Chapter 2
Students' Development of Measures

Christian Büscher

Abstract Knowledge is situated, and so are learning processes. Although contextual knowledge has always played an important role in statistics education research, there exists a need for a theoretical framework for describing students' development of statistical concepts. A conceptualization of measure is introduced that links concept development to the development of measures, which consists of the three mathematizing activities of structuring phenomena, formalizing communication, and creating evidence. In a qualitative study in the framework of topic-specific design research, learners' development of measures is reconstructed on a micro level. The analysis reveals impact of the context of a teaching-learning arrangement for students' situated concept development.

Keywords Concept development · Design research · Situativity of knowledge
Statistical measures · Statistical reasoning

2.1 Introduction: Concept Development as a Focus for Research

In recent years, the ability to draw Informal Statistical Inferences (ISI) (Makar and Rubin 2009) has become a focal point of statistics education research (see the ESM special issue on sampling, Ben-Zvi et al. 2015). ISI emphasizes the use of statistical concepts in drawing 'probabilistic generalizations from data' (Makar and Rubin 2009, p. 85) and in making claims about unknown phenomena. In order to describe the type of reasoning used in drawing ISI, Makar et al. (2011) propose a framework of Informal Inferential Reasoning (IIR) . This framework reveals the complexity of IIR; the components include knowledge of statistical concepts as well as contextual knowledge and general norms, habits, and patterns of action (Makar et al. 2011).

C. Büscher (✉)
TU Dortmund University, Dortmund, Germany
e-mail: christian.buescher@math.tu-dortmund.de

Learning to draw ISI is thus conceptualized by the development of IIR, placing strong emphasis on the *use* of statistical concepts in complex activity.

Learners however will first need to *develop* the statistical concepts to be used in IIR through activity in teaching-learning arrangements that are designed to facilitate such concept development. A key assumption of the research related to this study follows the ideas of Freudenthal (1991) that learners can develop formal, general concepts out of their informal, singular activity, when their learning processes are carefully guided. The framework of IIR gives only limited guidance for such a design, detailing the goal, but not the path, of concept development. A language is required that can be used to describe learners' situated understandings, the individual concepts guiding their actions, the relation of these concepts to formal statistical concepts, and the complex interplay with elements of the design of a teaching-learning arrangement.

Whereas similar studies have focused on the concepts of distribution (Lehrer and Schauble 2004) and shape (Gravemeijer 2007), this study focuses on the concept of measure. Since data-based evidence plays a major role in ISI and IIR and measures are a common form of such evidence, this focus could provide not only insights into concept development but also make connections to the development of IIR.

2.2 Theoretical Background

2.2.1 The Situativity of Knowledge

A long-standing perspective in cognitive psychology concerns the situativity of knowledge (Greeno 1998). A conceptualization of measure that takes learning processes into consideration needs to pay attention to the fact that knowledge emerges from situations. Vergnaud (1990, 1996) proposes a theory of conceptual fields as an epistemological framework. To Vergnaud, the perception of situations and the understanding of mathematical concepts stand in a dialectic relationship: "cognition is first of all conceptualization, and conceptualization is specific to the domain of phenomena" (Vergnaud 1996, p. 224). In this way, mathematical knowledge emerges through actions in situations. This knowledge is not to be understood as consisting of situation-independent abstractions but rather as an operational invariant across different situations.

The two most important types of operational invariants are concepts-in-action and theorems-in-action. Concepts-in-action are "categories (objects, properties, relationships, transformations, processes etc.) that enable the subject to cut the real world into distinct elements and aspects, and pick up the most adequate selection of information according to the situation and scheme involved" (Vergnaud 1996, p. 225). Thus, they organize what students focus on and in this case how they structure phenomena unknown to them. Theorems-in-action are defined as "propositions that [are] held to be true by the individual subject for a certain range of situation variables" (Vergnaud 1996, p. 225). They are intricately connected to the learners' concepts-in-action:

theorems-in-action give meaning to concepts-in-action, which in turn give content to the theorems-in-action.

A conceptualization of measure that takes into account the situativity of knowledge thus will need to provide a clear focus on the use of measures in situations. The median is a measure of center, but this does not explain its use in terms of operational invariants.

2.2.2 Functions of Measures

Although measures are a prominent concept in statistics and statistics education, few explicit definitions or conceptualizations exist that explain this construct. At least three different functions of measures can be identified based on literature: (a) structuring phenomena, (b) formalizing communication, and (c) creating evidence.

Structuring phenomena. Bakker and Gravemeijer (2004) distinguish between data (the individual values) and distribution (a conceptual entity). Two perspectives on data and distribution emerge: The 'upward perspective' consists in seeing data as a means to calculate measures (median, range, …) of aspects of a distribution (center, spread, …). The 'downward perspective' consists of looking at the data from the standpoint of distribution, with aspects of center and spread as organizing structures already in mind.

In this way, measures function as lenses that allow access to distributional properties. This resonates with the idea of an 'aggregate view' on data (Konold et al. 2015): perceiving data as a conceptual unit with its own emergent properties, which can be accessed through the use of measures. In data investigation, measures thus impose distributional properties on phenomena, creating structure in previously unstructured phenomena.

Formalizing communication. Structuring phenomena alone does not conclude statistical investigation; findings must also be communicated to a wider audience. Through their standardized procedures of calculation, measures can provide such a means of communication. They create intersubjectivity, allowing for communication about phenomena across distance and time (Porter 1995; Fischer 1988).

Creating evidence. One of the characterizing features of ISI given by Makar and Rubin (2009) is the use of data as evidence. Whereas they do not explicitly relate this role of evidence to measures, it is possible to think of the form of this evidence as consisting of measures. Abelson (1995) states that the discipline of statistics supports principled arguments that aim at changing beliefs and which therefore need to be convincing to others. Simple unspecified reference to data would not serve this goal of convincingness. Instead, specific aspects have to be 'singled out' that explicate what exactly is convincing in the data. This is a role played by measures.

2.2.3   A Conceptualization of Measure

Although the list of functions of measures presented above possibly is not complete, it illustrates some common facets of the use of measures on which each function places different emphasis, in different terms, with varying grades of explicitness.

Measures are *grounded in data*. Although this facet on its own is not terribly surprising, the role of measures becomes clearer when related to another facet: measures *describe phenomena*. They bridge the gap between data and phenomenon. A phenomenon behind some data can be accessed through the use of measures that operate on that data. This can lead to new insights into the phenomenon and is a prerequisite for communication about that phenomenon. Measures however can never capture the full phenomenon but provide *discrete* descriptions. They separate phenomena into relevant and irrelevant parts, highlighting only very specific *aspects* of phenomena. This is the reason why they can provide convincing principled arguments and give new, but also possibly incomplete, insights into phenomena.

From these considerations, this study draws a conceptualization for the concept of measure: a measure is a *data-based description of one aspect of a phenomenon*. This definition builds on a broad understanding of the term 'phenomenon'. 'Aspect of a phenomenon' can refer to any part of a phenomenon that is held to be relevant for a specific question in a specific situation. An example could be the daily ice growth used by climate scientists as a measure of the volatility of the melting process of Arctic sea ice (Fig. 2.1). Another aspect of the same phenomenon could be the general well-being of the Arctic ice sheet, addressed through the measure of monthly average extent (Fetterer et al. 2002). While these aspects are phenomenon-specific, measures can also refer to more general aspects like the central tendency. A distinction can be drawn between *general measure*s that focus on *general* aspects of phenomena
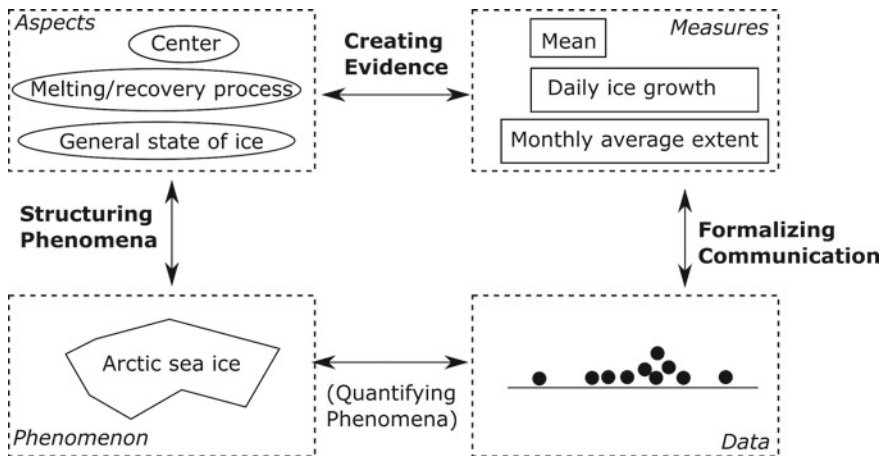

Fig. 2.1   The relations between phenomenon, data, measures, and aspects of phenomena

(center, spread, …) and *situative measure*s that focus on phenomenon-specific *situative* aspects. General measures consist of all measures commonly referred to in formal statistics, whereas situative measures address phenomenon-specific aspects such as the melting process of Arctic sea ice. The meaning of situative measures is often situation-specific, whereas general measures provide situation-independent tools for structuring phenomena. This does not mean that the use of general measures is strictly situation-independent: general measures can also be used to address phenomenon-specific aspects.

2.2.4 The Development of Measures

Whereas statisticians are able to use general measures such as the median to address the center of arbitrary phenomena, the situated nature of knowledge implies that learners will have to resort to phenomenon-specific situative measures when starting out in their learning trajectory. This puts the learners into an inconvenient position. They will need to structure phenomena by identifying aspects, while simultaneously finding situative measures to address just these aspects. Learners need to *develop* their measures. During their learning process, learners will need to answer questions corresponding to a measures' functions of structuring phenomena, formalizing communication, and creating evidence.

As emphasized by Vergnaud (1996), a theory of learning needs to give a prominent place to learners' activities. In order to illustrate how formal ideas can emerge from informal activity, the functions of measures are now (in reference to Freudenthal 1991) interpreted as *mathematizing activities* carried out by the learners while developing measures. When engaging in the mathematizing activity of structuring phenomena, learners make sense of a situation through their concepts-in-action. Their contextual knowledge of the phenomenon plays an important part, as they have not yet developed general measures for structuring phenomena. The mathematizing activity of formalizing communication focuses on a measure's formal characteristics, such as definition, calculation, and rules of application. In the beginning of learning processes, visual identification (i.e. 'just seeing') would be an adequate way of finding an situative measure. However, such visual identification could hardly provide intersubjectivity; finding standard procedures of calculation instead could be an act of formalizing communication. During the mathematizing activity of creating evidence, learners decide the actual aspects and measures to be chosen for argumentation. Again, contextual knowledge can play an important part for clarifying which aspects are relevant for which questions regarding the phenomenon and thus, which line of argumentation should be supported by what evidence.

Through the investigation of different phenomena, operational invariants over different situations can emerge, making the use of situative measures less phenomenon-dependent. In this framework, learning takes the form of developing situative measures into general measures through mathematizing activity across

different situations: broadening the aspects of phenomena addressed by measures, explicating formal characteristics, and supporting argumentation through evidence.

2.2.5 Research Questions

The starting point for this study was the need for a conceptualization of measure that allows for the design of a teaching-learning arrangement that draws on learners' situated understandings and can lead to the development of statistical concepts. Such a teaching-learning arrangement needs to elicit the mathematizing activities of structuring phenomena, formalizing communication, and creating evidence. Since the role of those activities was based on theoretical observations, it remains unclear how actual learning processes are constituted in these activities and how a teaching-learning arrangement can support them. Although all three mathematizing activities play a part in the development of measures, this study limits itself by focusing on the activities of structuring phenomena and formalizing communication in order to provide a more in-depth view of the learning processes. The empirical part of this study thus follows the following research questions:

(RQ1)   How can design elements of a teaching-learning arrangement elicit and support the mathematizing activities?
(RQ2)   How do learners' situative measures develop through the mathematizing activities of structuring phenomena and formalizing communication?

2.3 Research Design

2.3.1 Topic-Specific Didactical Design Research as Framework

Design research as methodological frame

The presented study is part of a larger project in the framework of topic-specific didactical design research (Prediger et al. 2012). This framework simultaneously aims at two different but strongly interconnected goals: empirically grounded theories on the nature of topic-specific learning processes and learning goals (i.e. what and how to learn), and design principles and concrete teaching-learning arrangements for learning this topic (i.e. with what to learn). This is achieved by a focus on learning processes (Prediger et al. 2015). Special attention is given to the careful specification and structuring of the learning content as well as to developing content-specific local theories of teaching and learning (Hußmann and Prediger 2016).

Research is structured into iterative cycles consisting of four different working areas (see Fig. 2.2). In a first working area, the learning content is specified and struc-
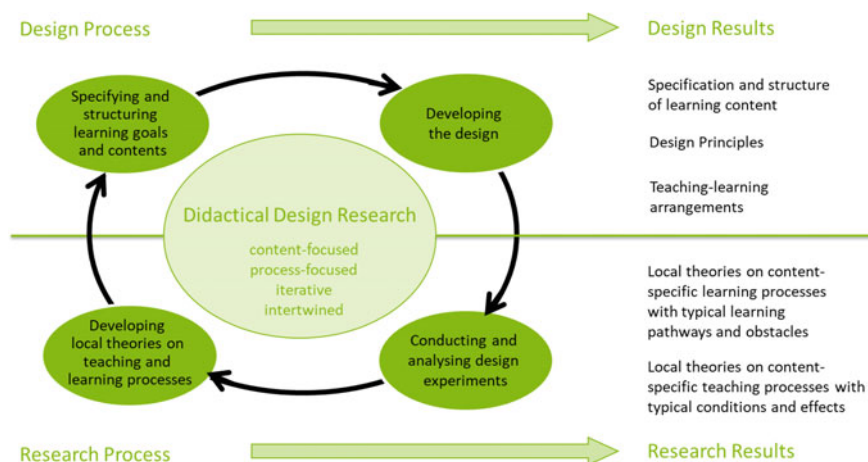
Fig. 2.2 The cycle of topic-specific didactical design research (Prediger et al. 2012; translated in Prediger and Zwetzschler 2013, p. 411)

tured, identifying central insights into the content that learners need to achieve and structuring them into possible learning pathways. This can be based on epistemological considerations such as a didactical phenomenology (Freudenthal 1983) as well as on empirical insights into possible learning obstacles and students' conceptions. The second working area consists of designing a teaching-learning arrangement to be used in the third working area, conducting design experiments (Cobb et al. 2003). The learning processes initiated in the design experiments are then analyzed and serve as a basis in developing local theories about these teaching and learning processes. A main strength of the framework of didactical design research is the interconnectedness of these working areas: in the next cycle, the local theories developed can inform the re-specification and re-structuring of the learning content. This re-structuring in turn influences the design principles enacted in the teaching-learning arrangement and thus, the initiated learning processes. Through this process, theory and design get successively more refined in each cycle.

2.3.1.1 Participants and Data Collection

This study reports on findings of the third cycle of design experiments of the on-going design research project (for other results see Büscher 2017, 2018; Büscher and Schnell 2017; Schnell and Büscher 2015). The design experiment series in the third cycle took place in laboratory settings with five pairs of students in a German middle school (ages 12–14). Each pair took part in a series of two consecutive design experiment sessions of 45 min each. The participating students were chosen by their teacher as performing well or average in mathematics, which includes statistics education in German curricula. At the time of the experiments, the students had very little expe-

rience with statistics besides learning simple measures such as the arithmetic mean and median a year before in grade 6. They were familiar with frequency distributions but only on a rather superficial level (e.g. reading out information on maximum and minimum), without comparing them strategically. They were not familiar with stacked dot plots or measures of spread.

All experiments were completely videotaped (altogether 450 min of video in the third cycle). Here, the case of two pairs of students is presented, selected due to the richness of their communication and mathematizing activities. Their design experiment sessions were fully transcribed.

2.3.1.2   Data Analysis

The qualitative data analysis aims not at solely assigning students' utterances to the general statistical concepts but instead at capturing the individual emergent, situative concepts. In order to capture the richness and heterogeneity of the students' individual reasoning, this study chose a category-developing approach (cf. Mayring 2000) using open and interpretative approaches (cf. Corbin and Strauss 1990) for identifying individual concepts-in-action and theorems-in-action (Vergnaud 1996) based on the students' utterances and gestures. This methodological foundation of the analytical framework by Vergnaud's constructs allow the data analysts to capture the situativity of knowledge and learning. The identified individual concepts-in-action and theorems-in-action on measures are not necessarily in line with general statistics concepts but rather mirror their own situative structure of phenomena. In the analysis, concepts-in-action are symbolized by ‖…‖ and theorems-in-action by <…>.

2.3.2   Design Principles

During the five design experiment cycles, several design principles were implemented and iteratively refined that played a role in initiating concept development. Three design principles play an important part in this study (for a complete overview see Büscher 2018); each of the design principles focused on eliciting a different mathematizing activity.

Investigating realistic phenomena. A teaching-learning arrangement focusing on the development of measures needs to elicit the mathematizing activity of structuring phenomena. Since most students do not yet have access to phenomenon-independent measures to structure arbitrary unknown phenomena, the choice of the phenomenon to be investigated has to be carefully considered. This study uses phenomena such as variability in the weather that are close enough to students' reality so that they can informally and intuitively structure the phenomenon.

Scaffolding the use of measures in argumentation. Previous cycles of the project showed how students did use situative and occasionally even general measures when comparing distributions. Whereas there was a lot of potential in this, their uses stayed elusive: students lacked the language to specify the addressed aspects and formal characteristics of their situative measures—they struggled to formalize their communication. This led to an insecure use of measures, so that they sometimes simply had already forgotten their train of thoughts when prompted by the researcher or other students. This raised the need to scaffold the use of measures by explicating their use in giving arguments about phenomena. This design principle was implemented through the use of so-called report sheets (see below)

Contrasting measures. Central to the measure-focused approach of this study is the insight that different measures for the same distribution can result in different views on the situation by emphasizing different aspects. Thus, engaging in the activity of creating evidence can mean to contrast and evaluate different measures with respect to (a) their usefulness regarding specific investigations, (b) their correspondence to learners' experienced reality, (c) their applicability in different situations, or (d) their advantages or disadvantages in argumentation. This design principle was realized by contrasting different report-sheets (see below).

2.3.3 Task Design

The design of the two sessions of the design experiments consisted of two different tasks, the *Antarctic weather task* (Session I) and the *Arctic sea ice task* (Session II). Each task was structured into different phases, with progressions between phases initiated by the researcher when certain requirements were met.

The Antarctic weather task

The goal of this task was to introduce the students to the idea of measures and the design elements central to the whole design experiment. The task was structured into three phases.

Phase I.1. The students were given dot plots of temperature distribution at the Norwegian Antarctic research station *Troll forskningsstasjon* (Fig. 2.3, data slightly modified from Stroeve and Shuman 2004) and introduced to the setting of the task: as consultants to researchers planning a trip next year, they were charged with giving a report of the temperature conditions. Since the students were unfamiliar with dot plots, special attention was given to make sure that students understood the diagrams. The data were presented to the students on a tablet with a screen overlay software to allow for drawing visualizations of their situative measures directly onto the screen. *Tinkerplots2.0* (Konold and Miller 2011) was used to create the diagrams, without giving the students access to interactive functionalities of the software. When sufficient understanding of the diagrams was achieved and the students had given some informal predictions for next year, the task progressed to Phase I.2.
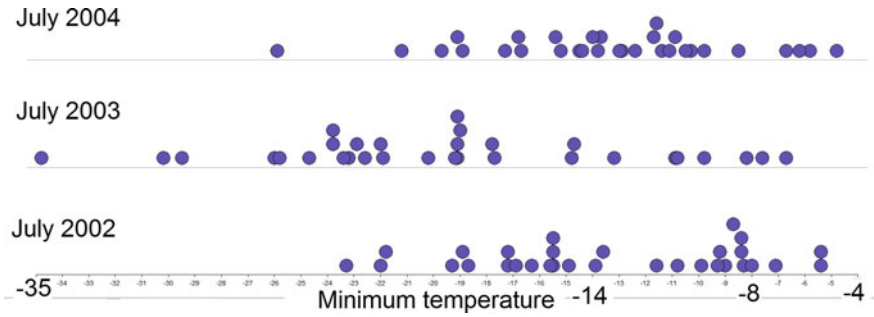
Fig. 2.3 Distributions of the Antarctic weather task (translated from German)

Fig. 2.4 Empty report sheet (translated from German)

Phase I.2. In this phase, the students were introduced to the central design element of the design experiments, the *report sheets* (Fig. 2.4). These report sheets served as a scaffold for argumentation with measures, combining a graphical representation with measures and a brief inference about the phenomenon of Antarctic weather. The students were asked to fill out a report sheet to be used as a report for the researchers. The measures employed were given to them without explanation, so that they needed to find their individual interpretation of minimum, maximum, and typical. Typical here served as an situative measure for a yet unspecified situative aspect, which could be interpreted by the students as incorporating some aspect of variability (similar to Konold et al. 2002) Since formal characteristics and the meanings of the measures were left unspecified, this task aimed at eliciting the mathematizing activities of structuring phenomena and formalizing communication.

Phase I.3. After the students had created their own report sheet, they were given fictitious students' *filled-in report sheets* (Fig. 2.5). These report sheet differed in their interpretations of the measures employed and thus focused on different aspects of the phenomenon. The students were asked to evaluate these report sheets and to possibly adapt their own report sheet.
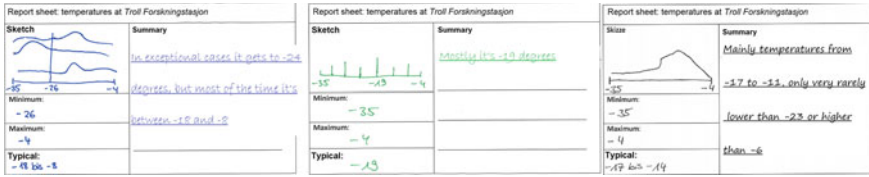
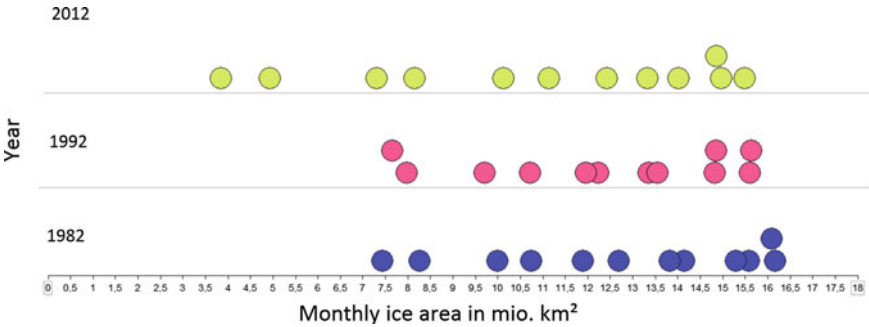Fig. 2.5   Fictitious students' filled-in report sheets (translated from German)

Fig. 2.6   Distribution of sea ice extent in Session II (translated from German)

The Arctic sea ice task

Phase II.1. The Arctic sea ice task followed a similar progression as the Antarctic weather task. This time the students were put into the roles of researchers of climate changes. Students were given distributions of monthly lowest Arctic sea ice extent for the years 1982, 1992, and 2012 (Fig. 2.6; data slightly modified from Fetterer et al. 2002) and were asked to give a report whether, and how much, the ice area had changed. This phase again aimed at ensuring the students' understanding of the diagram and the context. They were not yet asked to create a report sheet.

Phase II.2. Following the introduction of the setting, the students again received filled-in report sheets (Fig. 2.7). These report sheets now allowed for arbitrary measures and again presented different formalizations of measures and abstractions of the phenomenon. This time the different measures lead to radically different perceptions of the phenomenon of Arctic sea ice, with report sheets proclaiming either no change or radical change in Arctic sea ice (Fig. 2.7). Discussion revolved around which report sheet was right, and what a researcher would need focus on when reporting on Arctic sea ice, thus eliciting the mathematizing activity of creating evidence.

Phase II.3. Following the discussion, the students were again asked to create their own report sheet. Whereas the students were free to choose their measures for the report sheet, the students were expected to adapt elements of the filled-in report sheets for their own report sheet. This initiated further mathematizing activities, as the students were asked to justify their choice of measures.
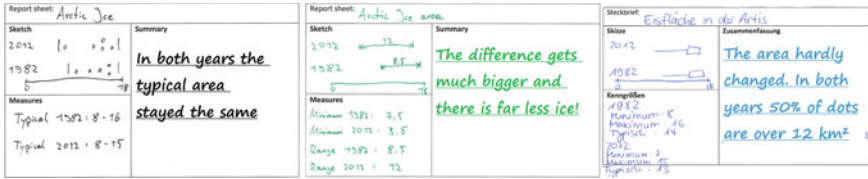
Fig. 2.7 Filled-in report sheets for Session II (translated from German)

2.4 Empirical Results

This study identifies students' mathematizing activities to investigate their developing measures. The first part of this section follows the learning processes of two students, Maria and Natalie, through both sessions of the design experiment. Due to the rapid changes of the roles in the students' interaction, the transcript has been partially cleaned up to increase readability. The analysis focuses on their use of the situative measure of Typical (reference to the situative measure Typical indicated by capital-T), from its unspecified beginning in Session I to its more formalized version at the end of Session II. During the design experiment, the students get increasingly precise in addressing different aspects of phenomena and in structuring the phenomenon. This is then briefly contrasted with the processes of another pair, Quanna and Rebecca, focusing on Session II and highlighting similarities and differences in the two pairs' use of Typical.

2.4.1 The Case of Maria and Natalie

Session I: The Antarctic weather task

The first snapshot starts with Phase I.1 of the Antarctic weather task. After giving some informal predictions of the weather, Maria (M) and Natalie (N) try to explicate their view on the data to the researcher (I).

1 M We are pondering what the relationship, like, how to…
2 N Yes, because we want to know what changes in each year. And we said that there *[points to 2003]* it came apart.

[…]

8 M Yes, I think it *[points to 2004]* is somehow similar to that *[points to 2002]*, but that one *[points to 2003]* is different.
9 N Like here *[points to 2004, around −12 °C]* are, like, like the most dots, and here *[points to 2002, −12 °C]* are almost none. And there *[points to 2002, −8 °C]* are the most and here *[points to 2004, −8 °C]* are almost none.
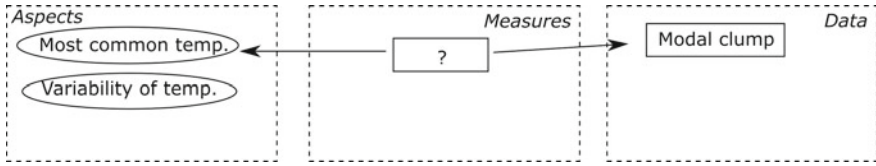
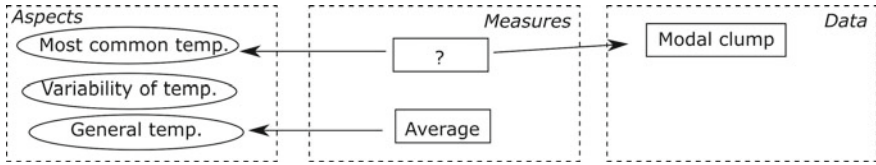**Fig. 2.8** Maria and Natalie's use of measures (Part 1)

Fig. 2.9 Maria and Natalie's use of measures (Part 2)

This excerpt serves as an illustration of the starting point in the students' reasoning. The students are trying to characterize the differences observed in the distributions. In order to do this, they structure the phenomenon by identifying two aspects: the ||*most common temperatures*|| (where "the most" temperatures lie, #9), and the ||*variability of temperatures*|| (how they "came apart", #2). Whereas the students are able to use modal clumps as a way to address the ||*most common temperatures*||, they seem to lack ways of addressing the ||*variability of temperatures*|| (Fig. 2.8).

A few minutes later, the students find a way to better address the difference between the distributions.

21   M   Well, we first should look at how many degrees it has risen or fallen. Generally. In two years.

[…]

27   N   You mean average, like…
28   M   The average, and then we look at how the average changed in two years.

By identifying the aspect of a ||*general temperature*|| ("Generally", #21), the students are able to re-structure the phenomenon to reduce the complexity of the temperatures. For this aspect, they appear to already know an adequate general measure: the ||*average*||. To the students, <*the average addresses the general temperature of a distribution*>. This ||*general temperature*|| does not necessarily correspond to the ||*most common temperatures*|| addressed earlier. In this way, the phenomenon gains additional structure (Fig. 2.9).

The design experiment progresses through Phase I.2, in which the students create their own report sheet (Fig. 2.10). The analysis picks up at beginning of Phase I.3, with the students comparing the different interpretations of Typical in the filled-in report sheets.
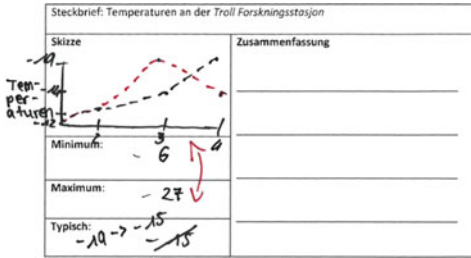
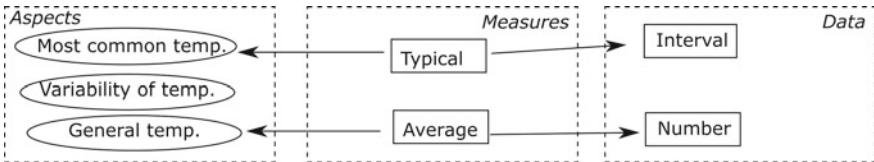Fig. 2.10   Maria and Natalie's first report sheet

Fig. 2.11   Maria and Natalie's use of measures (Part 3)

Comparing the different interpretations of Typical, Maria and Natalie are intrigued by the possibility to use an interval to formalize Typical. This consideration leads them to reflect on their use of the average.

41   N   But the average temperature isn't really typical, is it?
42   M   What, typical? Of course the average temperature is the typical.

[…]

46   M   Well, no. Typical is more like where the most… no…
47   M   The average temperature isn't the typical after all. Because it's only the general, the whole. The typical would be for example for this *[2004]* here *[points to −14 on the 2004 dot plot]*.
48   N   Typical I think simply is what is the most or the most common.

The students differentiate between average and Typical to address different aspects: The ||*general temperature*|| is addressed by the general measure ||*average*|| ("the general, the whole", #47), and the ||*most common temperatures*|| addressed by situative measure ||*Typical*|| ("the most common", #48). At this point it is not yet clear if the situative measure Typical consists of a number or an interval—it is still in need of formalization. However, introduction of this situative measure seems to have allowed the students to reconnect to the aspect of ||*most common temperatures*|| (first expressed in #9) that got swept aside by the more formalized average (Fig. 2.11).

Some minutes later, Natalie summarizes her view on the relation between Typical and average.

61 N Average is pretty imprecise, because it doesn't say anything about a single day. And with Typical, I'd say, that it's a span between two numbers, because that way you can better overlook how it is most of the time.

In the end of Session I, the measures Typical and average address two different aspects of the phenomenon of Antarctic weather. Whereas the average addresses the general temperature, Typical describes the most common temperatures. The average can be used to compare distributions, whereas Typical gives an insight into a range of 'normal' or 'expected' temperatures, to which any single day can be compared. Central to this distinction was the formalization of Typical as an interval.

Session II: The Arctic sea ice task

Most of Session II revolved around the question how to further formalize Typical, and how to distinguish it from the average. This excerpt starts in the middle of Phase II.2, and takes place over a period of eight minutes. In the preceding minutes, the students had used the average to propose a general decline in Arctic sea ice.

1 I Last time we talked about Typical, and here Typical is also drawn in. Do you think that's helpful, or not?
2 M Typical, wait a second, there *[report sheet 3, 1982]* Typical is 14 right? Huh, but why is 13 Typical here *[report sheet 3, 2012]?*
3 N Huh, Typical can't be 13, because Typical actually is a range, isn't it?

[…]

6 I What would you say what one should choose?
7 N I would definitely say a range, because that just tells you more. Because you can't say that it's 11 degrees typical.

Maria and Natalie are irritated by the same report sheets showing different values for the measure ||*Typical*|| (#2). This leads them to question whether Typical should be formalized as a number or an interval ("range", #3).

In Session I, the students opted for the interval. Natalie draws on this knowledge, postulating that<*Typical cannot be a number, because numbers do not describe Typical temperatures*> ("you can't say that it's 11 degrees typical", #7). In this way, she uses the situative aspect of ||*most common temperatures*|| from Session I to formalize the situative measure Typical in another phenomenon as an interval. Because this transfer of phenomena happens frequently throughout the session (see below), it could be seen as the emergence of operational invariants across situations, rather than a simple mistake in wording.

Some moments later, after they have again considered the average, the students compare the two measures.
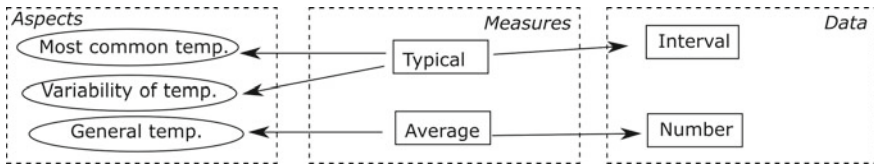
Fig. 2.12 Maria and Natalie's use of measures (Part 4)

21	I	And if you would create such a report sheet, would the average suffice?
22	N	No. Well I think the average is important, isn't it? But a range, what's typical, that just tells you more about single days than if you take the average.
23	N	Because if the average is like 12, then one day could be 18 degrees, or −10 degrees or something. And the average better tells you what generally happened, and I think a range better tells you what happened generally.
24	N	Because if the average was 8 degrees, but it also happened to get to 18 degrees or −10 degrees, then the range would rather be from 5 degrees to – I don't know.

Natalie distinguishes between two aspects: what "generally happened", and what "happened generally" (#23). These are two different (yet unnamed) aspects, because the distinction serves as an explanation of the distinction between average and Typical (sometimes referred to by Natalie as "range", #23). Natalie seems to lack the vocabulary to clearly differentiate between the two aspects. In her explanation however she again seems to draw on an aspect of the previous session: the ‖*variability of temperatures*‖, as she states that <*a high variability of temperatures can be seen in the Typical range*>(temperatures from 18 to −10 would somehow be reflected in the "range", #24), whereas <*the average is not impacted by the variability of temperatures*>(the average would stay at 8 degrees, #24). Again, the formalization of the measure Typical progresses by drawing on the structuring of another phenomenon (Fig. 2.12).

Following this exchange, after some minutes, the students return to the problem of finding the Typical interval.

41	N	I don't know how to calculate Typical. I think you start from the average, and then looks at the lowest and highest temperatures, and from that you take a middle value. Like between the average, and…
42	M	And the lowest and the highest… we are talking about temperatures the whole time, but those aren't temperatures.
43	N	Yes but if we took temperatures, then you take the average and the coldest and then again take the average.

[…]

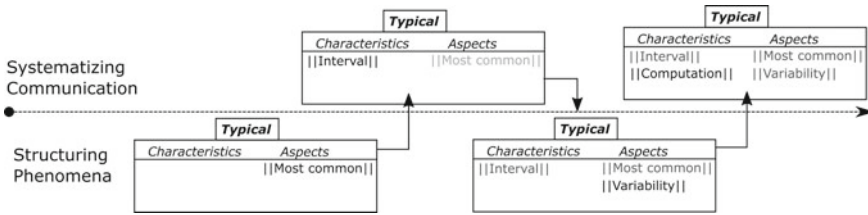| 48 | N | And then the average from the average is the Typical. Between this average and that. |

Fig. 2.13 Maria and Natalie's development of Typical

With the aspects addressed by the two measures now firmly separated, the students find a way to calculate their Typical interval: Taking the average of the whole distribution ("start from the average, #41), splitting the distribution into two halves at this point ("look at the lowest and highest temperatures, #41), calculating the average for each of those halves ("again take the average, #43), and then taking the interval between those two averages ("between this average and that", #48). This shows a highly formalized use of the average: the ||*general temperature*|| addressed by the measure ||*average*|| seems to also apply to only halves of distributions. This excerpt is the first time one of the students becomes aware of their substitution of the phenomenon of Arctic sea ice with Antarctic temperatures (#42). The casualness of Natalie's dismissal of this fact ("yes but if we took temperatures", #43) however seems to suggest that the operational invariants of Typical in the end seems to encompass both situations, temperatures and sea ice.

Summary. During Session I, Maria and Natalie structure the phenomenon of Antarctic temperatures into ||*most common temperatures*||, ||*general temperature*||, and ||*variability of temperatures*||. They also determine formal characteristics of the measure Typical by formalizing it as an interval, in contrast to the average. This distinction is transferred to another phenomenon in Session II, but not without problems: again, the characteristic of Typical as an interval must be justified. In the end, the students even arrive at a way of finding the Typical interval that is similar to that of finding the interquartile range. During the whole learning process, the situative measure of Typical develops in interrelated mathematizing activities of structuring phenomena and formalizing communication. Figure 2.13 provides an overview on this development.

2.4.2   The Case of Quanna and Rebecca

The following empirical snapshot follows the students Quanna (Q) and Rebecca (R) in Session II of the design experiment. The excerpts stem from a conversation of about 15 min. The snapshot starts in Phase II.3 with the students filling out their own report sheet (shown in Fig. 2.14, but not completed until turn #40) after they have discussed the filled-in report sheets.
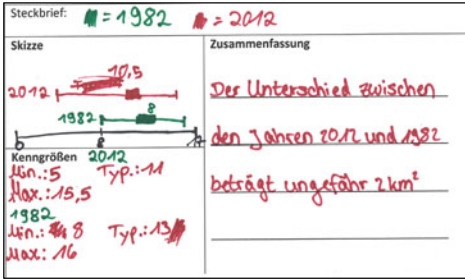
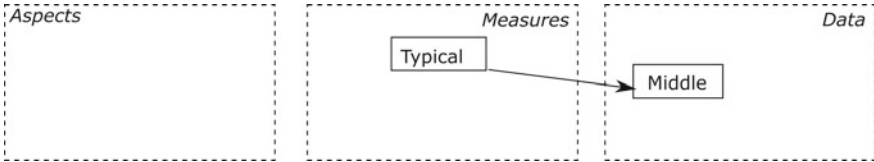**Fig. 2.14** Quanna and Rebecca's report sheet

Fig. 2.15 Quanna and Rebecca's use of measures (Part 1)

1	Q	*[while filling out own report sheet]* And Typical…
2	R	Typical […] it could be, like, the middle or something?
3	R	I would say the middle and a bit higher.

Although they could have referred to other measures, ||*Typical*|| is the main measure organizing their view on the phenomenon. Without paying attention on the aspects to be addressed, the students are formalizing ||*Typical*|| as located in the ||*middle*|| of the distribution: *<Typical is located a bit higher than the middle>* (#3) (Fig. 2.15).

Some minutes later, the students are about to write their summary for the report sheet.

20	Q	Okay, now the summary.
21	R	The numbers got *[points to own report sheet]* – look – more ice melted away.
22	Q	*[shakes head]* the difference is – is around 2.5.
23	R	Always?
24	Q	Yes, right here *[points to own report sheet]* of Typical.

Rebecca seems to have difficulties with combining the phenomenon (the melting ice) with the task of giving a short data-backed summary. At this point, Quanna is able to utilize their measure of Typical. In the meantime, the students had decided that *<Typical is a number>*, which they intuitively identified for the distributions of 1982 and 2012 as 11 and 13.5. These numbers show a difference of 2.5, which can now be used in their summary to report on the Arctic sea ice decline: *<Typical*
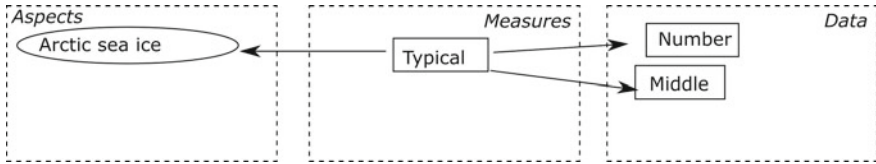
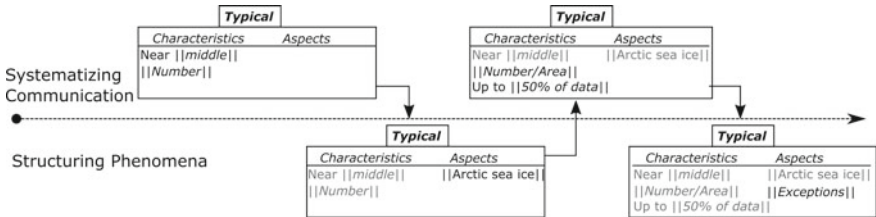**Fig. 2.16** Quanna and Rebecca's use of measures (Part 2)

Fig. 2.17 Quanna and Rebecca's development of Typical

can be used to address the state of Arctic sea ice> (the melting ice, #21, addressed through the difference of Typical, #22). However, it remains unspecified what exactly is meant by this aspect of a general ||*state of Arctic sea ice*|| (Fig. 2.16).

The characteristics of Typical still being unclear, the researcher challenges them to explain their use of Typical.

40   I   I see you decided to use only one number for Typical, in contrast to this report sheet, where they used an area *[points to filled-in report sheet]*. Is that better or worse, what do you think?

41   R   Well Typical is more of a single…

42   Q   *[simultaneously]* more of an area…

43   R   Now we disagree. […] Typical is more of a small area, or you could say a number. Like here, from 10 to 12. […] If the area is over 100, it may be over 10. […] But never more than the half.

The claim *<Typical is a number>* becomes disputed, as ||*number*|| and ||*area*|| both are possible characteristics of Typical, as evidenced by the filled-in report sheet in Fig. 2.17. This initiates further processes of formalization, resulting in more pronounced formal characteristics of Typical. Whereas there still is no full definition, there are criteria for its correct form: *<Typical is an area that at most covers half the data>* ("never more than the half", #43) and *<Typical can be signified by a number, if the area is small>* ("small area, or you could say a number", #43).

Some minutes later in the discussion, Rebecca tackles the question whether one is allowed to omit data points that could be seen as exceptions when creating report sheets.

61 R   Well, you can do that, but it depends. You have to make sure it fits. If you do it like here *[points to own report sheet]* you should not consider the isolated cases […] because then it gets imprecise. But if the Typical area was the same on both sides, I think you can do that.

Whereas there still is no full definition of Typical, another situative aspect has been added that is addressed by Typical. Typical not only functions as a description of a general ||*state of Arctic sea ice*||, but also addresses ||*rule and exceptions*|| of the Arctic ice: <*If the Typical area of two distributions is the same, one can use Typical to address exceptions*>. In this way, the formalization of Typical as an interval in the middle of the distribution has allowed for addressing a previously unstructured aspect of the phenomenon.

Summary. Throughout this episode, the students expand the aspects addressed through Typical as well as the situative measures' formal characteristics. In the end, they use Typical to address a wide range of aspects that could also be addressed through general statistical measures (Fig. 2.17). The differentiation of aspects of phenomena and the growing explicitness in formal characteristics of Typical took place in interlocking mathematizing activities of structuring phenomena and formalizing communication: after Typical has become sufficiently formalized, it could be used to structure the phenomenon of Arctic sea ice into ||*rule and exceptions*||.

2.5   Conclusion

This study started out from the need for a conceptualization of learners' situated understandings and the development of statistical concepts through their activities in learning processes. The concept of measure was introduced, distinguishing between general and situative measures: measures that address phenomenon-specific aspects without necessarily showing explicit formal characteristics. Learning took place during the development of learners' situative measures through the three mathematizing activities of structuring phenomena, formalizing communication, and creating evidence. An empirical study was then used to illustrate (a) how learning processes can be understood through this conceptualization of measure and (b) how the design of a teaching-learning arrangement can influence these learning processes.

2.5.1   The Development of Measures

The analysis shows the students to be fully engaged in the mathematizing activities, which presented themselves as being intricately connected. Structuring phenomena into aspects provided the reason for formalizing the measures, and additional formal characteristics found for measures initiated further structuring of the phenomenon.

The more the students formalized their measure, the more situations were included in the operational invariants of the measure.

The interpretative approach to the analysis revealed the phenomenon-specificity of the students' measures. Maria and Natalie did not use the general measure of the average to address a general aspect of center, but an situative aspect of general temperature. This was then contrasted with the situative measure Typical, which was used to address a range of expected temperatures. Using Typical to structure the new phenomenon of Arctic sea ice took explicit reference back to the phenomenon of Antarctic temperatures. In this way, the students' knowledge of the structure of phenomena influenced their development of measures. They did not simply use measures to make sense of phenomena, but knowledge of situation and measure emerged at the same time.

One strategy that emerged for Maria and Natalie was the comparison of measures with differing degrees of formalization. Because the students knew the formal characteristics and aspects addressed by the general measure average, they could use it to develop the situative measure Typical. The average could even be employed in the calculation of Typical, leading to a measure that addressed aspects that could not adequately be addressed previously.

One idea postulated in the framework was the possibility of development of situative into general measures. Although the learning processes investigated in this study ended before the development of general measures, the findings suggest that this would indeed be possible. Both pairs of students ended with an situative measure Typical that resembled the general measure of the interquartile range. Quanna and Rebecca used Typical to describe an area in the middle of the distribution, consisting of no more than half the data points, indicating the location of the densest area, partitioning the distribution into rule and exception. Maria and Natalie calculated Typical by finding multiple averages, which would have resulted in the interquartile range had the average been substituted by the median. In their combination of average and Typical, Maria and Natalie manage to coordinate different measures, showing the possibility of creating understanding even for conceptually rich representations such as boxplots (cf. Bakker et al. 2004).

2.5.2   Supporting Mathematizing Activities

Central to the design of the teaching-learning arrangement was the choice of phenomenon to be investigated. Theoretical considerations led to the design principle of choosing realistic phenomena to be investigated. The choice of Antarctic weather and Arctic sea ice proved to be a fruitful one: in the case of Maria and Natalie, the students could identify aspects of phenomena regarding the natural variability and central tendency of weather. Through identification of operational invariants across the phenomena, the corresponding measures could then be broadened to also structure the phenomenon of Arctic sea ice.

Another design principle was the scaffolding of the use of measures in argumentation through the report sheets. This provided the students with different situative measures both pairs could appropriate for their individual reasoning. Since these measures were provided without explanation, and with different formal characteristics, the students needed to choose and commit to certain characteristics. In this way, this design principle of contrasting models led to the activity of formalizing communication.

2.5.3 Limitations and Outlook

With the study limited by its own situativity in the design of the teaching-learning arrangement and number of students analyzed, careful consideration has to be given to the generalizability of the results. The investigated development of measures has to be understood in the context of the design: the mathematizing activities were influenced by design elements, students, and the researcher. Any change in these factors could result in very different learning processes.

Yet the nature of this study was that of an existence proof of concept development and an illustration of a theoretical concept revealing a richness within the students' learning processes. Aiming for ecological validity (Prediger et al. 2015), this richness observed with only two pairs of students calls for analysis of additional pairs. Some results already indicate a wealth of strategies and conceptions, along with similarities in the development of measures (Büscher 2017, 2018; Büscher and Schnell 2017).

The analysis also showed the importance of the phenomenon not only as a motivating factor, but as integral to concept development itself. Further research could also be broadened to include other phenomena to be investigated. Since the learning processes was bound to the phenomena, a task design that focuses on other phenomena than weather and ice could provide other starting points for the development of measures.

References

Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ: Erlbaum.

Bakker, A., Biehler, R., & Konold, C. (2004). Should young students learn about boxplots? In G. Burrill & M. Camden (Eds.), *Curricular development in statistics education* (pp. 163–173). Voorburg, The Netherlands: International Statistical Institute.

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

Ben-Zvi, D., Bakker, A., & Makar, K. (2015). Learning to reason from samples. *Educational Studies in Mathematics, 88*(3), 291–303.

Büscher, C. (2017, February). *Common patterns of thought and statistics: Accessing variability through the typical*. Paper presented at the Tenth Congress of the European Society for Research in Mathematics Education, Dublin, Ireland.

Büscher, C. (2018). *Mathematical literacy on statistical measures: A design research study*. Wiesbaden: Springer.

Büscher, C., & Schnell, S. (2017). Students' emergent modelling of statistical measures—A case study. *Statistics Education Research Journal, 16*(2), 144–162.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9–13.

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology, 13*(1), 3–21.

Fetterer, F., Knowles, K., Meier, W., & Savoie, M. (2002, updated daily). *Sea ice index, version 1: Arctic Sea ice extent*. NSIDC: National Snow and Ice Data Center.

Fischer, R. (1988). Didactics, mathematics, and communication. *For the Learning of Mathematics, 8*(2), 20–30.

Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht, The Netherlands: Reidel.

Freudenthal, H. (1991). *Revisiting mathematics education: China lectures*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Gravemeijer, K. (2007). Emergent modeling and iterative processes of design and improvement in mathematics education. In *Plenary lecture at the APEC-TSUKUBA International Conference III, Innovation of Classroom Teaching and Learning through Lesson Study—Focusing on Mathematical Communication* . Tokyo and Kanazawa, Japan.

Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist, 53*(1), 5–26.

Hußmann, S., & Prediger, S. (2016). Specifying and structuring mathematical topics. *Journal für Mathematik-Didaktik, 37*(S1), 33–67.

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics, 88*(3), 305–325.

Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R., et al. (2002, July). *Students' use of modal clumps to summarize data*. Paper presented at the Sixth International Conference on Teaching Statistics, Cape Town, South Africa.

Konold, C., & Miller, C. D. (2011). *Tinkerplots: Dynamic data exploration*. Emeryville, CA: Key Curriculum Press.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal, 41*(3), 645–679.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82–105.

Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning, 13*(1–2), 152–173.

Mayring, P. (2000). Qualitative content analysis. *Forum Qualitative Social Sciences*, *1*(2). Retrieved from http://www.qualitative-research.net/index.php/fqs/issue/view/28

Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.

Prediger, S., Gravemeijer, K., & Confrey, J. (2015). Design research with a focus on learning processes: An overview on achievements and challenges. *ZDM Mathematics Education, 47*(6), 877–891.

Prediger, S., Link, M., Hinz, R., Hußmann, S., Thiele, J., & Ralle, B. (2012). Lehr-Lernprozesse initiieren und erforschen—fachdidaktische Entwicklungsforschung im Dortmunder Modell [Initiating and investigating teaching-learning processes—topic-specific didactical design research in the Dortmund model]. *Mathematischer und Naturwissenschaftlicher Unterricht, 65*(8), 452–457.

Prediger, S., & Zwetzschler, L. (2013). Topic-specific design research with a focus on learning processes: The case of understanding algebraic equivalence in grade 8. In T. Plomp & N. Nieveen (Eds.), *Educational design research—Part A: An introduction* (pp. 409–423). Enschede, The Netherlands: SLO.

Schnell, S., & Büscher, C. (2015). Individual concepts of students Comparing distribution. In K. Krainer & N. Vondrová (Eds.), *Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education* (pp. 754–760).

Stroeve, J. & Shuman, C. (2004). *Historical Arctic and Antarctic surface observational data*, *version 1*. Retrieved from http://nsidc.org/data/nsidc-0190.

Vergnaud, G. (1990). Epistemology and psychology of mathematics education. In P. Nesher (Ed.), *ICMI study series. Mathematics and cognition: A research synthesis by the International Group for the Psychology of Mathematics Education* (pp. 14–30). Cambridge: Cambridge University Press.

Vergnaud, G. (1996). The theory of conceptual fields. In L. P. Steffe (Ed.), *Theories of mathematical learning* (pp. 219–239). Mahwah, NJ: Lawrence Erlbaum Associates.