ARTHUR BAKKER AND KOENO P. E. GRAVEMEIJER

AN HISTORICAL PHENOMENOLOGY OF MEAN AND MEDIAN

ABSTRACT. Using Freudenthal's method of historical phenomenology, the history of statistics was investigated as a source of inspiration for instructional design. Based on systematically selected historical examples, hypotheses were formulated about how students could be supported in learning to reason with particular statistical concepts and graphs. Such a historical study helped to distinguish different aspects and levels of understanding of concepts and helped us as instructional designers to look through the eyes of students. In this paper, we focus on an historical phenomenology of mean and median, and give examples of how hypotheses stemming from the historical phenomenology led to the design of instructional activities used for teaching experiments in grades 7 and 8 (12–14-years old).

KEY WORDS: history of statistics, inspiration for instructional activities, preparation phase of design research

1. Historical phenomenology

Various authors have suggested that studying the history of a topic is good preparation for teaching that topic (Fauvel and Van Maanen, 2000; Gulikers and Blom, 2001). The obstacles that people in the past grappled with are interesting to teachers and designers because students often encounter similar obstacles (*cf.* Brousseau, 1997, on epistemological obstacles). However, students also know things that people in the past did not know and therefore need not recapitulate the history of mathematics or statistics to learn about the key concepts (*cf.* Radford, 2000). In Freudenthal's (1991) view, students should be guided such that they can reinvent the concepts at issue using the knowledge they already have. He wrote that ideally,

> The young learner recapitulates the learning process of mankind, though in a modified way. He repeats history not as it actually happened but as it would have happened if people in the past would have known something like what we do know now. It is a revised and improved version of the historical learning process that young learners recapitulate.
>
> 'Ought to recapitulate' – we should say. In fact we have not understood the past well enough to give them this chance to recapitulate it. (Freudenthal, 1983b, p. 1696)

This paper is concerned with the question of what we can learn from a historical study for designing instructional materials in statistics education,

Educational Studies in Mathematics (2006) 62: 149–168
DOI: 10.1007/s10649-006-7099-8

where guided reinvention is taken as a general guideline. To explore the design of "a revised and improved version of the historical learning process," we studied the early history of statistics, in particular of average values, sampling, distribution, and graphs, which were the core ingredients of the intended instructional sequence for students aged 12–14.

The method used was Freudenthal's (1983a) historical phenomenology, a method for studying the relation between phenomena that were of interest historically and the statistical concepts that historically arose for organizing and analyzing such phenomena. In Freudenthal's (1983a) view, mathematical thought objects (e.g. concepts, structures, ideas, methods) serve to organize phenomena, both from daily life and from mathematics itself. A *historical* phenomenology is a study of the historical development of a concept in relation to the phenomena that led to the genesis of that concept. The objective of an historical phenomenology is to identify both the problem situations that created a need to organize certain phenomena, and (precursors to) the concepts that have been invented to get a handle on these phenomena. A *didactical* phenomenology is a study of concepts in relation to phenomena with a didactical interest. The challenge is to find problem situations in which phenomena "beg to be organised" (Freudenthal, 1983a, p. 32) by the thought objects that students, in our view as educators, need to develop.

By carrying out an historical phenomenology of statistical key concepts and graphs, we expected to be able to identify potentially productive activities around concepts that nowadays seem to be long-established and not open to question. We also took the stance that studying history can help us see certain phenomena through the eyes of people who did not have the same concepts and techniques as we have nowadays.

We first collected as many historical phenomena with a statistical flavor as possible, mainly about average values, sampling, distribution, and graphs that are commonly used in statistics. The next step, the selection of historical examples, was guided by the educational potential we saw in them; we only selected examples that could be regarded as precursors to statistical concepts with possible relevance for the design of instructional materials. For example, an estimation of the number of years between the first and last Egyptian king reached by some recognizably systematic method that could be interpreted as an intuitive version of an average was included in our historical phenomenology, whereas a simple guess of a large number would not be included.

The present paper is dedicated to an historical phenomenology of the mean and median only (sampling, distribution and graphs will be the topic of another paper). The following two sections present historical examples of a particular use of a concept and hypotheses about how students' reasoning with that concept could be supported. These hypotheses are part of what Simon (1995) calls hypothetical learning trajectories, which involve expectations of what the mental activities of students will be when they will participate in the envisioned instructional activities. Some of these hypotheses were tested in the design research of Bakker (2004), by analyzing whether the suggested instructional activities actually fostered the presumed forms of thinking and reasoning (see also Gravemeijer, 2002). The hypotheses that proved productive in that research are indicated with an asterisk and are illustrated with an example in Section 4. Last, Section 5 is a discussion of the process of generating those hypotheses and using those for instructional design.

2. THE MEAN

The term "average value" typically refers to several different measures of center: arithmetic mean, median, mode, midrange, and in this paper we also use it for precursors to these. Because we study the early history of statistics as a source of inspiration for instruction to young students, we do not make more modern technical distinctions between, for example, the mean of a population and the sample mean to estimate the center of a distribution.

2.1. *Average values to estimate a total*

The oldest historical examples that we considered as relevant for the historical phenomenology all had to do with estimation of large numbers. Two of them are presented below to illustrate early stages of the mean.

2.1.1. *Example 1. Number of leaves on a branch*

In an ancient Indian story, which was written down in the fourth century AD, the protagonist Rtuparna estimated the number of leaves and fruit on two great branches of a spreading tree (Hacking, 1975). He estimated the number of leaves and fruit on the basis of one single twig, which he multiplied by the number of twigs on the branches. He estimated 2095, which after a night of counting turned out to be very close to the real number. Although it is uncertain in the story how Rtuparna chose the twig, it seems that he must have chosen an "average-sized" twig, since that would lead to a good estimation.

The educational potential we saw in this example was that such an implicit use of a representative value could be an intuitive precursor of the arithmetic mean, because the twig one needs to choose has to represent all other twigs. In connection with its role in estimating the total number, the choice of the average twig thus involves notions such as representativeness and compensation. In modern terminology, the representative value a had to be chosen such that $n \times a$ is the sum of all x_i, with n being the number of twigs and x_i the numbers of leaves on each twig. Even if Rtuparna did not exactly use the method we think he used, the historical problem situation inspired our instructional design to let students reinvent such a method (see Section 4).

2.1.2. *Example 2. Crew size on ships*

Rubin (1971) has found old examples of statistical reasoning in the work of Thucydides (*circa* 460–400 BC), one of the first scientific historians. The reader is invited to decide how he or she would translate this excerpt from his *History of the Peloponnesian War* into modern statistical terms.

> Homer gives the number of ships as 1,200 and says that the crew of each Boetian ship numbered 120, and the crews of Philoctetes were fifty men for each ship. By this, I imagine, he means to express the maximum and minimum of the various ships' companies ... If, therefore, we reckon the number by taking an average of the biggest and smallest ships ... (Rubin, 1971, p. 53)

In this example, Thucydides possibly interpreted the given numbers as the extreme values, so that the total amount of men on the ships could be estimated by taking the middle of these two extremes, which is called the *midrange* – defined as the arithmetic mean of the two extremes. This technique of using the midrange instead of the mean can be justified if certain assumptions are fulfilled, for instance that the underlying distribution is approximately symmetrical.

The first instructional idea that arose from these historical examples was the following.

*H1. Estimation of large numbers can challenge students to develop and use intuitive notions of mean.**

2.2. *Mean values in Greek geometry*

Explicit use of mean values and names for these values are found in ancient Greek mathematics. In Pythagoras' time, around 500 BC, three mean values were known: the harmonic, geometric, and arithmetic mean (e.g. Heath, 1981). Though these mean values were not used in a statistical way, we can still learn from the Greek definition of the arithmetic mean, which reads as follows: the middle number b of a and c is called the arithmetic mean if and only if $a - b = b - c$. Note that this definition refers to only two values, and also differs in formulation from the equivalent modern one, $(a + c)/2$. The Greek version locates the mean in the middle of the two

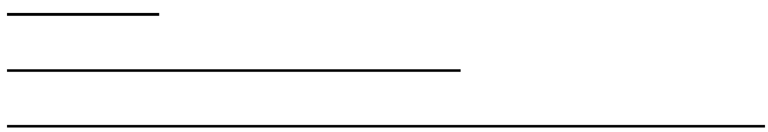

Figure 1. Greek representation of magnitudes as bars (2, 6, and 10).

values and is difficult to generalize, whereas the modern version emphasizes the calculation and is easy to generalize. With instructional design in mind, we also observe that for the Greeks magnitudes were represented by lines (Euclid, 1956) as in Figure 1.

Such a representation might help students realize that the part of the longest bar that "sticks out" (compared with the middle bar) compensates the corresponding part of the shortest bar.

*H2. The Greek bar representation may help students realize that the mean is in-between extreme values (intermediacy) and may scaffold a compensation strategy of visually estimating the mean.**

2.3. *Average, midrange, and generalization of the mean*

Apart from estimation and geometrical definitions, the concept of average also has other origins. For instance, the term "average" itself stems from maritime law, in connection with insurance and the fair share of profit and loss. The Oxford English Dictionary (Simpson and Weiner, 1989), under the heading of "average," states that one of the meanings of average in maritime law is "the equitable distribution of expense or loss, when of general incidence, among all the parties interested, in proportion to their several interests." More generally, the average came to mean the distribution of the aggregate inequalities (in quantity, quality, intensity, etc.) of a series of things among all the members of the series, so as to equalize them. In its transferred use the term average thus came to signify the arithmetic mean.

H3. Learning the mathematics that is involved in fair share and insurance (e.g. proportions and ratios) is good preparation for learning about the arithmetic mean. Fair share is also a suitable context to practice such skills (cf. Cortina, 2002; Mokros and Russell, 1995; Strauss and Bichler, 1988).

In line with the example on estimating the number of people on ships (example 2 in Section 2.1.2), another possible precursor to the arithmetic mean of more than two values is the *midrange*, which was used for example in Arabian astronomy of the ninth to eleventh centuries, and also in metallurgy and navigation (Eisenhart, 1974). Taking into account that we tend to presume implicitly a symmetrical distribution of errors, it is understandable that the midrange was used in those situations. Because the midrange was probably a precursor to the mean, we may also expect students to use the midrange in situations where a mean is more appropriate from a modern perspective.

*H4. Students may use the midrange as a precursor to more advanced notions of average.**

Not until the sixteenth century was it recognized that the arithmetic mean could be generalized to more than two cases: $a = (x_1 + x_2 + \cdots + x_n)/n$. Székely (1997) supposes that Simon Stevin's contribution to the decimal system in 1585 facilitated such division calculations. This generalized mean proved useful for astronomers who wanted to know a "true" value, such as the position of a planet or the diameter of the moon. Using the mean of several measured values, scientists assumed that the errors added up to a relatively small number when compared to the total of all measured values. This method of taking the mean for reducing observation errors was mainly developed in astronomy, first by Tycho Brahe (Plackett, 1970). From the late sixteenth century onwards, using the arithmetic mean to reduce errors gradually became a common method in other areas as well (Eisenhart, 1974). This implies for instructional design:

H5. Repeated measurement may be a useful instructional activity for developing understanding of the mean. (cf. Konold and Pollatsek, 2002; Petrosino et al., 2003)

2.4. *The mean as an entity in and of itself*

The historical examples up to about the nineteenth century mostly have to do with approximating a "true" or best value on the basis of repeated measurements. Thus in those examples, the mean was used as a means to an end. It took a long time before the notion of a representative value as an entity in and of itself emerged. In the nineteenth century, Quetelet, famous as the inventor of *l'homme moyen*, the average man, was one of the first scientists to use the mean as the representative value for an aspect of a population. This transition from the real value to a representative value as a statistical construct was an important conceptual change (Porter, 1986; Stigler, 1986). One of the problems scientists had with this use of the mean, as transferred from its use in astronomy, was the distinction between discrete and continuous values. Peirce (1958) wrote the following about this issue:

> In studies of numbers, the idea of continuity is so indispensable, that it is perpetually introduced even where there is no continuity in fact, as where we say that there are in the United States 10.7 inhabitants per square mile, or that in New York 14.72 persons live in the average house. [Footnote:] This mode of thought is so familiarly associated with all exact numerical consideration, that the phrase appropriate to it is imitated by shallow writers in order to produce the appearance of exactitude where none exists. Certain newspapers, which affect a learned tone, talk of "the average man," when they simply mean *most men*, and have no idea of striking an average. (CP, Vol. 2, Section 646)

Note that the examples Peirce gives, such as 10.7 inhabitants per square mile, imply what Cortina (2002) calls a "mean-as-a-measure," with which one superimposes an assumption of linearity on reality, and in doing so consciously ignores variation.

What is relevant for the historical phenomenology and our hypothetical learning trajectories is that there are several stages in understanding the mean as a representative value.

*H6. Using an average value in estimations of large numbers and using the mean for reduction of errors are probably easier for students than understanding the mean as an entity in and of itself, that is as a representative value for an aspect of a population. Estimation tasks should therefore be done before activities that involve the mean as an entity in and of itself.**

3. The median

We had three reasons to investigate the history of the median more carefully than we had done prior to the teaching experiments in grade 7 (1999–2000, comprising 12–15 lessons). First, although we had used instructional activities aimed at using the median as a representative value, the Dutch students we worked with did not view the median as such. This is in line with the initial problems Cobb et al. (2003) report about a group of American students. In their long-term teaching experiments (37 lessons in grade 7 and 41 in grade 8), the median eventually did come to the fore as a means of describing distributions in the form of an indicator of where the density of the data is the highest in a unimodal distribution. Moreover, the median and quartiles proved useful as a means of structuring data when comparing two data sets.

Second, during the instructional design phase it turned out difficult to find phenomena that "beg to be organised" (Freudenthal, 1983a, p. 32) by the median as a measure of center or a representative value. The historical phenomenology of the median presented in this section was thus intended to help us understand better why the median is difficult for students and to find phenomena that require using the median. Third, we had found very few historical studies of the median (Godard and Crépel, 1999; Harter, 1977; Monjardet, 1991), perhaps because its history is mostly buried under larger issues such as the normal distribution, Bayes' theorem, the central limit theorem, and the method of least squares.

We organize this section by focusing on the phenomena and contexts in which the median arose historically. It turned out that the median mostly emerged as an alternative to the mean.

3.1. *Theory of error*

The most important context in which average values were used was the theory of error (Sheynin, 1996). The problem at first was to find the assumed true value from the available observations, later to find the best estimate of such a value. The Greeks often used a value that fitted the theory instead of real observations to "save the phenomena," as they called it (Pannekoek, 1961). Another method was to choose a value that seemed reliable, for example from a middle cluster or from values measured under favorable conditions (Steinbring, 1980).

The first possible instance of the *median* that Eisenhart (1974) has found was in a book by Edward Wright of 1599 on navigation. Wright wrote about the determination of location with a compass:

> Exact truth amongst the inconstant waves of the sea is to be looked for, though good instruments be newer so well applied. Yet with heedful diligence we come so near the truth as the nature of the sea, our sight and instruments will suffer us. Neither if there be disagreement betwixt observations, are they all by and by to be rejected; but as when many arrows are shot at a mark, and the mark afterwards away, he may be thought to work according to reason, who to find out the place where the mark stood, shall seek out the middle place amongst all the arrows: so amongst many different observations, the middlemost is likest to come nearest the truth. (Eisenhart, 1974, p. 52, spelling modernized)

It is not certain that Wright really meant the median, since he gave no numerical examples. Eisenhart argued that since Wright wrote "Neither ... are they all by and by to be rejected" it is possible that he recommended the middle-most observation, the median, and not the middle place, the midrange, since then most observations would not be used.

A clearer example of the median, in the context of measurement errors, is found much later in the work of Boscovich (around 1755). The interesting point of his work for the history of the median was the set of conditions he proposed in the search for true values, in particular a line of best fit through observations (Stigler, 1984). One of these conditions was that the sum of absolute errors should be minimal; in our notation: $\Sigma|x_i - a|$ is minimal. This condition turns out to be equivalent to the statistical median (David, 1998; Eisenhart, 1977). Note that the condition that the errors should add up to zero is equivalent with the arithmetic mean: $\Sigma(x_i - a) = 0 \Leftrightarrow \Sigma x_i/n = a$. In fact, historical discussions on the mean in error theory led to the development of the concept of distribution (Steinbring, 1980).

H7. If the theory of errors (e.g. repeated measurements) is taken as a context for developing statistical ideas of center and distribution, it may be advantageous to let students formulate their own intuitions about errors: Do the errors add up to zero? Is the chance that the measurement is too small equal to the chance that they are too large? Do errors occur symmetrically?

3.2. *Probability*

Another context in which the median arose as a counterpart of the mean was probability theory. We give two examples of how the development of the median was connected to probability. The first is a paradigmatic example of how an intuition of something, in this case a middle value, became differentiated into two concepts: median and mean lifetime. The second example stems from Legendre and Laplace, who distinguished two possibilities of finding a true value, one of which we now call the median.

3.2.1. *Example 1. Mean and median lifetime*

In 1669, the Dutch brothers Chistiaan and Lodewijk Huygens had an informal correspondence about their father's life expectancy. They used the mortality tables in the famous *Bills of Mortality* by Graunt (1662), but then disagreed about certain calculations (Véron and Rohrbasser, 2000). It was Christiaan who realized that there was a difference between expected remaining lifetime and the lifetime that half of the people would reach. On November 28, 1669, he wrote to Lodewijk:

> There are thus two different concepts: the expectation or the value of the future age of a person, and the age at which he has an equal chance to survive or not. The first is for the calculation of life annuities, and the other for wagering. (Huygens, 1895, Volume 6, letter to Lodewijk Huygens; translation from French: Hald, 1990, p. 106)

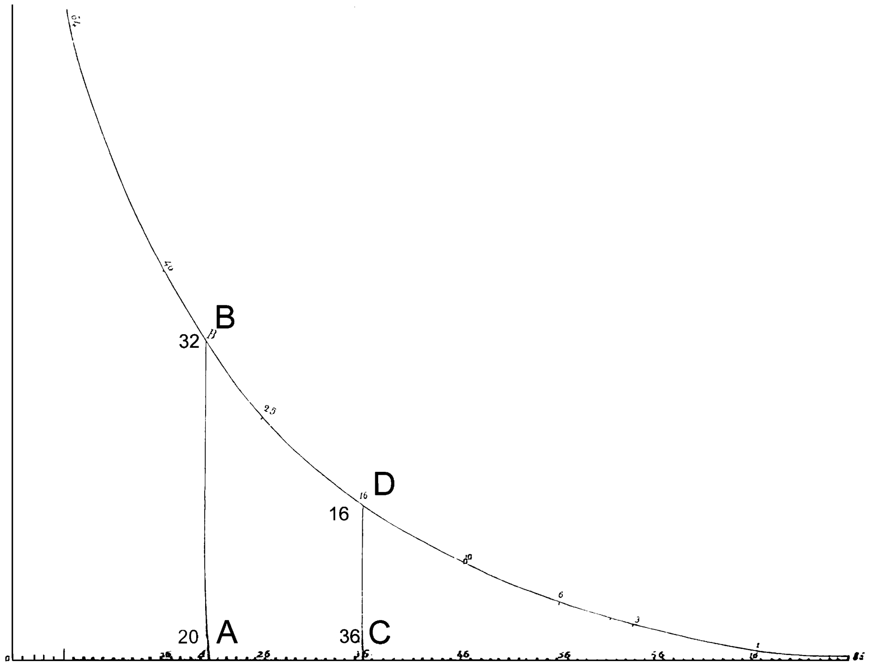

Figure 2. Huygens' theoretical line graph of mortality data (Huygens, 1895, between page 530 and 531). The letters and numbers in the original are enlarged for readability.

Christiaan made a graph from which we can read the median lifetime. In Figure 2 we can see that a 20-year-old person (A) had a median lifetime of 36 years: take the half of AB and find CD further in the graph. The French terms that Christiaan used for what we now call median life time were *apparence* (likeliness) and *vie probable* (probable life), since the person has equal chance to survive to this age or not. The chance of a half appears a natural point to look at, though it was not very useful except for chance-like problems such as wagering. More useful was what Christiaan called *espérance* and what we now call mean or expected lifetime. In terms of the historical phenomenology, the phenomenon of predictions about life times asked for a distinction between mean and median lifetime due to skewed distribution of age (*cf.* Stamhuis, 1996).

3.2.2. *Example 2. Milieu de probabilité*
The second example of the connection between the median and probability stems from Legendre (1805) and Laplace (1812/1891). They and their contemporaries used the term *milieu de probabilité*, the middle of the probability, which is a suggestive name for the median in the context of

probability functions. Cournot (1843/1984) was the first to use the term "median" (*valeur médiane*) for this value (David, 1995, 1998; Stigler, 1986). He defined the median as the value x_0 for which the distribution function F satisfied $F(x_0) = 1/2$ and he explained that it is the value for which the area under the graph is the same on the left and on the right. Furthermore, he wrote:

> Two players, one betting of the value smaller than x and the other larger than x, would bet with the same chances. With a very big number, the quotient of the larger (or smaller) values than x and the total number of values will not differ much from the fraction 1/2. (Cournot, 1843, p. 83; translation from French)

*H8. To support students' understanding of the median it is helpful to let them visually estimate the median in a dot plot (Figure 5) and look for which value the areas on the left and right are the same.**

The key point of this section is that the median is closely related to probability theory, especially with the chance of a half:

H9. To support students' understanding of the median it is worth designing problem situations in which it is reasonable to compare halves, for instance in chance situations.

3.3. *Statistical reasons to use the median*

Throughout the nineteenth century, scientists had different reasons to use the median as an alternative to the mean. In 1874, Fechner used the median (*Centralwerth,* today spelled as *Zentralwert)*, in an attempt to describe many sociological and psychological phenomena with methods that had proven to be useful in astronomy. He advocated the *ease of calculation* of the median, but he also had more theoretical reasons for using other measures of center than the mean, which we address later in this section.

Galton used the English term "median" for the first time in 1882 (David, 1995) and caused the breakthrough of the concept (Godard and Crépel, 1999). Important reasons to use the median were its ease of calculation and its *intuitive clarity* (Stigler, 1986). Most phenomena Galton (1889) studied were roughly symmetrical, so the median would not differ much from the mean, which is laborious to calculate. Throughout his book *Natural Inheritance* he therefore used the median M and quartile distance Q, with $Q = 1/2(Q_3 - Q_1)$, Q_1 being the first and Q_3 the third quartile (in the modern terminology). He rarely mentioned the mean and modulus ($\sqrt{2}$ times the standard deviation), probably to reduce calculation efforts and not scare away scientists without the necessary statistical background. Yet another reason for using the median could have played a role. Galton studied variables such as intelligence and reputation that he measured in an *ordinal way* (in which case the mean cannot be calculated).

Edgeworth, a younger contemporary of Galton, preferred the median to the mean because of its insensitivity to outliers; probably this was due to his interest in economics, which has less regular data than is common in astronomy for instance, and he was not the only one to prefer the median (Harter, 1977). Nowadays, the median's resistance to outliers is one of the major reasons to use it, especially when the data are irregular as is common in social sciences and economics.

H10. Using irregular data with outliers can motivate students to reason with the median instead of the mean as a measure of center, provided they already know about outliers and measures of center.

For a long time, distributions of error were assumed to be symmetrical. In 1838, Bessel was probably the first to doubt the assumption of symmetry (Kotz and Johnson, 1981; Steinbring, 1980). In contrast to most other scientists of his time, Fechner even assumed that *most* distributions of data were asymmetric. It turned out that the median minimizes the sum of absolute deviations to the first power and the mean the sum of deviations to the second power. Consequently, both measures of center are special cases of Fechner's *Potenz-mittelwerthen*, the values that minimize the sum of the deviations to the n-th power: $\Sigma|x_i - a|^n$ is minimal. Fechner used these generalized measures of center to describe the *skewness* of distributions. Edgeworth also used the difference of mean and median, divided by a normalizing factor, and this measure is still used today as an indication of skewness (Kotz and Johnson, 1981; Stigler, 1986). We therefore conjecture:

*H11. Skewed distributions can be used to make the usefulness of the median a topic of discussion.**

Despite the efforts of scientists such as Cournot, Fechner, Galton, and Edgeworth, the median was mostly neglected and the mean favored in the nineteenth century. Nowadays, the median is used in order statistics, since the mean cannot be used for ordinal data. The median is also used in robust statistics, since it is far more robust than the mean (Box was the first, in 1953, to use the term 'robust' for insensitivity to outliers – David, 1995). Since statistics is applied in more and more areas with irregular data, the median has become more popular (Portney and Koenker, 1997).

4. Translation to Instructional Design

The historical phenomenology presented in the previous sections offers various elements of a didactical phenomenology of mean and median. These are: the implicit use of the mean to estimate large numbers, the strategy of calculating the mean as a means to eliminate measurement errors, the change of perspective by which the mean becomes an entity in and of itself, the role of mean-as-a-measure, and the difference between the natural sciences (where there seems to be an implicit expectation of symmetrical distributions) and the social sciences (where the apparent skewed distributions create the need to describe those distributions, which creates a role for the median). Using those elements to construe such a didactical phenomenology, however, is a research task in itself that surpasses the scope of this paper.

Instead we exemplify how the historical phenomenology inspired the instructional design in the preparation phase of Bakker's (2004) design research by giving a few examples of instructional activities that were inspired by hypotheses H1, H2, H4, H8, and H11 mentioned in previous sections. The other hypotheses are discussed in the Section 5.

In line with H1 (on the estimation of large numbers), we used estimation tasks as the starting point of the statistics unit in grade 7 to support a process of guided reinvention of average (and sampling). For example, in the first lesson, we asked seventh graders to estimate the size of a large herd of elephants in a picture. Many of them made a grid (Figure 3), counted what

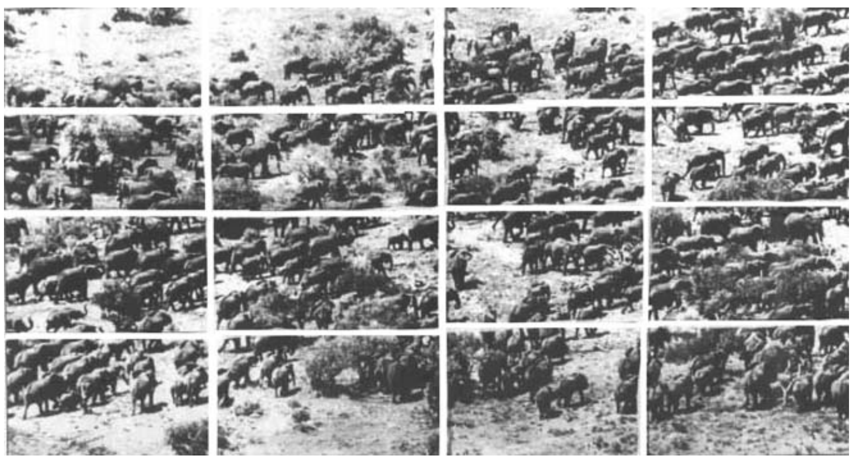

Figure 3. Many elephants in a picture that a student divided into 16 boxes. (Picture reprinted with permission from Britannica.)

they called an "average box," and multiplied that average number by the number of boxes in the grid. As mentioned before, the hypotheses are part of a hypothetical learning trajectory, and thus offer us tools to influence the students' thinking. What we see as valuable in the estimation task is that students have to take into account both the individual values and the way they are distributed in order to choose the appropriate value with which to calculate the total. This implies that with such activities we – as instructional designers – can create a need for the students to reason about the "average value" in a qualitative manner. We did so by making the choice of the "average box" a topic of discussion. It turned out that students had different informal views of average including mode, mean, and midrange.

This supported H4 (on the midrange). For example, several students explained their "average box" as the mean of the emptiest and the fullest box of the grid. One student argued that if one box contained 1 and all remaining boxes 100, 50 would not be a good average. In this way, the midrange was dismissed as a strategy for finding the right total number. In the first and third lessons, we also used skewed distributions such as in Table I to challenge students' midrange strategy. The teacher said, "Something else is being counted. Assume we know the numbers in each box, what would have been an average box?" Similarly, there are indications that such skewed distributions can be used to create a need for a distinction between mean and median (H11).

In this manner, implicit notions of compensation and representativeness were developed further. In addition, we exploited the graphical representation that was mentioned in hypothesis H2 (on the bar representation and compensation strategy of visually estimating the mean.) It turned out that students could easily interpret bars as representations of data values such as in a value-bar graph (Figure 4), especially if the variable at issue has a time dimension (life span) or a one-dimensional physical connotation (e.g., wing span, height, braking distance). In line with the research by Cobb et al. (2003), we introduced students to value-bar graphs, which were available in the first applet of a series of three so-called "Minitools" (Bakker and Gravemeijer, 2004) that we used during the teaching experiments.

TABLE I
Assume we know the numbers in each box, what would have been an average box?

35	58	91
93	83	89
98	97	68
76	82	11

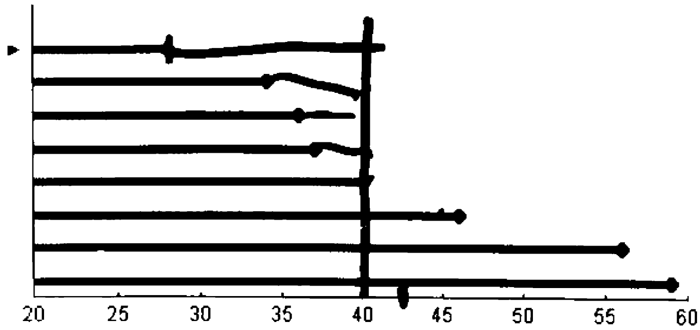

Figure 4. A seventh grader's scribbles to explain his compensation strategy.

The bar representation of data values indeed helped seventh-grade students estimate "average boxes" from data sets by using a compensation strategy. Figure 4 shows a seventh-grader's scribbles to explain his compensation strategy, during the second lesson. He started with 40 as the mean, but discovered that 42 would have been a better estimate. What we think is important here is the implicit notion of balancing and compensating. A key element of the hypotheses is, as we mentioned before, the support these offer in anticipating student thinking, as we will illustrate with one other example.

The second applet we used during the teaching experiments was Minitool 2 (Figure 5). It provides a dot plot in which the data values can be organized in different ways such as into two equal groups (median), four equal groups (precursor to the box plot), and fixed interval widths (precursor to the histogram). When students in grade 8 after ten lessons had to estimate the median in a sketch of a right-skewed distribution (Figure 6),

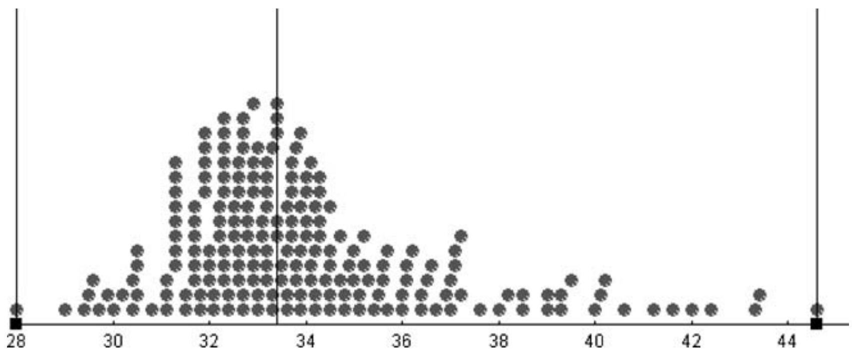

Figure 5. Dot plot in Minitool 2 with the data grouped into "two equal groups".

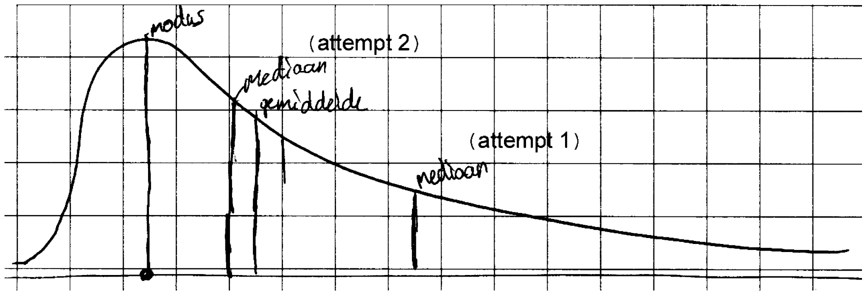

Figure 6. Sketch in which eighth-grade students estimated the median (mediaan = median, modus = mode, and gemiddelde = mean).

they initially mistook the median for the midrange (attempt 1), but by thinking back of the two-equal-groups option in the dot plot they had used (Figure 5), several were able to make an accurate estimate of the median by imagining dots under the continuous sketch (attempt 2 in Figure 6). This supports H8 that one way to visually estimate the median in a continuous sketch or in a dot plot is to look for which value the areas or the number of dots on the left and right are the same.

5. Discussion

In this paper, we have focused on an historical phenomenology of mean and median, and have supplemented it with elements of a didactical phenomenology. What we want to highlight here is that the hypotheses formulated during the historical phenomenology can be interpreted as elements of hypothetical learning trajectories that can be tested by analyzing how students reasoned during the instructional activities that were inspired by the historical phenomenology.

The ideas that were most fruitful for education were H1, H2, H4, H8 and H11, as exemplified in Section 4. Of course, similar instructional ideas can originate in other sources than an historical phenomenology. For instance, others have used estimation tasks as well (e.g. Perry and Kader, 1998). Yet we think that the historical phenomenology provides extra reasons for why estimation tasks as presented in the previous section may lead to conceptual understanding of the mean in relation to distribution and can also offer hints for ordering for didactical purposes the conceptual aspects that need to be addressed.

Some of the hypotheses presented in Sections 2 and 3 were not actually used for instructional design. For example, one reason not to use H3 on fair share was that we decided to focus on the conceptual aspect of a measure of center being a representative value and not on the computational aspects of the mean. In the case of H5, we have not tried out the idea of using repeated measurements ourselves, but support for this hypothesis can be found in Konold and Pollatsek (2002) and Petrosino et al. (2003). H6 (using the mean for estimating is easier than as an aspect of a population) did not directly lead to the design of an instructional activity but played an important role in the hypothetical learning trajectories for two reasons. First, it supported the ordering of activities: we successfully started the seventh-grade sequence with estimation of large numbers. Second, during preliminary interviews it indeed turned out that most students found it hard to answer questions such as: What does it mean that an average family consists of 2.5 people? Many thought, "the point 5 was a child." The results of the interviews thus supported H6. We have not used H7 on distribution of errors in our teaching experiments mainly because we considered these questions too abstract for the age group we were working with. Yet we think it could be useful for older students.

We had no opportunity to test H7, H9, and H10 concerning the median, i.e. to design instructional activities based on those hypotheses and to judge if they supported the learning about the median. It seems clear that the median is important in probability contexts and phenomena that lead to non-normal distributions such as is often the case in social sciences. More research is required to better understand how students can develop an understanding of the median and we hope Section 3 provides other researchers with fruitful starting points.

In our experience, developing an historical phenomenology has proved a useful source of inspiration for instructional design and it helps to distinguish problem situations, related concepts and precursors to particular concepts. It can be demanding for instructional designers and teachers to put aside their knowledge of mathematical and statistical concepts (as historical "end products") and take a student perspective: what may seem a minor step might have taken centuries to develop historically and might also be difficult to develop for students. In other words, an historical phenomenology can help us as designers look through the eyes of the students.

ACKNOWLEDGMENTS

The research was funded by the Netherlands Organization for Scientific Research under grant number 575-36-003B. We thank Paul Cobb, Cliff Konold, and Stephen Stigler for their helpful comments on earlier versions of this manuscript, and Phillip Kent for his editing help.

References

Bakker, A.: 2004, *Design Research in Statistics Education: On Symbolizing and Computer Tools*, CD Beta Press, Utrecht, The Netherlands.

Bakker, A. and Gravemeijer, K.P.E.: 2004, 'Learning to reason about distribution', in D. Ben-Zvi and J. Garfield (eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*, Kluwer, Dordrecht, The Netherlands, pp. 147–168.

Brousseau, G.: 1997, 'Theory of didactical situations in mathematics', in N. Balacheff, M. Cooper, R. Sutherland, and V. Warfield (eds. and trans.), Kluwer, Dordrecht, The Netherlands.

Cobb, P., McClain, K. and Gravemeijer, K.P.E.: 2003, 'Learning about statistical covariation', *Cognition and Instruction* 21, 1–78.

Cortina, J.L.: 2002, 'Developing instructional conjectures about how to support students' understanding of the arithmetic mean as a ratio', in B. Phillips (ed.), *Proceedings of the Sixth International Conference of Teaching Statistics* [CD-ROM], International Statistical Institute, Voorburg, The Netherlands.

Cournot, A.A.: 1843/1984, *Exposition de la théorie des chances et des probabilités [Treatise of the theory of chances and probabilities*], Librairie philosophique J. Vrin, Paris.

David, H.A.: 1995, 'First (?) occurences of common terms in mathematical statistics', *The American Statistician* 49, 121–133.

David, H.A.: 1998, 'First (?) occurences of common terms in probability and statistics – A second list, with corrections', *The American Statistician* 52, 36–40.

Eisenhart, C.: 1974, 'The development of the concept of the best mean of a set of measurements from antiquity to the present day', 1971 ASA Presidential Address. Unpublished manuscript.

Eisenhart, C.: 1977, 'Boscovich and the combination of observations', in M.G. Kendall and R.L. Plackett (eds.), *Studies in the History of Statistics and Probability*, Vol. 2, Charles Griffin, London.

Euclid: 1956, The thirteen books of The Elements. Translation with introduction and commentary by Sir Thomas L. Heath (T. H. Heath, trans.). New York: Dover.

Fauvel, J. and Van Maanen, J. (Eds.).: 2000, *History in Mathematics Education*. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Freudenthal, H.: 1983a, *Didactical Phenomenology of Mathematical Structures*. Dordrecht, the Netherlands: Reidel.

Freudenthal, H.: 1983b, 'The implicit philosophy of mathematics: History and education', in *Proceedings of the International Congress of Mathematicians*. Warsaw, pp. 1695–1709.

Freudenthal, H.: 1991, *Revisiting Mathematics Education: China Lectures*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Galton, F.: 1889, *Natural Inheritance*. Macmillan, London.

Godard, R. and Crépel, P.: 1999, 'An historical study of the median', in *Proceedings of the Canadian Society for History and Philosophy of Mathematics*, Canadian Mathematical Society: Ottawa, pp. 207–218.

Gravemeijer, K.P.E.: 2002, 'Emergent modeling as the basis for an instructional sequence on data analysis', in B. Phillips (ed.), *Developing a Statistically Literate Society. Proceedings of the International Conference on Teaching Statistics [CD-ROM]*. Cape Town, South-Africa, July 7–12, 2002.

Gulikers, I. and Blom, K.: 2001, "An historical angle', a survey of recent literature on the use and value of history in geometrical education', *Educational Studies in Mathematics* 47, 223–258.

Hacking, I.: 1975, *The Emergence of Probability. A Philosophical Study of Early Ideas About Probability, Induction and Statistical Inference*. Cambridge University Press, London.

Hald, A.: 1990, *A History of Probability and Statistics and their Applications before 1750*, John Wiley and Sons, New York.

Harter, H.L.: 1977, *A Chronological Annotated Bibliography on Order Statistics* (Vol. 1): US Air Force.

Heath, T.H.: 1981, *A History of Greek Mathematics*. Dover, New York.

Huygens, C.: 1895, *Oeuvres Complètes de Christiaan Huygens* (Vol. VI). Den Haag, the Netherlands: Nijhoff.

Konold, C. and Pollatsek, A.: 2002, 'Data analysis as the search for signals in noisy processes', *Journal for Research in Mathematics Education* 33, 259–289.

Kotz, S. and Johnson, N.L. (eds.): 1981, *Encyclopedia of Statistical Sciences*, Wiley, New York.

Laplace, P.S.: 1812/1891, *Théorie analytique des probabilités*. Oeuvres complètes Vol. 7. Gauthier-Villars, Paris.

Legendre, A.M.: 1805, *Nouvelles méthodes pour la détermination des orbites des comètes*, Courcier, Paris.

Mokros, J. and Russell, S.J.: 1995, 'Children's concepts of average and representativeness', *Journal for Research in Mathematics Education* 26, 20–39.

Monjardet, B.: 1991, 'Élements pour une histoire de la médiane métrique', in J. Feldman, G. Lagneau and B. Matalon (eds.), *Moyenne, Milieu, Centre: Histoire et Usages*, Ecole des Hautes études en Sciences Sociales, Paris, pp. 45–62.

Pannekoek, A.: 1961, *A History of Astronomy*, Allen and Unwin, London.

Peirce, C.S.: 1958, *CP, Collected Papers of Charles Sanders Peirce*. Harvard University Press, Cambridge, MA.

Perry, M. and Kader, G.: 1998, 'Counting penguins', *Mathematics Teacher* 91(2), 110–116.

Petrosino, A.J., Lehrer, R. and Schauble, L.: 2003, 'Structuring error and experimental variation as distribution in the fourth grade', *Mathematical Thinking and Learning* 5, 131–156.

Plackett, R.L.: 1970, 'The principle of the arithmetic mean', in E. Pearson and M.G. Kendall (eds.), *Studies in the History of Statistics and Probability*, Vol. 1, Griffin, London.

Porter, T.M.: 1986, *The Rise of Statistical Thinking, 1820–1900*. Princeton University Press, Princeton.

Portney, S. and Koenker, R.: 1997, 'The Gaussian hare and the Laplacian tortoise: Computability of squared-error versus absolute-error estimators', *Statistical Science* 12, 279–300.

Radford, L.: 2000, 'Historical formation and student understanding of mathematics', in J. Fauvel and J. van Maanen (eds.), *History in Mathematics Education: The ICMI Study*, Kluwer Academic Publishers, Dordrecht, the Netherlands.

Rubin, E.: 1971, 'Quantitative commentary on Thucydides', *The American Statistician* 1971 (December), 52–54.

Sheynin, O.: 1996, *The History of the Theory of Errors*, Verlag Dr. Markus Hänsel-Hohenhausen, Egelsbach, Germany.

Simon, M.A.: 1995, 'Reconstructing mathematics pedagogy from a constructivist perspective', *Journal for Research in Mathematics Education* 26, 114–145.

Simpson, J.A. and Weiner, E.S.C. (eds.).: 1989, *The Oxford English Dictionary*, 2nd edn., Vol. 1, Clarendon Press, Oxford.

Stamhuis, I.H.: 1996, 'Christiaan Huygens correspondeert met zijn broer over levensduur [Christiaan Huygens corresponds with his brother about life span]', *De Zeventiende Eeuw* 12(1), 161–170.

Steinbring, H.: 1980, Zur Entwicklung des Wahrscheinlichkeitsbegriffs – Das Anwendungsproblem in der Wahrscheinlichkeitstheorie aus didaktischer Sicht. Bielefeld: Institut für Didaktik der Mathematik der Universität Bielefeld.

Stigler, S.M.: 1984, 'Boscovich, Simpson and a 1760 manuscript on fitting a linear relation', *Biometrika* 71(3), 615–620.

Stigler, S.M.: 1986, *The History of Statistics. The Measurement of Uncertainty Before 1900*, Harvard University Press, Cambridge, MA.

Strauss, S. and Bichler, E.: 1988, 'The development of children's concepts of the arithmetic mean', *Journal for Research in Mathematics Education* 19, 64–80.

Székely, G.: 1997, 'Problem corner', *Chance* 10(4), 25.

Véron, J. and Rohrbasser, J.-M.: 2000, 'Lodewijk et Christiaan Huygens: La distinction entre vie moyenne et vie probable', *Mathématiques et Sciences Humaines* 38(149), 7–21.

ARTHUR BAKKER
Institute of Education
University of London
23-29 Emerald Street
London WC1N 3QS
United Kingdom
Tel.: +44-20-7763 2175
Fax: +44-20-7763 2138
E-mail: a.bakker@ioe.ac.uk

KOENO P. E. GRAVEMEIJER
Freudenthal Institute
Utrecht University
P.O. Box 9432, 3506 GK Utrecht
The Netherlands
Tel.: +31-30-263 55 49
Fax: +31-30-266 04 30
E-mail: koeno@fi.uu.nl