more or fewer games installed on their smartphones than 4th graders. The question was chosen as it resembled the investigative questions students came up with in class. Furthermore, the children had expressed interest in the smartphone variables during the teaching unit and were generally perceived to have pre-existing knowledge and thus possibly contextual expectations on this topic.

Figure 1 shows the TinkerPlots display of stacked dot plots for the distributions for 3rd graders (bottom) and 4th graders (top). The data reveal a large overlap between the two groups so that the analysis process requires sophisticated reasoning by comparing different measures and dealing with controversies in the data: While the maximum in the group of 4th graders is higher (36 games; 30 games for graders), median (4th graders: 4, 3rd graders: 8) and mean (4th graders: 5.4; 3rd graders: 9.2) are in favor of 3rd graders having more games on their phones.

The interview consisted of four phases: (1) Before looking at the data set, participants were asked about their initial conjectures and what they expect as an answer to the investigative question. (2) We picked up the idea of Biehler (2001) and asked the participants what they expected the distributions to look like. For this, the children placed 21 red and 21 blue cutout dots on a given diagram background, mimicking the TinkerPlots appearance (cf. Figure 4). Each dot represented one fictive student and their number of games on the phone. (3) Next, the pairs explored the investigative question using the given data set with TinkerPlots. Here, they were encouraged by the interviewer to apply and discuss different measures, including hills and the median. (4) Lastly, the students were asked to formulate a final conclusion for answering the investigative question.

During the interviews (duration between 30 and 40 min), all on-screen actions, as well as students' gestures and communication, were recorded and all written notes were collected.

Central constructs and data analysis

As mentioned above we define contextual knowledge as knowledge and beliefs about the students' daily life and experiences about the media use habits, especially about the number of games on smartphones/tablets. An example of drawing on contextual knowledge is the contextual expectation "they [4th graders] are allowed to play a bit more because they are older" (Case of Johannes).

To investigate these contextual expectations, this study uses the methodological proxy of *conjectures*. By this term, we refer to an inference about the context—uttered during the interview—before the beginning of the data analysis. Specifically, there were two stages during the interview to elicit conjectures: First, in phase 1,

Fig. 2 Initial conjecture by Nils

IC: 4th graders have more games on the smartphone because they are a few years older (case of Nils)

students were asked immediately after being presented with the statistical question which outcome they expected. We refer to the answer and the expressed reasoning behind it as *initial conjectures (IC)*. Figure 2 displays an example.

Second, in phase 2, students sketched distributions for both 4th and 3rd graders with cut-out dots, to which we refer with the term *conjectured distribution* (cf. Figure 4). These *conjectured distributions* served as a means to investigate a more statistical representation of the students' expectations. While the *initial conjecture* can be quite elusive and vague, the placement of the cut-out dots requires more commitment and makes certain decisions not only more traceable, but also more negotiable between the students. *Initial conjectures* as well as *conjectured distributions* can be expressed or backed up with varying degrees of confidence.

Conclusions drawn during the data exploration will be labeled as *data-centered insights (DCI)*, which can include descriptions, explanations, or inferences regarding the given data set. They are mostly based on different manipulations of the data set in TinkerPlots such as identifying the maxima or examining center and spread by drawing hills (cf. Figure 3).

To make the delicate reasoning processes and ideas visible, the study draws on interpretative methods (Jungwirth, 2005). The interviews were fully transcribed, discussed, and divided into episodes (Voigt, 1984). In the next step, episodes were analyzed for each pair, using a turn-by-turn approach, and summarized and analyzed along with the categories *initial conjecture*, *conjectured distributions*, and *data-centered insights*. To evaluate the students' confidence in their *initial conjectures*, *conjectured distributions*, and *data-centered insights*, specific attention was given to the qualifying, stochastic

Fig. 3 Data-centered insight 1 by Johannes including the statistical concepts backing up the DCI

DCI-1: 4th graders have more games on the smartphone. (Case of Johannes)

Backed-up by: spread, maxima

Fig. 4 Conjectured distributions of "number of games on the smartphone"; scale from 0 to 36, the upper half is 4th graders

terms students used when describing them (such as “maybe”, “surely” and “it is possible”). For *data-centered insights*, we identified the statistical concepts and/or elements of the data set (e.g., maximum/minimum, hills, number of dots, median, etc.) that were used to back them up.

Next, applying the method of constant comparison (see Jungwirth, 2005) by comparing each interpretation with previous findings, analyses aimed at suggesting the inner logic in the children’s utterances and then summarizing the emerging findings on the interplay of *conjectures* and *data-centered insights*. Lastly, all interpretations were consensually validated with the research groups of the authors.

Findings

In this section, we first discuss and compare findings concerning the first research question on students’ confidence and uncertainty regarding the expression of their initial conjecture and its relation to the conjectured distribution. Afterward and building upon that, we present the data analysis processes by the three pairs of students and investigate the interplay between their *initial conjectures* and *data-centered insights*.

Students’ initial conjectures and conjectured distributions

The selected three pairs of students verbalized the *initial conjecture* that 4th graders tend to have more games on their smartphones immediately after learning about the investigative question. Their reasoning behind this *initial conjecture* differed a little bit and was based on their contextual knowledge:

Nils and Johannes

Nils: Because they are [...] a few years older than 3rd graders. And that’s why they have more games.

Johannes: And they are allowed to play a bit more because they are older.

Linda and Paula

Linda: Fourth/ Both somehow.

Paula: Yes, 4th graders [...] Well, 4th graders probably have had a smartphone for a longer time than 3rd graders. They got it at the same age, maybe, but 4th graders were a bit older then. And yes, therefore they probably have more games.

Fiona and Sandra

Fiona: Last time, the others had more. So, 4th graders. [...] because maybe they are older or (*incomprehensible*).

Sandra: They are older. They are allowed to do more things. A bit. Such as stay up longer.

All pairs first referenced age as the reason for differences between the two groups. Additionally, they came up with further explanations drawn from their everyday knowledge ('when you are older, parents allow you more things'; 'the number of games increases with the time that you have it'). Looking at how they express their *initial conjecture* and the reasoning behind it, it is noticeable how the confidence of the pairs seemed to differ: While Nils, Johannes, and Sandra made firm statements (as can be found in Ben-Zvi et al., 2012), the other girls used terms such as "somehow", "maybe" and "probably", which can be considered a probabilistic language indicating their uncertainty (Ben-Zvi et al., 2012). It remains unclear, however, if they were unsure about the validity of the *initial conjecture* itself or about the reasoning they drew on to back it up.

After stating the *initial conjecture*, students were asked to come up with a *conjectured distribution* of 21 fictional students for each grade. While the three pairs of students constructed the distributions following their *initial conjecture* in favor of 4th graders (upper half in the representation, Fig. 4), their appearances differed.

Overall, **Nils and Johannes'** *conjectured distribution* has a strong tendency towards 4th graders having more games on their smartphones (Fig. 4, left). Visually, this becomes especially apparent in the section for a high number of games: They placed 6 dots between 30 and 36 games for 4th graders and only one in this range for 3rd graders. This is in line with their strongly stated *initial conjecture*. **Paula and Linda's** *conjectured distribution* shows a rather small difference between the two groups, which seems similar to their probabilistic language and apparent little confidence in their *initial conjecture*.

The largest difference between their *conjectured distributions* was the maximum number of games per distribution: the girls placed two dots with 22 games in 4th grade and one dot with 20 games in 3rd grade. The girls' *conjectured distributions* only differ concerning the center measures; from a spread perspective, they would expect similar distributions for 3rd and 4th graders.

Fiona and Sandra (Fig. 4, right) showed a strong shift to the right (i.e., more games) for the distribution of the 4th graders so that a third of the 3rd graders' dots were equal to or lower than the minimum of the 4th graders. Interestingly, their *conjectured distributions* are similar concerning shape and spread but the center of the 4th graders' distribution is shifted to the right. From the *conjectured distributions*, the difference between the two groups is expressed more confidently than especially Fiona's verbalization of the *initial conjecture*. The reason might be that Sandra was already more convinced of the validity of the *initial conjecture* and influenced the appearance of the *conjectured distributions* more.

In summary, we could not only observe a reflection of the statement of the *initial conjecture* in the *conjectured distribution* but also of the level of confidence (cf. Ben-Zvi et al., 2012) with which they were expressed seemed to be mirrored in the *conjectured distributions*. Assuming this as a basis for the data analysis process, Nils and Johannes had a strong conjecture of the advantage of 4th graders, Fiona and Sarah seemed quite convinced of it and Paula and Linda seemed rather unsure.

Students' analysis processes and data-centered insights.

We will now have a closer look at the students exploring the data in TinkerPlots. The results from the analysis of the cases of Nils and Johannes, Paula and Linda, and Fiona and Sandra are organized along with crucial episodes we identified in our analysis process, which are presented in chronological order so that the development of the *data-centered insights* and their relation to the *initial conjectures* can be fully traced. The *data-centered insights* (DCIs), i.e., conclusions regarding the investigative question that are made during or after working with the data set, are contrasted with the conjectures (i.e., *initial conjecture* and *conjectured distribution*). After the analysis of each pair and with the focus on the research question we will summarize how each pair dealt with survey data that is opposed to their initial conjectures and will present a schematic representation of how the students' *data-centered insights* related to the *initial conjecture* and how the *data-centered insights* are backed-up.

The case of Nils and Johannes

Episode 1: First analysis of the data

When presented with the large survey data set, Nils and Johannes had no trouble navigating the software TinkerPlots and created a stacked dot plot for 3rd graders and 4th graders (the same as in Fig. 1 without the measures). The research assistant (RA) then initiated an interpretation of the results.

RA: Can you also say something in regard to the investigative question? 3rd graders or 4th graders – who has more games?

Nils: Fourth graders.

Johannes: Fourth graders. Oh no, I would say that the 3rd graders and the 4th graders, too. For the 4th graders, it is more distributed than for the 3rd graders. 3rd graders have up to 30 games and 4th graders have up to 36 games.

Nils: Yes, I would say 4th graders.

Johannes: Yes, me too.

Nils and Johannes both at first continued in line with their conjecture that 4th graders have more games on their smartphones compared to 3rd graders (*DCI-1*). Johannes then offered in line 9 a more elaborate analysis by referring to the spread (“more distributed”) of 4th graders and the maxima of both groups. With “Oh no, I would say that the 3rd graders and the 4th graders, too” (*DCI-2*), it remained unclear if he viewed the groups to have the same or similar amounts of games or if he tried to express the overlap of the data.

As he starts with “oh no”, he might have perceived a discrepancy between the conjecture, the *DCI-1*, and this new finding (*DCI-2*). Since Nils began line 10 with “yes”, he might have understood Johannes’s statement in line 9 as in favor of 4th graders. Thus, he perceived *DCI-1* as supported by data and specifically by the spread and the maxima.

Episode 2: Comparing the hills

In the next episode, the boys were asked to draw the hills in both distributions (Fig. 5).

RA: Okay, with regard to your hills, what have you found out in regard to your investigative question? Who has more games on the smartphone?

Johannes: Fourth graders, 100%, Fourth graders.

RA: How can you tell this from the hill?

Nils: Because they have more. And also more here, up to 36.

Johannes: Yes.

RA: How can you tell this from the hill?

Johannes: Because the hill is located more to the right. Ends more to the right.

Asked to draw a conclusion from the hills, Johannes very confidently interpreted them in favor of the 4th graders (line 13) and thus following the boys’ *initial conjecture* and *DCI-1*. Nils supported this by referring to a larger range or the higher maximum (“up to 36”) in the distribution of 4th graders (line 15). Johannes then added a correct interpretation of the 4th graders’ hill ending farther to the right in support of *DCI-1* (line 18).

From the looks of the hills, one could suggest that they were possibly already drawn with the intention to support the *initial conjecture*. Johannes first drew the hill of the 4th graders and immediately afterward without commentary the one of the 3rd graders with much sharper edges than the previous one. Contrasting the *initial conjecture* and the *data-centered insights* we can say that Nils and Johannes gain more confidence in their *initial conjecture* after comparing the hills as an intermediate stage by additionally backing the *DCI-1* up with the arguments of range and shift of the hills – both arguments which are related to a local view (range) and an intermediate view (hills) are interpreted by Nils and Johannes in favor for the 4th graders.

Fig. 5 Group comparison display of Nils and Johannes with drawn hills in TinkerPlots

Episode 3: Comparing the medians

With some technical support from the research assistant, the boys display the median in both distributions in TinkerPlots (the same as in Fig. 1). Johannes then justifies why the median for the 4th graders is so far to the left compared to the overall diagram.

Johannes: But here (*points in the 4th graders' diagram to the dots left of the median*), it is much smaller, that is why he made it this far (*points to median*). Because here (*points to the right side of the 4th graders' diagram, moves his hand left to right*) is also a large number of dots distributed. [...].

Johannes: The median for 4th graders is halfway to the left of the median for 3rd graders. 3rd graders are twice as far to the right. [...].

RA: What did you find out for your investigative question? 3rd or 4th graders? Who has more games on their smartphones?

Johannes: (*points at the computer*) For 100%. The 4th graders.

RA: And if you focus on the median? On both medians?

Johannes: That the 4th graders have more games in one pile than the 3rd graders.

RA: What do you mean by this?

Johannes: That the 4th graders have more games here (*points to the drawn-in hill of 4th graders, marks its boundaries left and right with his fingers*) and that they (*points to the diagram of 3rd graders*) have distributed them everywhere (*skims with his finger over the bottom diagram from the median on to the right*). That the 4th graders have squeezed together a whole lot of games at once.

Fig. 6 Initial conjecture and data-centered insight 1 and 2 by Nils and Johannes after comparing the medians

Johannes began the interpretation with a more technical description of the median in line 19 and the correct comparison of the medians in line 20. From this, however, he did not draw a conclusion in favor of 3rd graders although his sophisticated multiplicative comparison (cf. Frischmeier, 2017; “halfway”, “twice as far”) of both medians in line 20 reveals a big difference. Instead, he again confidently claimed that 4th graders tend to have more games, supporting the *initial conjecture* and *DCI-1* (line 22). His reasoning in line 24 seemed to relate more to the hills (“more games in one pile”) which is supported by his gestures in line 26. Therefore, while he began with the analysis of the medians, neither of the boys seemed to take them into account when arriving at the conclusion. Despite that, they appeared to interpret the medians in favor of *DCI-1* and gaining even more confidence.

Summarizing these episodes, Nils and Johannes had a strong *initial conjecture* and a *conjectured distribution* clearly in favor of 4th graders. Even though the real data suggests a different conclusion and the boys were able to make some in-depth analyses (spread, center maximum, median, etc.) and correct statements on the measures (multiplicative comparison of the medians), they tailored all *data-centered insights* to support the *initial conjecture*. *DCI-2*, mentioned in the beginning by Johannes, offered a pathway for a more comprehensive analysis and conclusion. However, it was immediately disregarded in favor of *DCI-1*, which was not challenged afterward but backed with different arguments (spread, maxima, hills, and medians) (see Fig. 6).

The case of Paula and Linda

Episode 1: First analysis of the data

Paula and Linda also started with the *initial conjecture* that 4th graders tend to have more games on their smartphone than the 3rd graders. The way they stated it and their *conjectured distribution* suggested that they were not very confident in its validity. When the girls started exploring the real data set in TinkerPlots and displayed the stacked dot plot, they immediately commented on the conflicting look of the two overlapping groups.

Paula: So, here, it also looks like the 4th graders had more. But some how also as if the 3rd graders had more.

Linda: (*incomprehensible*) (*pointing at the left half of the 4th graders' distribution*) 4th graders have fewer games than them. But back here (*hovering her hand back and forth over the right half of the diagram*) it changes, so that 4th graders have more. Here you think if you look at it first/ (*makes a repeated down and up movement with her hand as if cutting the distributions off between 15 and 19 games, beginning on the 4th graders*).

Paula: (*interrupting, incomprehensible*) So, 4th graders are all the way through to the end, and 3rd graders go until 30 roughly.

Linda: From here on (*makes the same up-and-down movement as before, beginning at the 3rd graders' distribution*) you first think 3rd graders have more. But then when you look at this row (*points at the end of the scale and at 18 games*) it really looks like it.

Paula described a conflict that both 3rd graders and 4th graders could be seen to have more games on their smartphones, which can be understood as two *data-centered insights*: (*DCI-1* in favor of 4th graders, and *DCI-2* in favor of 3rd graders). These were supported by Linda's more in-depth analysis: With her hand gesture, she 'cut' the distributions in two different sections (similar in Frischemeier, 2017; Hammerman & Rubin, 2004) and drew conclusions in terms of the statistical question for each section. As stated in line 30, the right part (from the gap between 14 and 18 to the maximum of 36 games) was interpreted in favor of the 4th graders and the left part in favor of the 3rd graders. Looking at the data, both assumptions are correct (concerning arithmetic mean and median) but there are 35 children from grade 3 and only 14 from grade 4 in the right group (17 games and more).

Most importantly, the left group contains the majority of each data set with over 200 dots in each distribution. The number of dots per section, however, was not addressed by both students over the course of the interview. In her comparison, Linda seemed to give both parts the same weight. While she did not express a final conclusion, her last statement in line 30 ("it really looks like it") could be a confirmation that 4th graders have more games on their phones. Thus, there seemed to be a slight preference for *DCI-1* (which would be in line with the *initial conjecture* and the *conjectured distribution* in Fig. 4). In line 29, Paula offered an additional argument: by "to the end" and "until 30", she might have referred to a comparison of the maximum of 36 games for 4th graders and 30 games for 3rd graders or to a comparison of the ranges of both distributions which both rather tend to show a local view on distributions. This would also be in favor of 4th graders having more games on their phones and thus in line with the girls' *IC* as well as *DCI-1*.

Episode 2: Comparing the hills

In the next episode, the children were asked to draw hills (modal clumps) as they had done in the classroom activity. While Paula and Linda drew the hills immediately (see Fig. 7), they did not explain why they chose their specific shapes and

widths. In the following, the girls were asked to interpret the hills in terms of the investigative question.

RA: Okay. Good. Now, what did you find out in regards to your question when you look at the hills. 3rd graders or 4th graders – who has more games on the smartphone? [...].

Linda: 4th graders, definitely.

Paula: Because 4th graders moved farther to the right with their dots and if you take the exact center, now, only of the dots, [...] the median of the 4th graders would be farther in the middle than of the 3rd graders.

Linda: Yes, definitely. But really, if you take just half now (*makes up and down gesture as before in the middle of the diagram*), then it really looks as if 3rd graders furth/ as if they had more.

Paula: Well, if you look just at the hills, then 3rd graders would have more. But if you look at the whole, then it's 4th graders.

In this episode, Linda and Paula seemed to gain more confidence in *DCI-1* by supporting it with the median (though it was not displayed at this point) (line 33). While it was unclear at this point, later statements confirmed that both Paula and Linda used an informal conception of the median, which was not in line with the standardized version (cf. Schnell & Frischmeier, 2020): they believed the median to be “the exact center of the dots”, meaning the midrange. From this “median as midrange” perspective, their median argument was correctly interpreted in favor of the 4th graders and thus supporting *DCI-1*. When Linda repeated her strategy of ‘cutting’ the distribution into halves and evaluating them separately (line 34), Paula

Fig. 7 Group comparison display of Paula and Linda with drawn hills in both distributions

followed her idea and interpreted the hills (drawn-in by the girls) as in favor of the 3rd graders (*DCI-2*) and “the whole” (line 35) in favor of 4th graders (*DCI-1*). Here, “the whole” could either refer to each of the two distributions of all dots or to the whole range (minimum to maximum of 36 games) which would relate to Paula’s previously stated argument in line 29.

Overall, in this episode, the girls found again arguments both for their *DCI-1* (median, “the whole”) as well as for *DCI-2* (hills). Again, *DCI-1* seemed slightly more supported, as there was one measure more backing it up.

Episode 3: Comparing the medians

The research assistant then asked the students to display the medians in TinkerPlots, which they did correctly without any hesitation. Immediately after the medians show up, Paula reacted spontaneously with an interpretation:

Paula: They just took it of the hills. In the hills, 3rd graders have more. [...] So the median of the hills.

RA: The median of the hills? What is that/

Linda: They didn’t use the one of the whole. They just drew the hills. [...].

Paula: Yes, but otherwise, the [median] of the 4th graders would be much further in the middle and the one of the 3rd graders, too.

RA: So, when you only look at the median, who has more games on the smartphone?

Paula: Yes, well, in this case, where they used only the hills, it’s the 3rd graders. But if you did it without the hills, it would be the 4th graders.

In line 36, Paula probably referred to TinkerPlots by “they”, stating that the way the program calculated the median differs from the way they believe to be correct (determining the midrange instead of the median). Thus, the formal TinkerPlots median showing a clear indication that 3rd graders tend to have more games on their smartphones than 4th graders differed from their expected median. This could be an indication that the students not only had a contextual expectation, but also a statistical conjecture model of the median (cf. Dvir & Ben-Zvi, 2018). To account for this, they argued that the program displayed only the “median of the hills” (i.e., midrange of the ‘edges’ of the hills). The girls concluded that the TinkerPlots median was in favor of 3rd graders (*DCI-2*) while their median would be in favor of 4th graders (*DCI-1*) and thus in line with the *initial conjecture*.

Until the end of the interview, the students always offered arguments for either group having more games on the phone, depending on which measure was used (Fig. 8). This is also in line with the very first statement by Linda in line 3 when she stated her *initial conjecture* as “both”. Even though the girls settled on the *initial conjecture* “4th graders have more games on their smartphone” and the *conjectured distribution* was slightly more supporting this (cf. Figure 4), the uncertainty and the ambiguity seemed to stay persistent over the course of the interview. However, it seems that *DCI-1* is slightly more what the girls expected: When the TinkerPlots median is in favor of *DCI-2*, they neither challenge the program’s calculation nor revise their confidence in *DCI-1*, but rather agree with a co-existence of two types of

IC: 4 th graders have more games on the smartphone.	DCI-2: 3 rd graders have more games on the smartphone. <i>Backed-up by: left part of distribution, hills; TinkerPlots median</i>
	DCI-1: 4 th graders have more games on the smartphone. <i>Backed up by: right part of distribution, range, (midrange), median</i>

Fig. 8 *Initial conjecture and data-centered insight 1 and 2 by Paula and Linda after comparing the medians*

medians, which back-up the different *DCIs*. So overall, even though Paula and Linda came from a different starting point than Nils and Johannes with less confidence in the *initial conjecture* and a more sophisticated analysis integrating statistical and contextual information in which they constructed arguments for two different *DCIs*, they also found ways of interpreting the data differently to fit their expectation.

The case of Fiona and Sandra

While the girls shared the same *initial conjecture* as to the other pairs (4th graders tend to have more games on their smartphones), Sandra seemed at first more confident than Fiona. Their *conjectured distribution* (Fig. 4), however, was clearly in favor of 4th graders having more games on their smartphone and thus in line with their *initial conjecture*.

Episode 1: First analysis of the data

Sandra and Fiona created the stacked dot plot diagrams to compare the two groups in TinkerPlots (similar to Fig. 1 without the measures). Both had difficulties interpreting the diagram and extracting findings from the data. Neither Sandra nor Fiona offered an initial interpretation of the graphs (answering the question “can you see something in particular” with “no”), so that the research assistant guided them directly to investigate the medians. So after the first analysis of the data, Fiona and Sandra only offer an *initial conjecture* but no *data-centered insight*.

Episode 2: Comparing the medians

The girls then displayed the median in TinkerPlots (cf. Figure 1).

RA: Okay. Then describe and compare both medians. [...].

Sandra: 3rd graders [...], 3rd graders have more games.

Fiona: No, 4th.

Sandra: No, [...] look at the median; 4 and 8 (*points at both medians*).

Fiona: Right, then 3rd graders have more.

Fig. 9 Group comparison display of Sandra and Fiona with drawn hills in both distributions

Fiona and Sandra struggled at first with the interpretation of the median but finally settled correctly on the median of 3rd graders being higher and thus in favor of them having more games. As both children agree on this interpretation, an alternative *data-centered insight* was not constructed, even though *DCI-1* was not in line with the *initial conjecture*.

Episode 3: Comparing the hills

Next, Sandra and Fiona also used hills (Fig. 9) to compare the two distributions.

RA: Okay, when you look at the two hills, who has more games? [...].

Sandra: 3rd.

Fiona: 3rd, yes, 3rd.

RA: How do you see that? [...].

Fiona: Well, this is longer here (*moves finger quickly back and forth in horizontal line over the screen*), but there is more (*points to 3rd graders' distribution*).

Fiona's statement "this is longer" might have referred to the range of the 4th graders' distribution, even though her gesture is not clearly identifiable. "There is more" could mean the width of the hill of 3rd graders, but the exact interpretation remains unclear. However, Fiona seemed to understand the hills as supporting the *DCI-1*, i.e., in favor of 3rd graders (see Fig. 10).

Asked for a final conclusion by the research assistant at the end of the interview, both girls answered the investigative question with "3rd graders". Their reasoning behind this did not refer to the context, but rather the outcomes of the steps of the statistical analysis:

RA: Okay. Do you have an explanation why 3rd graders could have more?

Fig. 10 Fiona and Sandra's *initial conjecture* and *data-centered insight* after comparing the hills

Sandra: Somehow, I think, we always had 3rd graders. What you showed us, 3rd graders, 3rd graders. We noticed it everywhere. Thus 3rd graders.

Summarizing, Sandra, and Fiona at first had the *initial conjecture* that 4th graders tend to have more games on their smartphones. Even though they backed-up this conjecture with contextual knowledge ('older kids are allowed more') and constructed a *conjectured distribution* clearly supporting this, they never drew on it in during the data exploration process and did not integrate statistical and contextual information which is a crucial part of statistical thinking (Pfannkuch & Wild, 2004). Compared to the other two pairs, they were the only ones who did not verbalize a *data-centered insight* in line with the *initial conjecture* but used the median and hills only to back-up the *DCI* in favor of 3rd graders.

Summary

Overall, these three cases offered insight into how complex the question "who has more games on the smartphone?" is when it comes to comparing two distributions. This paper focused on interviews with three pairs of 3rd graders whose *initial conjecture* contradicted most of the outcomes of the data exploration, their *data-centered insights*.

In our case study, we investigated how our young students with conjectures of varying degrees of confidence develop *data-centered insights* and how they relate their *data-centered insights* to their *initial conjectures*. For the three pairs presented here, we observe that the children seemed to (intuitively) mirror their contextual expectations and their daily life knowledge not only in terms of the overall tendency (i.e., here in favor of 4th graders having more games on their smartphone) but also apparently in terms of the level of confidence they have in the conjectures (cf. Ben-Zvi et al., 2012).

While all three pairs had the same *initial conjecture*, they differed in the degree of confidence in its validity and the ways they deal with data that is opposed to their *initial conjectures*: as suggested by the framework of Chinn and Brewer (1993), Nils and Johannes, who have had the strongest conjecture at the beginning, represent the one extreme who do not ignore the anomalous data but interpret it so that all findings support their expectation in the sense of a confirmation bias (Jermias, 2001). This finding is also in line with Masnick et al. (2007) who illustrated childrens'

resistance to revise existing beliefs when experimental data did not match their expectations.

Fiona and Sandra also showed a quite strong conjecture in favor of the 4th graders but dealt with the data completely differently: rather than ignoring or re-interpreting the data, they completely ignored their *initial conjecture*. While the girls' conclusions are certainly not a whole theory in the sense of Chinn and Brewer (1993), their will to abandon pre-existing considerations completely relates to the other end of their framework.

The reactions of both pairs are not in line with the experimental findings by Legare and colleagues, who emphasized the role of information inconsistent with expectations as a motivation for deepening the investigations.

Paula and Linda can be perceived as in the middle as they interpret some measures in favor of one result and others in favor of the other. Even though their understanding of the median is not yet in line with formal definitions (cf. Schnell & Frischemeier, 2020), it shows how they have more sophisticated reasoning about the outcome of statistics by tolerating uncertainty (cf. Ben-Zvi et al., 2012; Legare et al., 2016) and integrating statistical and contextual information (cf. Pfannkuch & Wild, 2004).

Conclusions and implications

The presented paper investigated two related questions: (1) How – in terms of confidence and uncertainty—do students express their initial, contextual conjecture in regard to the outcome of a statistical investigation of survey data, and how is this related to their conjectured distribution of data? (2) Assuming the initial conjecture serves as a starting point for data investigations, how do young learners deal with findings from data explorations in survey data that are not in line with their initial conjectures?

In regard to the first research question, we found that some students expressed their conjectures with respect to their uncertainty by using probabilistic language while others formulated very firm deterministic conjectures (similar to Ben-Zvi et al., 2012). The degree of confidence was also mirrored in the conjectured distributions. Therefore, we assume that the initial conjecture served as a statistical conjecture model in the sense of Dvir and Ben-Zvi (2021) and influenced the data analysis process.

Regarding the second research question, we found three different approaches in the case analysis: tailoring data-centered insights to fit the initial conjecture, integrating initial conjecture and data-centered insights, and ignoring the initial conjecture by relying fully on the data-centered insights. In line with other studies and settings of exploring physical or psychological experiments (Busch & Legare, 2019; Legare et al., 2016; Masnick et al., 2007; Pfannkuch & Rubick, 2002), contextual expectations seem to play an important role in data explorations in survey data as in this case study.

While some of these studies report a positive effect of data diverging from expectations on the investigation processes (e.g., Busch & Legare, 2019; Legare et al., 2016), others emphasize the problems especially young students have with resolving

this situation (Masnick et al., 2007). The presented study shows that for a setting of investigating real, social data statistically, young students have very different approaches to deal with a perceived difference between their contextually rooted initial conjectures and the outcome of data analysis. While Dvir and Ben-Zvi (2018) argue that statistical conjectures and data models develop at the same time, our results show that a contextual expectation might not develop at all but stays persistent throughout the data exploration process. Our empirical findings show how the *initial conjecture* can not only influence how students expect data to look (similar as in Dvir & Ben-Zvi, 2021) but also the way they interpret outcomes from data analysis (similar as in Jermias, 2001). As has been shown in mathematics education (Smith et al., 1993), individual pre-existing conjectures seem to be able to overpower the results of the statistical analysis, especially when students have only begun to deal with statistics and have not yet gained trust in its results and/or their own statistical abilities.

Implications for practice and future research

The presented study highlights that learning to trust data and data analysis more than the own experience (Wild & Pfannkuch, 1999) is challenging and cannot be resolved only by exploring data. Rather, interpreting and drawing connections between statistical and contextual information have to be carefully scaffolded.

A starting point for an adequately designed learning pathway might be to provide students with the means to distance themselves temporarily from their own expectations and experiences to interpret results from data analysis more objectively. The subsequent comparison and in-depth discussion of both – pre-existing conjectures and statistical measures – is then a cognitively demanding but crucial step in acquiring statistical reasoning skills. However, the presented case of Paula and Linda indicates that even these young learners are capable of such a sophisticated discussion.

Overall, the study of three cases can only be the beginning of exploring the field. We illustrated different ways of dealing with insights from data analysis even though all students started with the same conjecture. Even though we focused on only three pairs of students, we believe that valuable insights can be gained from the in-depth analysis of these primary school students' reasoning processes. However, studies with more participants are needed to investigate which other ways exist and what is common for students of this young age.

Acknowledgements We are very grateful to Katie Makar, Kevin Larkin and the anonymous reviewers for very helpful and constructive comments and feedback on previous versions on this manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Research involving human and animal participants This study involved research with human participants (primary school students). The participants along with their legal guardians provided informed consent to participate in the study. All guidelines set forth in the informed consent agreement have been followed.

Ethical approval This study involved research with human participants (primary school students). The participants along with their legal guardians provided informed consent to participate in the study. Ethical approval for this study was not required in accordance with the conditions outlined by the German Research Foundation, where research that carries no additional risk beyond daily activities does not require Research Ethics Board Approval.

Consent to participate All participants and parents of all minor participants gave their informed consent.

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147–167). Kluwer Academic Publisher.
- Ben-Zvi, D. (2018). Foreword. In A. Leavy, M. Meletiou-Mavrotheris, & E. Paparistodemou (Eds.), *Statistics in Early Childhood and Primary Education* (pp. vii–viii). Springer.
- Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM - the International Journal on Mathematics Education*, 44(7), 913–925.
- Ben-Zvi, D. & Garfield, J. (Eds.). (2004) *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*. Kluwer Academics.
- Biehler, R. (2001). Statistische Kompetenz von Schülerinnen und Schülern - Konzepte und Ergebnisse empirischer Studien am Beispiel des Vergleichens empirischer Verteilungen. In M. Borovcnik, J. Engel, & D. Wickmann (Eds.), *Anregungen zum Stochastikunterricht* (pp. 97–114). Franz Becker.
- Burrill, G., & Biehler, R. (2011). Fundamental Statistical Ideas in the School Curriculum and in Training Teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching Statistics in School Mathematics – Challenges for Teaching and Teacher Education* (pp. 57–69). Springer.
- Busch, J. T. A., & Legare, C. H. (2019). Using data to solve problems: children reason flexibly in response to different kinds of evidence. *Journal of Experimental Child Psychology*, 183, 172–188.
- 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.
- Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, 104(9), 801–823.
- Dvir, M., & Ben-Zvi, D. (2021). The double-edged sword of conjecturing. *Mathematical Thinking and Learning*. <https://doi.org/10.1080/10986065.2021.1940427>
- Dvir, M., & Ben-Zvi, D. (2018). The role of model comparison in young learners' reasoning with statistical models and modeling. *ZDM Mathematics Education*, 50, 1183–1196.
- Engel, J. (2017). Statistical literacy for active citizenship: a call for data science education. *Statistics Education Research Journal*, 16(1), 44–49.
- Engels, A. (2018). *Design, Durchführung und Auswertung einer Umfrage zu Freizeitgewohnheiten von Grundschulern sowie Design einer möglichen Unterrichtssequenz zum Einsatz der erhobenen Daten im Mathematikunterricht der 4. Klasse*. Bachelor Thesis, University of Paderborn.

- Fielding-Wells, J. (2018). Dot plots and hat plots: supporting young students emerging understandings of distribution, center and variability through modeling. *ZDM Mathematics Education*, 50(7), 1125–1138.
- Frischemeier, D. (2017). *Statistisch denken und forschen lernen mit der Software TinkerPlots*. Springer Spektrum.
- Frischemeier, D. (2019). Primary school students' reasoning when comparing groups using modal clumps, medians and hatplots. *Mathematics Education Research Journal*, 31, 485–505.
- Garfield, J., & Gal, I. (1999). Teaching and assessing statistical reasoning. In L. Stiff (Ed.), *Developing mathematical reasoning in grades K-12: National Council Teachers of Mathematics 1999 Yearbook* (pp. 207–219). NCTM.
- Garfield, J., & Ben-Zvi, D. (2009). Helping Students Develop Statistical Reasoning: implementing a Statistical Reasoning Learning Environment. *Teaching Statistics*, 31(3), 72–77.
- Gattuso, L., & Ottavani, M.G (2011). Complementing mathematical thinking and statistical thinking in school mathematics. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education* (pp. 121–133). Dordrecht: Springer.
- Groth, R. E. (2007). Toward a conceptualization of statistical knowledge for teaching. *Journal for Research in Mathematics Education*, 38(5), 427–437.
- Hammerman, J. K. L., & Rubin, A. (2004). Strategies for managing statistical complexity with new software tools. *Statistical Education Research Journal*, 3(2), 17–41.
- Jermias, J. (2001). Cognitive dissonance and resistance to change: the influence of commitment confirmation and feedback on judgement usefulness of accounting systems. *Accounting, Organizations and Society*, 26(2), 141–160.
- Jungwirth, H. (2005). *Interpretative Mathematikdidaktik: Methodisches Und Methodologisches Am Beispiel Von Normen Im Mathematikunterricht*. <http://psydok.sulb.uni-saarland.de/volltexte/2005/449/pdf/jungwirth.pdf>
- Konold, C., & Miller, C. (2011). *TinkerPlots 2.0*. Key Curriculum Press.
- Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R., & Mayr, S. (2002). Students' use of modal clumps to summarize data. Paper presented at the *Sixth International Conference on Teaching Statistics*. Cape Town, South Africa.
- Langrall, C., Nisbet, S., Mooney, E., & Jansem, S. (2011). The role of context expertise when comparing data. *Mathematical Thinking and Learning*, 13, 47–67.
- Leavy, A., & Hourigan, M. (2018). Inscriptional Capacities and Representations of Young Children Engaged in Data Collection During a Statistical Investigation. In A. Leavy, M. Meletiou-Mavrotheris, & E. Paparistodemou (Eds.), *Statistics in Early Childhood and Primary Education* (pp. 89–108). Springer.
- Leavy, A., Meletiou-Mavrotheris, M., & Paparistodemou, E. (2018). *Statistics in early childhood and primary education: Supporting early statistical and probabilistic thinking*. Springer.
- Legare, C. H., Schult, C. A., Impola, M., & Souza, A. L. (2016). Young children revise explanations in response to new evidence. *Cognitive Development*, 39, 45–56.
- Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, 8(1), 82–105.
- Masnack, A. M., Klahr, D., & Morris, B. J. (2007). Separating signal from noise: Children's understanding of error and variability in experimental outcomes. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 3–26). Lawrence Erlbaum.
- Pfannkuch, M., & Rubick, A. (2002). An exploration of students' statistical thinking with given data. *Statistics Education Research Journal*, 1(2), 4–21.
- Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 17–46). Kluwer.
- Schnell, S. & Frischemeier, D. (2020). Primary school students reasoning about and with the median when comparing distributions. In U. T. Jankvist, M. Van den Heuvel-Panhuizen, & M. Veldhuis (Eds.), *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education (CERME11, February 6 – 10, 2019)* (pp. 1075–1082). Freudenthal Group & Freudenthal Institute, Utrecht University and ERME.
- Smith, J. P., diSessa, A. A., & Rochelle, J. (1993). Misconceptions Reconceived – A Constructivist Analysis of Knowledge in Transition. *Journal of Learning Sciences*, 3(2), 115–163.
- Voigt, J. (1984). *Interaktionsmuster Und Routinen Im Mathematikunterricht: Theoretische Grundlagen und Mikroethnographische Falluntersuchungen*. Beltz.

- Watson, J. M., & English, L. D. (2015). Introducing the practice of statistics: are we environmentally friendly? *Mathematics Education Research Journal*, 27(4), 585–613.
- Watson, J. M., & Moritz, J. B. (1999). The beginning of statistical inference: comparing two data sets. *Educational Studies in Mathematics*, 37(2), 145–168.
- Weiland, T. (2019). The contextualized situations constructed for the use of statistics by school mathematics textbooks. *Statistics Education Research Journal*, 18(2), 18–38.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223–248. <https://doi.org/10.1111/j.1751-5823.1999.tb00442.x>
- Yin, R. (2014). *Case study research: Design and methods* (5th ed.). Sage.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.