Individual concepts of students comparing distributions

Susanne Schnell[1] and Christian Büscher[2]

[1]TU Dortmund University, IEEM, Germany, susanne.schnell@tu-dortmund.de,
[2]TU Dortmund University, IEEM, Germany, christian.buescher@tu-dortmund.de

While distributions were long understood as "fundamental given of statistical reasoning" (Wild 2006, 10), recent research uncovered students' difficulties in acquiring the underlying conceptual structure and making statistically sustainable inferences when comparing distributions. Research so far describes informal strategies such as using 'clumps' as productive when learning to compare distributions. However, more insights are needed regarding why some of these strategies are chosen in certain situations in order to fully relate students' informal strategies to statistical concepts and measures in task design. This paper aims at clarifying the students' reasoning behind what to focus on for comparisons. We will present empirical snapshots from students in grade 8 (13-15 years old) who focus almost exclusively on absolute frequencies of dots and specifically determined intervals for comparing distributions.

Keywords: Statistical Reasoning; Comparing Distributions; Design Research

INTRODUCTION

Comparing empirical distributions has a high value for statistics education: "Concepts and judgments involved in comparing groups have been found to be a productive vehicle for motivating learners to reason statistically and are critical for building the intuitive foundation for inferential reasoning" (Ben-Zvi 2003, p. 1; also Bakker & Gravemeijer 2004). A combination of descriptive and inferential reasoning is needed in order to make sense of differences and commonalities of two (or more) distributions and go beyond the data at hand. Thus, this activity is an important part of data analysis. The necessary concepts and insights are "multifaceted" (Ben-Zvi 2003) as properties of and between distributions have to be considered. The aim of this paper is to explore students underlying informal resources and rationales as well as conceptual difficulties as starting points for establishing a suitable learning environment.

PROPERTIES *OF* AND *BETWEEN* DISTRIBUTIONS

From a normative perspective, comparing distributions statistically requires students to perceive a distribution as an "organizing conceptual structure with which they can conceive the aggregate instead of just the individual values" (Bakker & Gravemeijer 2004, p. 148). Wild (1999) calls a distribution a lens, through which variation is looked at by "set[ting] aside case labels" (p. 11). This short descriptions already points out that a distributions is in fact a net of different intertwined concepts: Centre, spread, density and skewedness are *properties of* a distribution and constitute its shape (ibid., Ben-Zvi 2003). “The concept of distribution has a complex structure, but this concept is also part of a larger structure consisting of big ideas such as variation and sampling (…). [One can] deal informally and coherently with all these big ideas at the same time with distribution in a central position.” (Bakker & Gravemeijer 2004, p. 149).

These properties can be approached formally (e.g. calculating arithmetic mean, mode and median as measures for the centre), but also in more phenomenological and visual ways (e.g. determining intervals with high density, gaps and clusters; cf. Pfannkuch et al. 2010 for the visual approach). For clarification, this paper uses ‘properties’ to refer to statistical concepts and ‘features’ to more visual aspects of a distribution.

Inherent in the statistical concept of distribution is the necessity to not only focus on single data points or small groups (so-called local view; Ben-Zvi & Arcavi 2001), but to perceive a distribution as a whole, allowing to “search for, recognize, describe and explain general patterns in a set of data” (so-called global view; Ben-Zvi & Arcavi 2001, p. 38). Especially the latter is fundamental to statistical reasoning, but also challenging for students to acquire (ibid.).

When comparing two or more distributions, properties have to be put *in relation between* the distribution, adding further relative insights such as overlap, shift and unusual features (e.g. outliers; Pfannkuch et al. 2010; cf. Ben-Zvi 2003 for comparing measures of variation within and between groups). The comparison can then also allow for new insights into the peculiarities of the initial distribution: For instance, looking at a set of temperatures from July 2014 on the mountain Zugspitze could become an indicator for the effects of global warming when put in relation to the distributions of the 1900s.

INDIVIDUAL APPROACHES TO COMPARING DISTRIBUTIONS

It is not surprising that this complex interplay of concepts is challenging for students: Recent research points out that taking a global view on distributions rather than focussing on single data points or groups is especially challenging (Ben-Zvi & Arcavi 2001; Bakker & Gravemeijer 2004). Problems persist even after instruction in statistics (Ben-Zvi 2003; Konold et al. 1997): students who are familiar with formal measures such as mean and median for single distributions do not make use of them when comparing distributions (e.g. Watson & Moritz 1999; Konold et al. 1997). As Konold et al. (1997) argue, this might indicate a lacking understanding of averages as properties that represent a distribution.

However, some informal strategies were repeatedly shown, which offer productive starting points for structuring learning pathways: Focussing on visually remarkable aspects of distributions (e.g. represented as dot plots), learners make use of informal concepts such as ‘clumps’, ‘hills’ or ‘chunks’ to describe and compare distributions

(Bakker & Gravemeijer 2004; Konold 2002; Cobb 1999). Konold (2002) for instance describes how students use ranges of data in the heart of the distribution (“modal clumps”), which he interprets as vehicles for describing the centre (average) and at the same time the variation of data points. Bakker & Gravemeijer (2004) show that students divide given distributions in three groups (low, middle and high), which are then interpreted in the given context and compared. They understand this as steps from a local to a more global view.

While many studies reproduced the use of informal descriptors such as ‘bumps’, it remains open what underlying rationale guides students in choosing or dismissing features such as modal clumps in situations involving the comparison of distributions. Understanding *why* certain foci are chosen to compare might provide further insights into how task design has to be structured to promote development of statistical reasoning. These questions call for research on the micro-level and reconstructing step by step the individual concepts activated by students and the foci they take on the distribution.

METHODOLOGY AND DESIGN OF THE CASE STUDY

The presented study is part of a larger design research project using the methodological framework of topic-specific Didactical Design Research (Prediger & Schnell 2014; Prediger et al. 2012), which has two intertwined aims: (1) designing a teaching-learning arrangement to facilitate the acquisition of the concept of distribution by comparing distributions and (2) deepening the understanding of the processes of conceptual development on an epistemological level. The design research is conducted by iterative cycles of design experiments, consisting of closely related phases of (re-)structuring learning goals, (re-)constructing the teaching-learning arrangement, conducting and analysing the design experiments and developing local theories. By combining process-oriented analysis and construction of teaching-learning arrangement, this framework provides for the need of research on the micro-level outlined above. Situated early in the research process, this first design experiment cycle aimed to explore how students in German middle schools (informally) compare frequency distributions (represented as stacked dot plots) and identify individual approaches as starting points for task design. Specifically, the research aimed at finding answers to the following questions:

(RQ1) Which individual concepts do students use to compare distributions?

(RQ2) Which rationales guide students in choosing certain foci for comparing distributions?

Data collection

To investigate the complex processes of comparing distributions, we conducted and videotaped design experiments (45 to 60 minutes) in a laboratory setting (cf. Cobb et al. 2003) with three pairs of students, aged 13 to 15. To make sure that students were familiar with distributions as a prerequisite for the activities, we chose students who had learnt about box plots in class a few weeks before the experiments. With statistics playing only a small role in German mathematics education, the students learned to construct boxplots and interpret them in a course over six lessons. The focus was on formal methods, e.g. for determining the five parameters. There was only limited attention given to informally examining variation in terms of centre and spread. The students were not familiar with stacked dot plots. Guiding the experiments was the July climate task, comparing stacked dot plots of temperature on a mountain in July in different years (Fig. 1).

Students were tasked with comparing the temperatures in July on the top of the Zugspitze in the years 2002, 2004 and later 2007 in order to determine the warmest month. Although the graphs were created in Tinkerplots, the students at this point had only access to the plots printed on paper to encourage informal statistical reasoning without focussing on pre-given measurement and tools. Later in the experiment, the students were also given boxplots that had to be matched with the according dot plots.

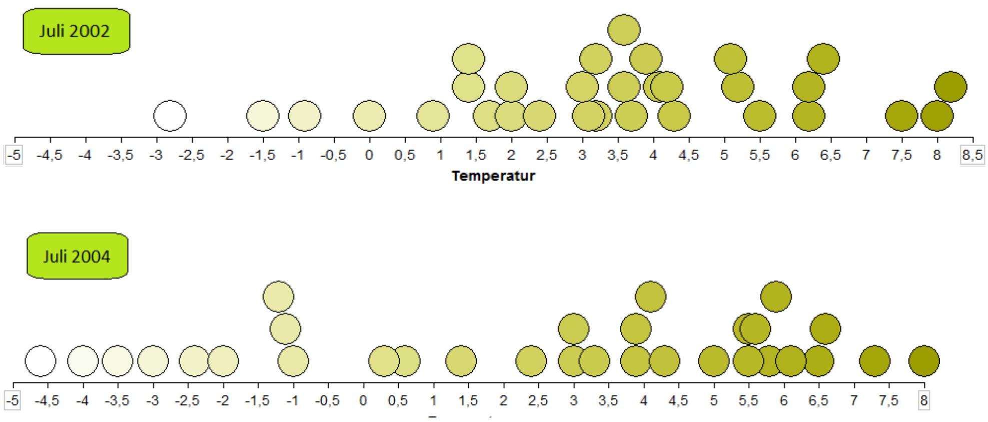

Fig. 1. The stacked dot plots of the July climate task (July 2002 and 2004)

Data analysis

The fine-grained analysis is conducted under an interpretative paradigm using the framework of Vergnaud's Theory of Conceptual Fields (Vergnaud 1996). To give insight into students' individual concepts, we adapted the theoretical construct 'concept-in-action', which is defined as "categories (objects, properties, relationships, transformations, processes, etc.) that enable the subject to cut the world into distinct elements and aspects, and pick up the most adequate selection of information" (Vergnaud 1996, 225). The reconstructed concepts-in-action are symbolised as ||…|| and can provide for different functions: They can be the guiding category in *how* to compare distributions such as the ||*absolute frequency*|| *of dots under zero* (see episode 1 below). Furthermore, we found concepts-in-action which guide the students in *why* they choose certain aspects to compare, such as an ||*individual* *representativity of chosen intervals for the specific properties of a distribution*|| (see episode 2 below).

Concepts-in-action are not necessarily in line with normative mathematical ideas but guide the students' individual process of making sense of the situation. They are shown through action (ibid.) and can be uncovered through interpretation of the students' behaviour.

The in-depth analysis is so far limited to the case of Annika and Bastian; preliminary analysis showed that the other pairs are comparable in terms of focussing on visual features of the dot plots and determining absolute frequencies, but were less able to explain their reasoning behind certain actions and communicate their ideas and strategies. In the first step of the analysis of our data, we reconstructed the nature of this case from video, identifying crucial episodes of the students' reasoning process. These scenes were transcribed verbatim and annotated by both researchers separately. The goal of the analysis was to infer a) students' individual concepts-in-action when comparing the given distributions and b) students' underlying reasons for choosing specific foci on the distribution(s). The results of the analysis were then compared and discussed until a consensus on the interpretations was reached.

EMPIRICAL RESULTS

Annika and Bastian finished a lesson on statistics using box plots, but are unfamiliar with the representation of stacked dot plots. In line with other studies, they use exclusively informal methods to compare the presented distributions of temperatures. While they mention that it would be convenient to "have the arithmetic mean" or "use box plots" for "a better comparison", they are not making an effort to generate them. When presented with boxplots in the end of the interview, they have no trouble interpreting and matching them with the dot plots. This activity is rather superficial though and stays on a level of formal procedures rather than connecting insights acquired in the informal analysis of the dot plots. As mentioned above, this is in line with recent research. We will thus focus exclusively on the comparing activities concerning the dot plots.

Throughout the interview Annika und Bastian make use of various concepts-in-action guiding for comparing the distributions, which we present in table 1 (RQ1). The concepts-in-action of ||*density*||, ||*value*||, and ||*spread*|| only appear rarely, as Annika and Bastian mostly focus on the ||*absolute frequency*|| and ||*position*|| of groups of dots. For this, they create so-called "sections" (intervals with groups of dots) within the distribution; this activity is mostly guided by the visually perceived 'hills' (a modal interval), which the students call "agglomeration area" (a geographical term, which might indicate that they also take the density into account). In Fig. 2, we marked the sections, which the students address verbally or by gestures[1]: They first focus on the visual hill in 2002 (see Fig. 2, section 1_{2002}). The number of dots in this interval is then compared with the number of dots in the same interval in 2004 (Fig. 2, section 1_{2004}; episode 1 below gives more details on this comparison activity). Section 2 is defined by the agglomeration area in 2004; section 3 consists of the 'most right dots' (also right of the border of section 2) and section 4 is defined as left of the agglomeration area of 2002. In other situations in the interview, the students also make use of the scale and the context by looking at dots under or around 0°C or "dots in the colder interval".

Concept-in-action	Activity
||Absolute frequencies of dots in certain intervals||	Comparing the (difference of) the amount of data points in chosen intervals
||Density in certain intervals||	Comparing the number of dots in relation to the width of the interval
||Position of certain (intervals of) dots||	Comparing the relative position of data points by 'left of'/'right of' or 'higher'/'lower than'
||Value of certain (intervals of) dots||	Comparing the temperature values of data points
||Spread of dots in certain intervals||	Comparing how spread out the data points are

Table 1: Concepts-in-action activated in comparing activities

The students focus almost exclusively on such groups of data points in a local view. However, the in-depth analysis uncovers that the students have features of the whole distribution in mind when comparing sections, as we will show in the next segment.

Episode 1: Comparing and equalising sections of distributions

25 minutes into the experiment, Annika is summarizing previous arguments of her and Bastian in favour of calling July 2002 the warmer.

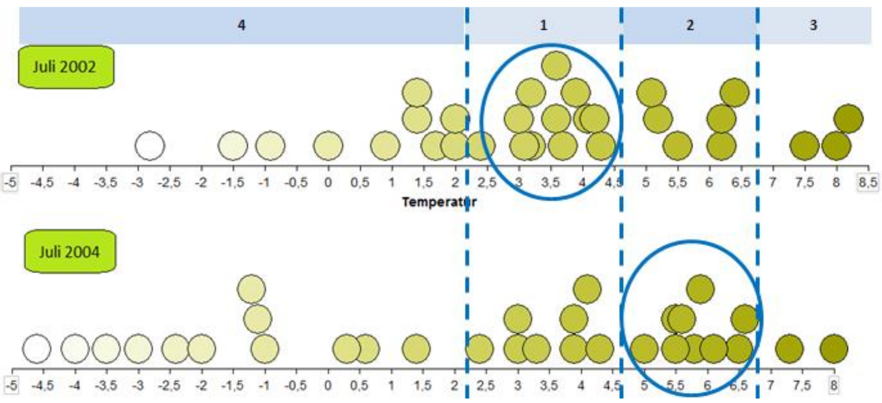

Fig. 2. Annika & Bastian's sections when comparing the distributions

[1] Dotted lines, circles and section numbers added by authors for clarification; circled are the previously identified 'agglomeration areas'; numbers indicate the order in which the sections are addressed by the students.

1	Annika:	If these were 11 dots in the agglomeration area *[circle in section 1_{2002}]*, and here are 8, then *[points to section 1_{2004}]*. Then that's a difference of 3 […]
2	Annika:	But .. Bastian says here *[points to section 2]* is also a difference of 3. So here in the agglomeration area *[in section 2_{2002}]* are 3 fewer and here in this agglomeration area *[in section 1_{2004}]* are also 3 fewer, so to say. Thus, it is even again, Bastian said.

In line 1, Annika compares the ||*differences of absolute frequencies*|| starting with the visual hills (circled in Fig. 2): Each distribution's agglomeration area has three dots more than the corresponding section of the other distribution. Not activating a concept of ||*values*||, Bastian's reasoning as repeated by Annika seems to be that the differences between the distributions thus even out for these two sections (line 2). Therefore, the agglomeration areas alone might not be suitable to determine the warmer month and – according to Annika – other features have to be taken into account:

3a	Annika:	Thus I argued that – up here *[points to section 3_{2002}]* has one dot more; they are higher, too.
3b	[cont'd]	And all these dots *[points to section 4_{2002}]* are much more spread downwards *[points to section 4_{2004}]*. That is why in my opinion 2002 is warmer.

Annika now compares the distributions by the ||*absolute frequency*|| and ||*position*|| *of dots with highest value* (maxima; section 3, line 3a). In line 3b, her gesture pointing at the dots in section 4 of 2002 and saying "all these dots are much more spread downwards [in 2004]" seems to indicate that she assumes the same absolute frequency for the sections. Thus, it is the ||*spread*|| in combination with the ||*position*|| of the dots which makes the difference and lets her back up her argument that 2002 is warmer.

We call Annika's strategy 'equalising': When agglomeration areas are not useful for comparison as they are 'equal' in terms of absolute frequency, the other intervals have to be taken into account. In regard to RQ1, this episode highlights how the students use different concepts-in-action to compare distributions. Concerning RQ2 and the question of underlying rationales, we uncover how Annika seems to choose the outer intervals of the distribution and concepts-in-action other than ||*absolute frequencies*|| because she perceives the agglomeration areas – in this case of 'equal absolute frequencies' – as not helpful for the comparison.

Episode 2: Discussing the width of intervals for comparing

The determination of sections plays a crucial role throughout the students' comparative actions. At first, Annika and Bastian choose them spontaneously according to the perceived visual features of the distributions. When after 36 minutes the third distribution is discussed, they are asked to give a general rule for comparing distributions. Annika explains her strategy as "count how many dots are in the lower section and how many are in the higher section and compare them", stating that it is not useful to look only at the agglomeration areas as was established in episode 1. The interviewer then prompts the students to consider the choice of sections explicitly:

4	Researcher:	And these sections you pick out, where do they come from? […]
5	Bastian:	What you just said about counting the dots *[points at Annika]*; I would definitely use the same distances. That means always 5, so to say: minus 5 to 0, 5 to 8.5 or 10 *[draws imaginary vertical lines through all distributions, indicated with dotted lines in Fig. 3]* […] because you always need equal sections […]
6a	Annika:	Yes... *[5 sec]* I have to think about that again. *[looks at distributions, 27 sec]*
6b	[cont'd]	Well, I'm kind of undecided. On the one hand, I think that it makes sense. On the other hand I think why not take sections of different size? But then I think, then you can't compare it that well with the <u>other</u> sections, but can compare it still with the sections of the other <u>years</u> *[moves hand vertically over the three distributions]*. […]
6c	[cont'd]	Well here, from minus 5 to 0, we simply use the step of 5 now. [The step] to [plus] 5 is not fitting though, because these dots *[points to dots just above 5° in 2007, circled in Fig. 3]* are close to those under 5. […] and that is why it is sometimes better to use different sections, because then points still just belong to it and are not already in another section.

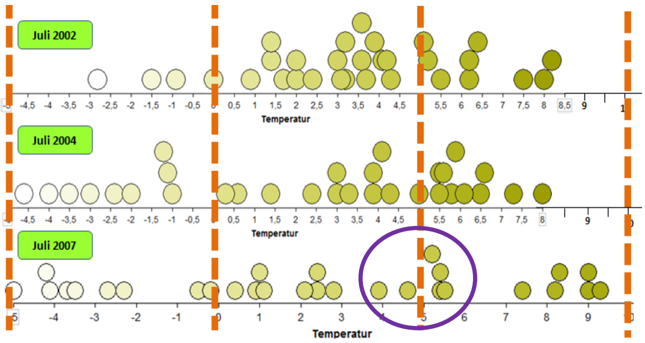

Fig. 3. Bastian's proposed sections

In line 5, Bastian proposes an approach of choosing sections within a distribution due to a fixed width of the interval of 5. Annika however seems torn between the ideas of fixed and dynamic interval width (line 6b). To her, there seem to be instances where interval width can be chosen arbitrarily (from -5 to 0, line 6c), and where interval borders have to respect features of the distribution (from 0 not to 5, line 6c). Her reason "these dots are close to those under 5" might refer to the visual impression of 2007: the group of five dots around 5° (circled in Fig. 3) are separated from others by gaps and thus form a visual unit. We interpret this as an indication of an underlying concepts-in-action: To compare different distributions, one has to take the gaps and groups into account. Thus, sections have to represent the specific visual features of a distribution, which we call an individual concept-in-action of || *representativity of a distribution's features*||.

In line 6b, Annika utters the underlying reason for her conflict: Annika seems to explicitly differentiate between comparing sections *within* a distribution and comparing sections *between* distributions: Sections of different widths (as in Fig. 2) are worse for comparing them within a distribution, but due to their ||*individual representativity*|| better for comparing them with other distributions. This indicates a global view on the distribution which is Annika is dealing with by the informal approach of determining absolute frequencies in sections of different width.

CONCLUSION AND OUTLOOK

Consistent with literature (e.g. Bakker & Gravemeijer 2004), the students organised the data through visual features such as modal clumps. The partition of the data however did not necessarily follow the structure of low-middle-high, but was informed by complex interplay of various concepts-in-action. The empirical snapshots show that these students are mostly focusing on the ||*absolute frequency*|| instead of the position of certain features in relation to each other and values of data points. Elaborate strategies such as 'equalising' combine different concepts-in-action and create individual rules of which features to compare in certain situation. The choice of sections in which absolute frequencies are determined are guided by an individual concept-in-action of ||*representativity*|| which does allow seeing the characteristic properties of one distribution but might be an obstacle to put different distributions in relation with each other.

The episodes shown in this paper are not intended to highlight deficits of the lessons students had for acquiring underlying concepts underlying box plots. Rather, the intention is to give insights into students' hidden individual concepts and reasons in order to understand the rationality of their activities. The presented reasoning processes have a lot of potential for a deep understanding and statistical reasoning, as a rich repertory of (in itself mathematically sustainable) concepts is activated and consciously combined.

Even though this phenomenon was so far discovered in the design experiment with only one pair of students, we take its possible impact on learning pathways seriously: In the next design experiment cycle, we will make room for students to explicitly address the question if and in how far it is necessary to represent a distribution's features for comparisons. We will carefully consider tasks to guide students in the shift from focusing on absolute frequencies to ordinal views of the relative position of features. Furthermore, our further research aims at uncovering other concepts-in-action guiding students in comparing distributions.

REFERENCES

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

Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representation. *Educational Studies in Mathematics Education*, 45, 36–65.

Ben-Zvi, D. (2003). The emergence of reasoning about variability in comparing distributions: A case study of two seventh grade students. In C. Lee (Ed.), *Reasoning about variability: a collection of current research studies* (pp. 42–63). Mount Pleasant, Michigan: Central Michigan University.

Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning*, 1(1), 5–43.

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

Konold, C., Pollatsek, A., Well, A., & Gagnon, A. (1997). Students analyzing data: Research of critical barriers. In J. B. Garfield & G. Burril (Eds.), *Research on the role of technology in teaching and learning statistics: Proceedings of the 1996 IASE Roundtable Conference* (pp. 151–167). Voorburg: International Statistical Institute.

Pfannkuch, M., Regan, Matt, & Wild, C. (2010). Telling data stories: Essential dialogues for comparative reasoning. *Journal of Statistics Education*, 18(1). Retrieved from http://www.amstat.org/publications/jse/v18n1/pfannkuch.pdf.

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

Prediger, S., & Schnell, S. (2014). Investigating the dynamics of stochastic learning processes: A didactical research perspective, its methodological and theoretical framework, illustrated for the case of the short term–long term distinction. In E. J. Chernoff & B. Sriraman (Eds.), *Advances in Mathematics Education. Probabilistic Thinking* (pp. 533-558). Springer Netherlands.

Vergnaud, G. (1996). The theory of conceptual fields. In L. P. Steffe, P. Nesher, P. Cobb, G. A. Goldin, & B. Greer (Eds.), *Theories of Mathematical Learning*. Hove (UK), Psychology Press.

Watson, J. M., & Moritz, J. B. (1999). The beginning of statistical inference: Comparing two Data Sets. *Educational Studies in Mathematics Education*, 37(2), 145-168.

Wild, C. J. (2006). The concept of distribution. *Statistics Education Research Journal*, 5(2), 10–26.