Teacher narratives in making sense of the statistical mean algorithm

Author:
Erna Lampen[1,2]

Affiliations:
[1]School of Education, University of the Witwatersrand, South Africa

[2]Research Unit for Mathematics Education, Stellenbosch University, South Africa

Correspondence to:
Erna Lampen

Email:
ernalampen@sun.ac.za

Postal address:
PO Box 455, Stellenbosch 7599, South Africa

Teaching statistics meaningfully at school level requires that mathematics teachers conduct classroom discussions in ways that give statistical meaning to mathematical concepts and enable learners to develop integrated statistical thinking. Key to statistical discourse are narratives about variation within and between distributions of measurements and comparison of varying measurements to a central anchoring value. Teachers who understand the concepts and tools of statistics in an isolated and processual way cannot teach in such a connected way. Teachers' discourses about the mean tend to be particularly processual and lead to limited understanding of the statistical mean as measure of centre in order to compare variation within data sets. In this article I report on findings from an analysis of discussions about the statistical mean by a group of teachers. The findings suggest that discourses for instruction in statistics should explicitly differentiate between the everyday 'average' and the statistical mean, and explain the meaning of the arithmetic algorithm for the mean. I propose a narrative that logically explains the mean algorithm in order to establish the mean as an origin in a measurement of variation discourse.

Introduction

This article explores the knowledge needed by teachers to enable meaningful *mathematical discourse in instruction* (Venkat & Adler, 2012) for the statistical mean. In Venkat and Adler's (2012) work, mathematical discourse in instruction comprises inter alia the explanations and discussions a teacher creates between the stated problem, the initial object, transformations of the object and applications of the result. Establishing coherence between these aspects is the learning task and enabling the construction of coherence by learners through tasks and discussions is the most important role of a teacher. Whilst mathematical discourse in instruction can be understood narrowly as a discourse that aims for local, micro-level coherence from one step of a transformation to another, ending when the problem at hand is solved, the mathematical discourse in instruction that I advocate builds on and is framed by a conceptual orientation (Thompson, Philipp, Thompson & Boyd, 1994) and aims at constructing meaning for statistical procedures that have the statistical horizon in mind, to paraphrase Ball (1993).

Thompson's (2013, p. 61) rendition of Piaget and Garcia's (1991) notion of meaning – 'meaning comes from an assimilation's implications for further action' – motivates for a conception of the mean that includes a rationale for its use in more advanced statistical processes such as calculating the standard deviation and linear regression. Thompson calls for research on teachers' mathematical meaning for teaching in recognition that developing mathematical meanings for teaching requires deep reflection on connections and organisations between mathematical objects and processes in relation to the larger mathematical project: that of providing opportunities for learning to think mathematically. In particular, this article wants to promote deep reflection on the connection between the use of the statistical mean as a central value for a data set and the mathematical procedure to calculate the mean. Such knowledge of the uses of mathematical procedures to create statistical tools is specialised content knowledge (Ball, Thames & Phelps, 2008) that will help teachers to conduct classroom discussions that promote statistical reasoning.

Problem statement

At school level Statistics is usually taught by mathematics teachers, whose studies may not have included courses in Statistics. Hence, the instructional discourse of Statistics tends to be restricted and mostly aimed at instruction for performing well-defined mathematical procedures, such as calculating the mean when it is asked for explicitly. In contrast, statistical thinking 'involves "big ideas" that underlie statistical investigations' (Ben-Zvi & Garfield, 2004, p. 7). Big ideas that have been made explicit in statistics education literature include the ideas of variation and distribution, where measures like the mean and median act as representative values and summaries of distributions.

The *statistical mean* derives its meaning from mappings between practical, everyday discourse about varying observations and mathematical discourse in which the algorithm for the arithmetic mean is understood to effect equal sharing. Evident from Statistics textbooks, the mean is at most reported as 'the average' in a context without any further attempt at explicating the meaning of average. On the one hand, the problem is that average has many contextual meanings that do not all map onto the statistical mean (Watson, 2006). On the other hand, the mathematical algorithm is adopted as the way to obtain the average, rather than logically explained. Teachers who are not aware that different meanings can be assigned to average in context may treat average and mean as synonyms in classroom discussions and fail to provide opportunities to shift classroom discourse from purely informal understandings of average towards statistically literate discourse.

Teachers who cannot logically explain the mean algorithm may fail to explain why it yields a statistically representative number and why the mean is an important statistic in more advanced procedures. Although there is a substantial amount of research about teachers' and learners' explanations of average and mean (Shaugnessy, 2007), an aspect that has not been researched explicitly is the conflation of the arithmetic mean and the statistical mean in teachers' discourses for instruction. I use the term arithmetic mean to refer to the mathematical structure of the mean algorithm and the use of the mean in other than statistical contexts. For example, the calculation of the gradient between two points, and division as equal sharing in typical school tasks, use the mean algorithm without viewing the resultant number as a measure of central tendency for a data set. In this article I explore the following research question: How do high school mathematics teachers reason about the relationships between average and mean and the structure of the mean algorithm? I analyse a group of high school teachers' discussion of the meaning of the statistical mean in relation to data contexts and the algorithm. I show that the teachers' narratives about the mean shift from limited contextual enactments of 'average' and 'middle' to using the mean as a norm to compare data values to. Cognitive conflict about the interpretation of the equal values obtained by fair sharing, whilst the measured values were variable, enabled discursive shift towards statistical discourse. The findings have implications for teachers' potential instructional discourses and suggest a need for an object definition of the statistical mean that takes account of the structure of the mean algorithm.

The research setting

The discussion that provides the data for this article took place in the third session of a semester course in introductory Statistics for high school teachers. The course formed part of an honours degree in mathematics education. I was the lecturer of the course and engaged the teachers as students in deep discussions of data contexts, engaging with and contrasting everyday reasoning with statistical reasoning in such contexts. Twelve students were enrolled in the course.

I arranged the students into three groups of four and video-recorded the discussions of two of the groups. I constituted the groups in a way that would reflect the language complexities of classroom discourse in South Africa, but also provide the best possible chance of promoting discussion. I mainly controlled for power issues related to age, gender and previous knowledge of Statistics. Group 1 comprised mature students who are experienced mathematics teachers, evenly divided according to gender and previous knowledge of Statistics. Two students (KH and RK) had taken Statistics as an undergraduate course. Only one student (KH) had English as a first language. Group 2 comprised young students, with little or no teaching experience. In this group only one student was male, but gender power issues amongst the younger students were unproblematic. Two students (SDS and GG) had English as their first language and three (SDS, NM and MM) had recently done a Statistics course in their B.Ed. programme. In total, five of the eight students in the video-recorded groups had done Statistics courses prior to this course and five of them were teaching Statistics at Grade 10 level at the time of the research. The third group was not included in the study as a separate group, although the contributions of these students were included in analysis of whole class discussions. I decided not to include the last group since they were least balanced in terms of my criteria. The discussions were transcribed from the video tapes and analysed together with the students' written work.

Ethical issues

I studied the group and classroom discussions during the course as part of my doctoral research. Ethical clearance for the study was duly obtained from the ethics committee of the relevant university's School of Education. After a contact session during which information about my research was provided and the conditions for consent were negotiated with the students, they gave informed consent that their recorded discussions and their written work may be used as research data and disseminated in scholarly conferences and publications. The conditions for consent were anonymity in the wider dissemination of the research and ensuring that their withholding consent would not influence their participation in the course or their assessments.

The research design

For this case study I undertook discourse analysis of three sessions of the course in order to investigate emergent statistical reasoning. I used Sfard's (2008) theory of commognition to inform the analysis of the uses of words and other symbols in different discourses. Key to commognition is the notion of thinking as communication and of learning as a process of shifting discourses. This theory allowed me to interrogate the participants' every day and informal statistical reasoning about the meaning of the mean, rather than discount it as idiosyncratic. In order to analyse shifts in discourses, the target discourses must be defined and operationalised. I conceptualised everyday discourse, informal statistics discourse and literate statistics discourse as follows: everyday discourse about average

and mean refers to concrete objects and observations of similarity amongst objects and bases arguments on practical considerations in context and personal experience and opinion. Informal statistics discourse about average or mean comprises narratives that informally explore and compare measurements of variable attributes to derive an informal value of central tendency, related to an informal measure of spread. Literate statistics discourse distinguishes between average as a contextual observation and mean as an abstract measure of central tendency of a data set, and relates the mean as a measure of central tendency to standard deviation as a measure of spread. These operational definitions of the discourses guided my data analysis. For example, everyday discourse was coded if a participant referred to a person as being 'average', or 'the average one', without recourse to measurements. Sfard (2008, p. 57) calls such flattened discourse instances of ontological collapse, in which a construct like the mean is treated as if it belongs to the world of direct observation. Informal statistics discourse was coded when participants indicated, through words, inscriptions or gestures, that average or mean is a position on a continuum which serves to facilitate informal comparison of many objects. Such objectification of average from being a property of an object to being a position on an informal scale indicates a shift to informal statistics discourse. The participants in my study did not provide narratives that could be coded as literate statistics discourse. Such discourse would, for example, refer to the need for a set of data, a formal calculation of the mean and a contextual interpretation of the number obtained.

Commognitive research requires in-depth analysis of the uses of words and discursive patterns in extended discussions. Words are concepts and the ways in which participants elaborate on word uses through other words or representations like gestures allow the researcher to make conjectures about participants' discourses and hence understanding of concepts.

Mean and average in validated discourses

The word usage of the participants in my research is not independent of culturally validated uses in different discourses. Hence, I begin by contrasting the meanings of average and mean as they are used in three discourses: everyday discourse evident from dictionaries, statistics discourse used in subject dictionaries and mathematics discourse as evident from the historical emergence of the arithmetic mean. Then I discuss literature about discourse on average and mean that emerge in teaching and learning situations.

Dictionary definitions of mean and average

A study of dictionary entries under 'average' and 'mean' reveals an opaque and circular relationship between the two terms. In Table 1 I compare the definitions of average from a dictionary of everyday usage: the Merriam-Webster Online Dictionary (Merriam-Webster, 2015), and a Statistics dictionary: Collins Dictionary of Statistics (Porkess, 2004).

A comparison of the everyday and statistics definitions of average in Table 1 indicates that average as being *typical or representative of a group* is a shared meaning in the two discourses. However, in everyday discourse average is 'an estimation or approximation to an arithmetic mean' whilst in statistics discourse average may refer to 'any (or none) of mean, mode, median and midrange'. Hence, the statistics point of view acknowledges that the term average derives meaning mainly from context and the everyday perspective acknowledges that what is average *may be approximately the same* as the value calculated by the mean algorithm.

A second observation is that in both discourses average is implicitly utilised as *a point for comparison*. In the examples provided for average as typical or representative (see entries numbered 1 in Table 1), objects are described in comparison to average as 'above average' or 'is average'. This use of average is not made explicit, yet I will argue later that the mean as a logical point to which to compare other measurements is a crucial narrative in a discourse about variation.

In Table 2, in everyday discourse the term 'mean' is explained as a middle position (though not necessarily a number) between extremes and as a *calculated value* that falls within a range of values. Similarly, in statistics discourse mean is defined as a *measurement* of average, with the vague concession that there are different ways to measure average appropriately.

The definitions of 'mean' in the Merriam-Webster Online Dictionary (Merriam-Webster, 2015) emphasise the 'laws' for calculating the arithmetic mean or the expected value and refrain from explicit contextual examples; these therefore belong to a more abstract discourse than the definitions of 'average'. This analysis and comparison of the sanctioned meanings of average and mean reveals a disjunct that begs explanation: intuitively and informally average is

TABLE 1: Comparison of definitions of average in everyday and statistics discourses.

Everyday discourse *Merriam-Webster Online Dictionary* (Merriam-Webster, 2015)	Statistics discourse *Collins Dictionary of Statistics* (Porkess, 2004)
1. (a) A single value (as mean, mode or median) that summarises or represents the general significance of a set of unequal values 1. (b) A level (as of intelligence) typical of a group, class or series - 'above the average'.	1. In everyday use the word average is often used loosely to mean typical or representative, as in a statement like 'William is average at football'. … According to context, it may be any (or none) of mean, mode, median and midrange (p. 14).
2. An estimation of or approximation to an arithmetic mean.	2. In technical use, average usually has the same meaning as mean or arithmetic mean. In certain contexts technical use (of average) requires other types of mean … geometric mean and harmonic mean.
3. A ratio expressing the average performance especially of an athletic team or an athlete, computed according to the number of opportunities for successful performance.	-

TABLE 2: Comparison of definitions of mean in everyday and Statistics discourses.

Everyday discourse *Merriam-Webster Online Dictionary* (Merriam-Webster, 2015)	Statistics discourse *Collins Dictionary of Statistics* (Porkess, 2004)
1. (adjective): Something intervening or intermediate.	-
2. A middle point between extremes	-
3. A value that lies within a range of values and is computed according to a prescribed law: as (1): arithmetic mean (2): expected value	1. A measure of an average value. There are several types of mean, used in appropriate circumstances, but unless stated otherwise the term 'mean' is usually taken to be the arithmetic mean (p. 150) [*The dictionary then proceeds to give the algorithms for the arithmetic, geometric, harmonic and weighted means in algebraic notation*]
4. Either of the middle two terms of a proportion	-

representative and serves as a point of comparison, yet these meanings are not carried over in the definitions of the mean. In particular, it is problematic for instructional discourses that Porkess's (2004, p. 14) statistical definition of the arithmetic mean as a 'measure of an average value' fails to explain *why* the calculation of the arithmetic mean is a measure of an average value or *how* it manages to be a middle, typical or representative value.

Research about understanding of the statistical mean in teaching and learning situations indicates that the conflation of average and mean is problematic for teaching, since it leaves the ontologies of the mean and the average unexplained. A teacher who needs to answer the question 'what is the statistical mean?' may invoke the calculation procedure to imply 'the mean is what is does', but, as the statistics education literature reports, the process-definition is open to varied interpretations.

Statistics education research: Understanding average and mean

In-depth interviews as well as large-scale studies that have researched the meanings learners and teachers assign to the mean provide wider context for the meanings of average and mean, which are reflected in dictionaries. It also illuminates the potential for confusion in statistics classrooms: literally, participants in a classroom discussion may not be talking about the same thing when they refer to average or to mean.

Everyday meanings of average

Various meanings of average in everyday discourse are described in Statistics education literature. Both teachers and learners routinely elaborate the meaning of 'average' as 'middle'. In turn, 'middle' is understood in more than one way: sometimes middle is determined by active ordering of measurements of some attribute, where after the middle *position* between minimum and maximum is assigned to 'average'. This meaning of average can be mapped on the statistical median or on the midrange. Sometimes, middle is achieved by excluding extreme values so that middle refers to an *interval* of similar values rather than a single value. This meaning of average can be a precursor of a measure of spread of similar values, rather than a measure of central tendency (Konold & Pollatzek, 2004; Makar & McPhee, 2009; Mokros & Russell, 1995; Watson & Moritz, 2000).

Average is also explained as 'typical' in everyday discourse. When data are available, 'typical' tends to be associated with the most frequent observation (Konold & Pollatzek, 2004) but also with a reasonable range of values (Makar & McPhee, 2009). In these meanings the confusion between average as a single value or a range of values is evident: average as the 'most frequent' observation can be mapped onto the statistical mode rather than the statistical mean, whilst a 'reasonable range' indicates early notions of spread of near-similar data points.

The complexity does not end here. Everyday meanings of average do not depend on the comparison of numerical values. Interpretations of average are often based on qualitative judgments of what is experienced as 'not extreme'. Hence, a person can be described as average in appearance, based on a qualitative judgement of appearance that lies between extremes, for example the extremes of ugly and attractive. 'Average' in context may be so tightly associated with normative contextual descriptions that it is associated with adjectives like good, bad (to score an 'average' mark is good or bad, depending on the value of the average mark), low, high, cheap or expensive, rather than reflecting a relationship between overt or covert measurements of an attribute of a collection of objects (Lampen, 2013).

These everyday meanings of average held by teachers and learners suggest that simply explaining the number obtained by the mean calculation as the average does not provide access to statistical discourse. Indeed, the equal sharing meaning suggested by the mean algorithm is not associated with average by people who do not know the algorithm (Mokros & Russell, 1995): in many everyday contexts where observations are not equal, the mean as an equal share makes little sense.

Didactical meanings of the statistical mean

Attempts to unpack the mean didactically as a statistical object have led to descriptive definitions such as an equal share, true value, signal in noise, balance point or representative value (Konold & Pollatzek, 2004). In these definitions the mean refers to a distribution of data, abstracted from a collection of contextual measurements. Studies of meanings assigned to the mean have not specifically asked participants to explain what they understand by these descriptions; rather the descriptions have been used by researchers to categorise ways in which participants interpret graphs and data sets. Only rarely have learners or teachers without formal statistical background responded in these statistically descriptive categories (Groth & Bergner, 2006; Watson &

Moritz, 1999) and there is consensus that such abstract meanings of the mean are difficult to develop (Konold & Pollatzek, 2004; Watson & Moritz, 2000).

Makar and Confrey (2004) concur that the statistical relationship between a distribution as an object and the mean as a measure of the object is opaque, whilst Mokros and Russell (1995) draw attention to the disjunct between understanding the process of measuring the distribution and the mean as an object when they say 'the mathematical relationship [of the mean algorithm and the uses of the statistical mean] itself remains opaque' (p. 22). Cortina, Saldanha and Thompson (1999) propose a conceptualisation of the statistical mean that consciously measures variation and yields an object:

> students need to create the mean as an adjustment on the measure of group performance … as one runs through the contribution of cases to the mean of the group. (p. 2)

However, in their conceptualisation, the mean as an object is a multiplicative concept that serves as a measurement of group performance, hence it foregrounds the relationship:

$$n\bar{x} = \sum_{i=1}^{n} x_i \qquad \text{[Eqn 1]}$$

Historical discourses: From the arithmetic mean to the statistical mean

Historically the concept of the mean can be traced back to estimation in order to solve practical, measurement-related problems and the geometric construction of different means in mathematics, namely the harmonic, geometric and arithmetic means. Statistical use of the mean can only be traced back to the 19th century (Bakker, 2004). In this section I draw on research about the historical development of the mean algorithm to show that the arithmetic mean and the statistical mean are different concepts, despite having the same algorithm. The difference lies in the discourses in which they are used.

The arithmetic mean

Bakker (2004) describes two different calculation procedures that were historical precursors of the mean algorithm, even if these processes were not named with terms related to average or mean. The historical enacted algorithms provide insight into the uses and therefore the concepts that have underpinned the concept of average.

The first procedure uses one representative value multiplicatively to estimate a large total number. Bakker (2004) gives two examples. In the first example[1] the number of leaves on a twig was multiplied by the number of twigs on the tree to estimate the number of leaves on the tree. In the second example, the thickness of a brick was estimated and multiplied by the number of layers of bricks in a wall in order to estimate the height of the wall[2]. In these early

historical examples the term average does not appear; instead the method or process of calculating some practical quantity was described in words. The goal was to determine a direct measurement for a physical object. Bakker interprets the relevance of these examples as incorporating notions of the arithmetic mean in relation to the statistical concept of representativeness (the number of leaves on one twig is representative of the number of leaves on all the other twigs). The totals in the examples were calculated according to the algorithm:

$$\text{(A representative object)} \times \text{(number of objects)} = \text{total number of objects.}$$

[Eqn 2]

Structurally, 'a representative object' represents the mean and its value can be calculated by a simple transformation of the relationship above. It is important to note that in this historical use of finding a total number of objects the mean was not an unknown or hypothetical value. It was the smallest component unit (a brick in a wall or leaves on a twig) that could be used to access measurements of larger, composite objects (rows of bricks and walls or leaves on a tree). Hence, there is no intuitive conceptual step to 'creating' the arithmetic mean by equal sharing. In practice, bricks are made to a standard size whilst the heights of walls vary; it does not make practical sense to ask how wide a brick must be to build a wall of a given height with a given number of rows.

The geometric concepts of arithmetic, geometric and harmonic means existed long before the statistical concept of mean and were studied in Pythagoras's time (around 500 BC). In ancient Greece, where these concepts were mathematically formalised, lengths were constructed with the use of compasses and straight edges and treated as concrete objects (to the extent that numerical discourse on square root lengths was problematic). Bakker (2004, p. 56) cites the theorem of Pappus in which the arithmetic mean, the geometric mean and the harmonic mean of two line segments were indicated in a single construction (see Figure 1). The construction placed the two line segments AB and BC as extensions of each other, so that the combined length was $a + c$ and formed
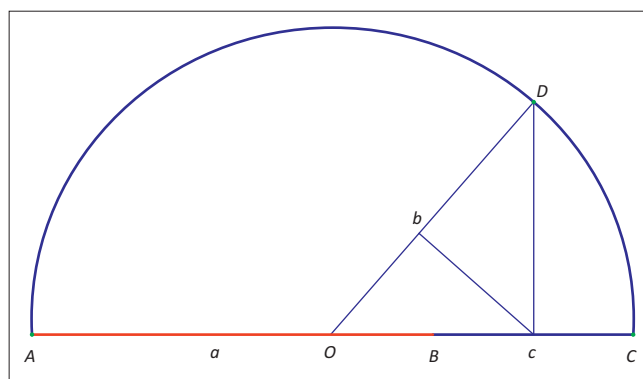
Source: Adapted from Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Unpublished doctoral dissertation. Centre for Science and Mathematics Education, Utrecht University, The Netherlands (p. 56). Available from http://dspace.library.uu.nl/bitstream/handle/1874/893/full.pdf?sequence=2

FIGURE 1: Theorem of Pappus: OD is the arithmetic mean of AB and BC.

---

1. An ancient Indian story reported by Hacking (1975).

2. From the history of the Peloponnesian War (431–404 BC). See Bakker (2004).

the diameter of a circle. Hence, the arithmetic mean was half of the diameter (the total length), which is the radius.

Through the construction of Pappus (ca. 320 AD) the arithmetic mean existed as an object with a measurable length. The formula that was used to calculate *b* as the average or middle length of two lengths *a* and *c* was:

$$a - b = b - c \qquad \text{[Eqn 3]}$$

In this equation it is clear that the mean length (*b*) is between the two lengths it has to average. Expressed in words, *b* is the length between *a* and *c* such that the difference between the lengths of *a* and *b* is the same as the difference between the lengths of *b* and *c*. However, reasoning about the lengths of geometrically constructed line segments as in Pappus's theorem does not lead to the mean algorithm, since the radius of a circle is always half the diameter, and not an *n*th part. Only in the 16th century, and possibly enabled by the development of the decimal system, was the arithmetic mean generalised to more than two cases (Bakker, 2004). Bakker draws attention to the historical process, since about 700 BC, of averaging the value of cargo losses at sea, so that such losses could be shared equally between merchants and shippers. This meaning of average is reflected in the following definition of average as a transitive verb: To divide among a number, according to a given proportion; as, to average a loss (Merriam-Webster, 1913).

According to Bakker, it is unclear how average in this sense came to signify the arithmetic mean and when and how the shift from the concept of the arithmetic mean to the statistical concept of representative value or balance point of a data set occurred. Such loose ends in overlapping discourses about average and mean are problematic in teaching for statistical reasoning.

The mean of a distribution

The use of mean in a discourse on variation, hence statistical discourse, developed quite recently in the history of mathematics. Until about the 19th century the calculation of the mean was used to find a 'real' value, a measurement of a physical object (e.g. the diameter of the moon or the number of leaves on a tree). Bakker (2004) dates the first use of the mean as 'the representative value for an aspect of a population' around 1835, when the Belgian statistician Quetelet invented the term *l'homme moyen*, the average man. This use of the mean as a representative value rather than a 'real' value, as in astronomy, was an important, yet difficult step in the development of variation discourse on the mean. Fifty years after Quetelet's invention, Charles Peirce, mathematician and philosopher, wrote in 1877 how problematic it was to map continuity of measurement onto situations where measurements are in discrete units, in order to report averages like 'there are in the United States 10.7 inhabitants per square mile' or to talk of 'the average man'. According to Bakker Peirce preferred 'most men' instead of 'the average man' (p. 61).

Conceptualising the relationship between average, arithmetic mean and statistical mean for instructional discourse

I now report on the meanings of the statistical mean that emerged in a discussion of the mean algorithm by a group of high school teachers, after which I reflect on connections between their narratives about the mean and average, and their understanding of the meaning of the division step in the mean algorithm; finally, I consider possibilities for integrated discourse for instruction of the mean as a statistical concept.

Framing the discussion of the meaning of the mean

Prior to the discussion of the meaning of the mean, the students had studied real data of samples of prices of used cars and drawn various graphs of the data with the aid of FATHOM™ in order to investigate shapes of distributions and to estimate measurements that could reasonably serve to represent and summarise central tendency and spread. They had also compared calculated values of the mean and the median to their estimations on graphs. Furthermore, the sensitivity of the mean to extreme values had been explored empirically and discussed as a reason for representing and comparing skewed data sets by the median rather than the mean. Hence, all the students knew how to find the median and how to calculate the mean.

I introduced the following prompt for the discussion of the meaning of the mean algorithm:

'What is the logic or common sense behind using the mean as a measure of centre?'

The aim of the discussion as a learning task was to engage the students in analysing the meanings of average and mean, and in constructing a logical connection between the syntax of the mean algorithm and the role of the mean as a statistical measure of centre. In my analysis of the discussions I looked for 'seed concepts' that could be used in discourses for instruction to develop statistical reasoning about the mean. In particular, I wanted to understand if and how the participants considered the enacted meanings of addition (putting together) and division (sharing or grouping) in their explanations of the mean algorithm. It transpired that their discourse maps well onto everyday discourses such as those evident from the dictionary entries. The students too explained mean as average and average as mean with 'middle' as the predominant spatial image. They were at a loss to give meaning to the mean algorithm, yet they developed a generative narrative of the mean as a norm or a value to which to compare measurements. This narrative holds the key to a new object definition of the mean. I will now report on seven meanings that emerged during group and whole class discussion of the meaning of the mean algorithm. The excerpts are provided in chronological order and provide the

opportunity to describe discursive shifts in the discussion. In order to establish confidence in the credibility of my own interpretive narratives (and hence the validity of my research) I provide extended transcripts of the discussions (Sfard, 2012, p. 8). Full transcripts of the discussions are available in Lampen (2013).

Results: Narratives about mean and average

Meaning 1: Mean is average

Throughout the group and class discussions the students explained the mean as the 'average' in contexts in which they imagined the mean could be used. The excerpt in Box 1 is an example. The numbered turns provide a chronological order for the students' utterances.

At first glance it appears that the students are treating mean and average simply as synonyms, yet in Turn 10 and Turn 15 KH's utterances suggest a primary ontological position for average. The students seem to share the common sense meaning of average that they believe 'people' have. The discussion about the mean as an object ('the mean is …') stops here. The ontological collapse in this narrative prevents the students from further reasoning. The requirement to further unpack the meaning of average seems ridiculous: the mean is 'just' the average as if the average was self-evident and no further explanation is needed.

Meaning 2: Average gives a general picture

In the excerpt in Box 2 the discussion shifts to why the mean is used as a measure of centre. The discussion is based on references to imagined contexts of real objects: that of a class of 'kids' of different heights and cars with different prices. Through its conflation with average the mean provides 'an impression' and 'a general picture' of a situation. In this narrative the mean provides one with a bird's eye view in which the differences between the imagined objects recede and the similarities remain.

Intertwined with the impression narrative in Box 2, a narrative about mean-as-middle develops. In contrast with the impersonal 'it gives …' (Box 2, Turn 18 and Turn 23), the 'middle' narrative in Box 3 draws the observer into the context: 'you have to order it'; 'you take the middle value'

and 'then you know'; 'exactly half are above that height and exactly half are below'. In the excerpt in Box 3 the use of middle in relation to average and median raises conflict.

Meaning 3: Average is middle

In the excerpt in Box 3 RK, who is the leading discussant, first describes average as a value in the middle of some interval where objects (kids) would converge if compared by a measurement like height (Turn 20). In Turn 23 RK insists that this average as a middle value gives a general impression of the situation. KH (Turn 26) initiates a discussion about middle as being representative and the procedure to find the middle value. She queries the assertion that average is the only middle value through her reference to the median.

RK's leading narrative about the mean as a 'middle value' is within everyday discourse in which physical examples and imagined contexts are used to give weight to the

BOX 2: Average gives a general picture.

Turn	Student	Utterance
17	KH	So *why* do we use the mean? [*Glances in textbook*].
18	RK	I think it's because it gives them the picture. … It captures … some particular number encapsulating … like the mean height.
22	RK	So how could we phrase it?
23	RK	It kind of gives the general picture of how tall the kids is.
24	KH	Yes, yes.
34	RK	In a class *definitely* all the kids can't have the same height. … But if you are asked the question, what is the average height of the kids.
36	RK	You give a *number*, you don't necessarily talk about the tallest one or the shortest one [*Gestures high and low with sharp hand movements*].
37	KH	Mm.
44	RK	The mean is the general picture.
45	KH	Yes, yes an impression.

Source: Lampen, C.E. (2013). *Learning to teach statistics meaningfully*. Unpublished doctoral dissertation, University of the Witwatersrand, Johannesburg. (p. 232). Available from http://wiredspace.wits.ac.za/handle/10539/13349

BOX 3: Average is middle.

Turn	Student	Utterance
20	RK	I want to say, if you say the average height of kids. … Let's say one meter two [*1.2 meters*] … you say that generally … you find kids *around* that (moves hand horizontally at the same height at which he indicated 1 meter 2).
23	RK	It [*the mean*] kind of gives the general picture of how tall the kids is.
26	KH	[*to RK*] Why doesn't the median do the same?
28	RK	I think the median is like … if you have it ordered.
30	RK	You take the middle value.
31	GK	The middle value is the median yes.
32	SM	That is the mean.
38	RK	[*Moves hand across horizontally*] The average gives you the middle.
39	KH	Exactly half are above that height and exactly half are below.
40	GK	And what is the median.
41	RK	That is the median.
42	GK	The half is the median.
43	KH	Aha, yes, OK.
44	RK	The mean is the general picture.

Source: Lampen, C.E. (2013). *Learning to teach statistics meaningfully*. Unpublished doctoral dissertation, University of the Witwatersrand, Johannesburg. (p. 232). Available from http://wiredspace.wits.ac.za/handle/10539/13349

BOX 1: Mean is average.

Turn	Student	Utterance
7	SM	All right, we can say that it [*the mean*] is the average price that you can pay me.
8	KH	Yes, so it's the … just the average.
9	SM	Yes just the average price.
10	KH	It's what people understand by average.
14	RK	… the common sense behind that [*the algorithm*]
15	KH	I think it is because … it is the average. When you talk to the general public, the mean is the average, they understand average. Median is a different aspect [*weighing movements with hands*].

Source: Lampen, C.E. (2013). *Learning to teach statistics meaningfully*. Unpublished doctoral dissertation, University of the Witwatersrand, Johannesburg. (p. 227). Available from http://wiredspace.wits.ac.za/handle/10539/13349

argument. KH's narrative, on the other hand, is anchored in statistical discourse, drawing on the procedural definition of the median. The students seem to have control over the median: they are certain they find the middle when they calculate the median position, whilst there is no such agency in their narrative about the mean. Since the logic by which mean becomes middle is not clear, the students are unable to resolve the conflict around the meaning of the mean-as-middle, and RK and KH (Turn 44 and Turn 45) retreat to the initial realisations of mean as 'the general picture' and 'an impression' of what is going on in a situation in which it is used. An underlying problem is that the objects that support the reasoning at this stage are a concrete, although imagined, collection of 'kids'. The mean does not have anything more to say about this collection; average is adequate. With no recourse to logical reasoning about the syntax of the mean algorithm in relation to average and average-is-middle, there is no opportunity to develop more abstract statistical narratives about the mean. As I mentioned before, the students knew how to calculate the mean and how to find the median; hence, their confusion between mean and median cannot simply be ascribed to lack of algorithmic knowledge.

Meaning 4: Average is most

In the excerpt in Box 3, Turn 20, RK pointed out that the mean is such that 'generally … you find kids *around* that', and is therefore a centre within an interval. In the excerpt in Box 4 (Turn 49 to Turn 50), another property of average is realised in everyday discourse, namely that average describes an interval that captures most objects.

In Turn 49 GK agrees with the narrative that the mean as the average gives a general picture of some aspect of a context. She then realises her understanding of the use of the mean algorithm. The result of 'add[ing] up the total and dividing it by the number' is realised as a frequency of occurrence 'how often you can get it'. With her verbal realisation of average as most, GK gestures grouping together of objects within brackets. In Turn 49 (Box 4) GK strengthens the realisation of average as a *place* rather than a measurement or a property of an object: 'Most of the learners are here … in a certain average'. Utterances of 'most' are interpreted in the statistics education literature as unrepresentative modal understandings of the mean (Mokros & Russell, 1995), but I interpret GK's combined verbal and gestural realisations as 'most will be around the mean, *because* they are average' (see also RK's utterance in Box 3, Turn 20). GK does not refer to a measurement that occurs most often (the mode), but to the majority of cases that were grouped together as 'average'. RK does not explicitly take up the notion of average as an interval; on the contrary, his emphasis on 'general' together with a sweep of the hand (Box 4, Turn 50) supports replacement of many measures by one.

At this stage in the discussion the student teachers do not have access to narratives that unpack the meaning of the mean; instead, their narratives compare uses of the statistical mean with the everyday, self-evident notion of average.

BOX 4: Average is most.

Turn	Student	Utterance
49	GK	[*To RK*] No, but I get what you are saying. You know when you say you got a total, let's say you want to find the average of something. You know you add up the total and you divide it by the number. In fact it's telling you the average, how often can you get that. Most of the learners are here [*makes brackets with her hands*] in a certain average.
50	RK	[*To GK*] Yeah. Say we had a thousand people, are you seriously going to investigate one-one? When you take the mean, the average height of everybody it gives you the *general* picture [*Sweeping gesture*].
51	GK	Mm-mm [*agreement*].
52	RK	How can we say it?

Source: Lampen, C.E. (2013). *Learning to teach statistics meaningfully*. Unpublished doctoral dissertation, University of the Witwatersrand, Johannesburg. (p. 239). Available from http://wiredspace.wits.ac.za/handle/10539/13349
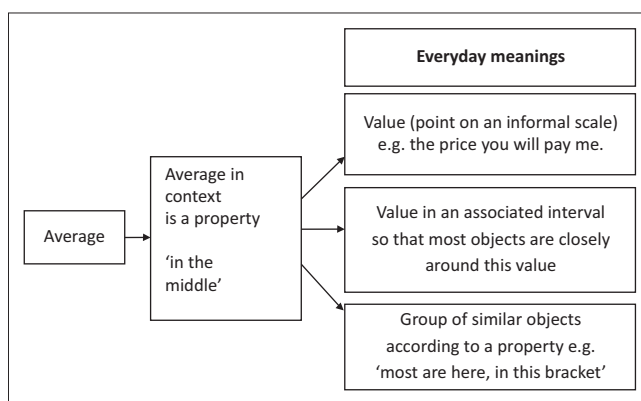

FIGURE 2: Three narratives about the mean as the average in everyday discourse.

Figure 2 summarises the available narratives that relate mean to average in context.

The ontology of the mean – what the mean is – is completely realised in intuitive everyday understanding of average in which similarity and extremity are observed properties of objects. The epistemology of the mean is similarly intuitive and practical: we come to know what the mean is through its uses in everyday contexts. Hence, both ontology and epistemology of the mean in these teachers' narratives are intuitive and restricted to everyday discourse. The meanings they assign to the mean as average are reflected in the dictionary definitions I mentioned earlier. The problem is that even the definitions in the statistics dictionary do not provide a way out of the conundrum of the conflation of mean and average.

In the ensuing discussion the conflation of mean and average is gradually resolved. By comparing measurements to the mean, the mean is useful to determine what is *not* average.

Meaning 5: The mean is a value to compare to

In order to focus the discussion on the syntax of the mean algorithm, I led the student teachers to think about the division step as equal sharing and then challenged: 'What does it help you to pretend they are all the same? They are not the same!' (in reference to the sample of car prices that was used in the group discussion). The students haltingly started to compare a state in which all the cars were hypothetically assigned the same price and the actual state of variable prices.

BOX 5: The mean is a value to compare to.

Turn	Student	Utterance
77	RK	Example of buying a car. I mean. If … if you typically buy a car, it tells you in this car shop, you know that this brand of car, the RunX I want to buy, it generally costs around this [mean] price. [GK makes weighing movements with her hands] I mean in terms of money I know what to prepare. This [mean] amount plus or minus [hand movements left and right of an imaginary point/line on the horizontal].

Source: Lampen, C.E. (2013). *Learning to teach statistics meaningfully*. Unpublished doctoral dissertation, University of the Witwatersrand, Johannesburg. (p. 249). Available from http://wiredspace.wits.ac.za/handle/10539/13349

In the excerpt in Box 5 RK replaces vague impressions of mean as average and middle by a narrative about the mean as a calculated number that is in the middle of the average values and a value that anchors the actual values mathematically: if the mean is known, the actual values can be found by addition or subtraction. This understanding can be related to the definition of the mean as a measurement of average in the statistics dictionaries (see Table 2) and stimulates the abstraction of the mean from average.

Meaning 6: Far from the mean is not average

Concurrent with the discussion of the first group reported so far, the second group of four students that were video-recorded raises the distance of a point from the mean as a means to judge in context whether an object is average or not.

In Turn 269 (Box 6) NM talks about her learners' marks and in Turn 273 GG talks about prices of used cars; the implication of the discussion is that distance from the calculated mean holds qualitative information about the object: a mark far from the mean may be judged (Turn 270) as good or bad, whilst a price that differs by R60 000 from the mean is 'way out of the average' and presumably too expensive in comparison to the rest. Equal sharing is the enacted concept that is related to the mean as a point of comparison. These narratives about distance from the calculated mean indicate a further shift in discourse from every day to informal statistics discourse as it allows the meaning of the mean as a 'constant' or a 'norm' to emerge.

Meaning 7: Mean is a constant and a norm

The discussion of the meaning of the mean algorithm closes with tentative object definitions of the mean as a constant amidst variable measurements and as a norm. The accompanying procedure is that of levelling out variable measurements.

In the excerpt in Box 7 RK (Turn 144) tentatively realises the mean as some constant value compared to the variable measures in a data set. This realisation signals a crucial shift in his discourse: without the mean, we are aware of relative variation amongst actual measurements; with the mean we become aware of deviation from a single hypothetical measurement. RK interprets this 'constant' as an approximation to the actual values in context. RK's choice of the term constant was meaningful. The Merriam-Webster online dictionary (Merriam-Webster, 2015) defines the noun

BOX 6: Far from the mean is not average.

Turn	Student	Utterance
269	NM	I think it's because … remember that total is coming from all of them, so sharing their effort, for example, that total that you have just before dividing, so if they were to share … [laughs, giving up] … if they share equally … I keep on going back to this: once they share equally we will be able to see if … they're far *from* the whole or if they are very close to the whole.
270	SDS	So then you can judge in terms of it [head in hands quite despondently].
273	GG	So for example, if we go to the car one [in the data set] … then it's like, if we find the mean of that [all the cars], say it's hundred and thirty thousand then we can see that the car that was hundred and ninety thousand is *way* out, was way out of the average … OK? Not average, was way out.

Source: Lampen, C.E. (2013). *Learning to teach statistics meaningfully*. Unpublished doctoral dissertation, University of the Witwatersrand, Johannesburg. (p. 256). Available from http://wiredspace.wits.ac.za/handle/10539/13349

BOX 7: Mean is a constant and a norm.

Turn	Student	Utterance
144	RK	It gives me a sense that uh, I think the value of the mean perhaps we should have a constant that we have, [The real values] it's [the mean] plus or minus *something*. You know what I'm saying (bashful), because, it's like [gestures shaking hand] an approximation.
232	SDS	For example the marks, if you add all these marks together, and you divide by 5 [points to her dot plot], then you're trying to get that each person had *this* mark [gestures bars of equal height]. If each one had to have the same mark they would have *that* mark. So it's like levelling out, find the norm.
246	NM	We are sharing the mark equally to all of them, the total mark after adding their marks up, we're sharing it equally [gestures].
247	SDS	Like in your histogram, when we were looking at the histogram, remember, like I was saying [drawing on the histogram] if you take this little piece and take that piece.
248	NM	Yes, add the mark.
249	SDS	*That*, [draws] if you level them all out, *that* gives you the mean

Source: Lampen, C.E. (2013). *Learning to teach statistics meaningfully*. Unpublished doctoral dissertation, University of the Witwatersrand, Johannesburg. (p. 254–255). Available from http://wiredspace.wits.ac.za/handle/10539/13349

'constant' as follows: 'a number that has a fixed value in a given situation or universally or that is characteristic of some substance or instrument'.

SDS's explanation (Turn 232 and Turn 247) of the result of evening out as norm supports the shift in the discourse from intuitive awareness of variation in context to comparing measurements to a fixed number. In these attempts to define the mean as an object, the position of the mean (in the 'middle') is not mentioned. Levelling out and fair sharing emerge as process meanings of the division step. Figure 3 provides a summary of the narratives of the meaning of the mean algorithm.

In the discussion of the meaning of the mean algorithm, the mean emerged as a hypothetical, abstract object that serves as an objective point of comparison amongst measurements. Hence, the conflation of average and mean is resolved and the students' narratives now belong to informal statistical discourse.

Discussion

The meanings of the mean and average that emerged in my study support findings in the literature that the mean algorithm
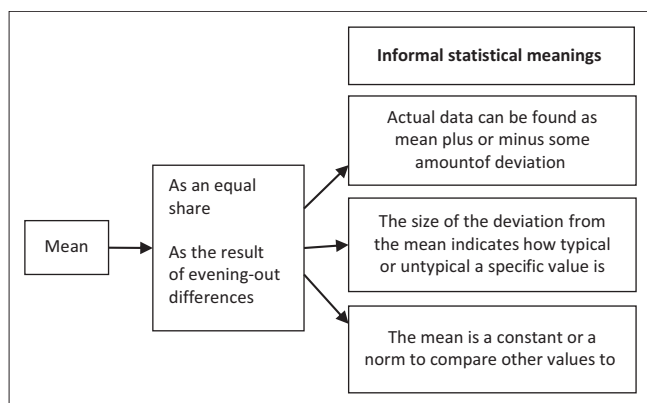
FIGURE 3: Informal statistical narratives on the meaning of the mean algorithm.

is badly understood by teachers. The tendency to accept the mean as a ready-made formula to assign a number to a variety of everyday meanings of average is pervasive and persistent. The reported discussion suggests that, unless teachers consciously work to separate the meanings of the calculated mean and the contextual average, their discourses for instruction will be limited to everyday, experiential meanings.

From the students' discussion I identified two seed narratives for developing connections between average, the mean algorithm and the statistical mean. The students' narratives presented the mean as *an evening-out process* and the mean as *an object, namely a norm to compare to*. I propose that these two narratives are conceptual process-object counterparts that can be developed to logically relate the arithmetic mean to the statistical mean. In the rest of the discussion I will illustrate a possible discourse for instruction towards this integration.

Evening out as a process to derive the mean algorithm

Evening out is reported in the literature as an intuitive process to find a mean value (Bakker, 2004). In the absence of data, evening out is used even by young learners when they can draw on case-value bar graphs. A case-value bar graph represents specific cases and their measurement values as bars with different lengths. In accompanying discourses for instruction teachers view the task as completed when the evening out of bars is achieved, but the process is not abstracted in relation to the mean algorithm. Furthermore, narratives about evening-out processes refer to the bars (case values) and not to the differences between the bars. Yet, evening-out processes are based on redistributing *differences* between bar lengths. I will illustrate how attention to the evening out of differences can be productively used in a measurement of variation discourse that shifts to the statistical mean.

The bars in a case-value bar graph can be ordered from small to large to support a narrative about ordered evening out. The process is illustrated in Figure 4.

As a narrative the algorithm proceeds as follows: even out the difference between the smallest and the second smallest
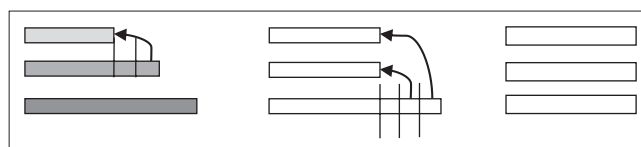

FIGURE 4: Evening out differences between ordered measurements.

Let $x_1 < x_2 < x_3$ be measurements.

Even out the difference between x_1 and x_2:

$$x_1 + \frac{1}{2}(x_2 - x_1) = \frac{x_1 + x_2}{2} \text{ and } x_2 + \frac{1}{2}(x_2 - x_1) = \frac{x_1 + x_2}{2}$$

Next, even out the difference between x_3 and the evened measurements.

Each of two the evened measurements is increased by $\frac{1}{3}\left(x_2 - \frac{x_1 + x_2}{2}\right)$:

$$\frac{x_1 + x_2}{2} + \frac{1}{3}\left(x_3 - \frac{x_1 + x_2}{2}\right)$$
$$= \frac{x_1 + x_2}{2} + \frac{x_3}{3} - \frac{x_1 + x_2}{6}$$
$$= \frac{2x_3 + 2(x_1 + x_2)}{6}$$
$$= \frac{x_1 + x_2 + x_3}{3}$$

Since the largest measurement x_3 was reduced by $2\left(\frac{1}{3}\left(x_3 - \frac{x_1 + x_2}{2}\right)\right)$ its value after evening out is

$$x_3 - 2\left(\frac{1}{3}\left(x_3 - \frac{x_1 + x_2}{2}\right)\right)$$
$$= \frac{x_1 + x_2 + x_3}{3}$$

All three quantites are now equal and hence the mean is $\frac{x_1 + x_2 + x_3}{3}$.

FIGURE 5: Algebraic derivation of the algorithm for the statistical mean.

measurement by taking away half of the difference between the measurements and adding it to the smallest measurement. Then the difference between the largest measurement and the two equalled measurements is shared equally amongst all three bars to achieve the mean measurement. This process can be extended to any number of measurements. Modelling the evening-out action closely, the algebraic process yields a mathematical narrative about the algorithm for the statistical mean, as shown in Figure 5.

Structural differences between the arithmetic mean and the statistical mean

The evening-out process to derive the statistical mean can be described as a first-divide-then-redistribute process, since in this enacted narrative division happens first and is effected on a single measurement at a time. Each bar is divided according to the proportion required to even out bars that are shorter. In this example, in the first step the difference between the shortest bar and the second shortest bar is halved, whilst in the second step, the difference between the length of the evened bars and the remaining long bar is divided into thirds. The redistribution between the bars is additive. Consequently, there is a disjunct between the mathematical structure of the mean algorithm (where division is the final action) and the meaning derived from the evening-out process. The disjunct demands a statistical redefinition of the object that is constructed by evening out. The object definition of the mean as a 'fair share' is not compatible with the process of sequential sharing between

two measures at a time. An object definition based on the narratives that emerged about the mean as a norm in my research is the following: the mean is *an origin of zero variation for the purpose of measuring variation.*

The statistical mean as a norm in relation to the mean algorithm

Statistics education literature abounds with reports of learners' inappropriate comparison of distributions according to a contextually meaningful measure, rather than a statistical measure of central tendency (Bakker & Gravemeijer, 2004; Ben-Zvi & Arcavi, 2001; Konold & Pollatzek, 2004). Various explanations are given for such non-statistical comparison, such as students' perceived roles in the task context (Bakker, 2004), their level of knowledge of the context (Pfannkuch, 2011) and local rather than global conception of distributions (Ben-Zvi & Arcavi, 2001). In addition, I argue that comparison to the mean is not logically motivated in a measurement of variation discourse. Measurement of variation raises the questions of where to measure from, that is, what value shall act as the 'zero' or 'origin', and what the unit is that shall be iterated. The answers to these questions do not lie in discourse about average in context, but fundamentally engage with the arithmetic mean as a statistical model. The evened-out value acts a standard of zero variation amongst varying measures in a data set. Just as any measurement tool has a zero value from which deviations are quantified, so the mean is the origin for measuring variation in a data set. The standard deviation, also based on the concept of a mean, can then be developed as the unit of measurement of variation.

Conclusion

In addition to reflecting on the connections between statistical concepts, a teacher who wishes to teach Statistics as a cycle of enquiry (Wild & Pfannkuch, 1999) needs to reflect deeply on the connections amongst three discourses: the everyday discourse in the real-world context in which the enquiry takes place, the how to and why discourses about the applications of the statistical concepts that are to be developed through this enquiry and the why discourse that logically motivates the mathematical tools that are used in statistics. The last discourse is neglected in Statistics education research and hence in the education of mathematics teachers who teach Statistics at school.

In this article I have argued that the teachers in my study could initially not create a narrative about the mean as a statistical object. Their explanations conflated mean with vague and varied ideas about average and middle in imagined situations. Through focused discussion of the mathematical structure of the mean algorithm they were able to construct narratives about the statistical mean as a constant and a norm or standard to which actual data can be compared. Such understanding of the statistical mean is a big idea in a discourse in which statistics is the science of measuring variation. Averaging in the sense of calculating a mean pervades the structure of more complicated statistical models. Therefore, for discussions of the mean to be statistical rather than informal the mean must be used with conscious consideration of variation and, most importantly, the endeavour to measure variation.

The implication of this study for teachers' statistical discourses for instruction is twofold:

1. Instructional discourse must consciously strive to separate the meanings of average in context and the statistical mean. The intuitive understanding of the mean as the middle value of an interval of average (not extreme) values in a data set should be taken up in a deviation discourse, which raises the need to measure variation. Hence, I draw the attention of teachers to another big idea, namely that statistics is concerned with the *measurement* of variation, rather than merely the description of variation. Without instructional discourses that consciously differentiate between average and mean, meaningful integration discourses about these concepts are not possible.

2. The object conception of the mean as a norm or a standard has the potential to construct clear narratives of the difference between the statistical mean and the arithmetic mean. In arithmetic narratives the mean is understood as a fair share, whilst in statistical narratives the mean is the origin or zero variation value from which variation is measured. I showed how intuitively accessible evening-out procedures can be ordered and used to derive the mean algebraically. The conception of the mean as a norm or standard is thus rich in connections to intuitive reasoning as well as formal statistical reasoning.

Further classroom-based research is needed to understand how teachers develop instructional discourses about measurement of variation and the mean as an origin for such measurement.

Acknowledgements

References

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools.* Unpublished doctoral dissertation. Centre for Science and Mathematics Education, Utrecht University, The Netherlands. Available from http://dspace.library.uu.nl/bitstream/handle/1874/893/full.pdf?sequence=2

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 147–168). Dordrecht: Kluwer Academic Press. http://dx.doi.org/10.1007/1-4020-2278-6_7

Ball, D. (1993). With an eye on the mathematical horizon: Dilemmas of teaching elementary school mathematics. *The Elementary School Journal, 93*(4), 373–397. http://dx.doi.org/10.1086/461730

Ball, D., Thames, M., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389–407. http://dx.doi.org/10.1177/0022487108324554

Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics, 45*(1/3), 35–65. http://dx.doi.org/10.1023/A:1013809201228

Ben-Zvi, D., & Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions and challenges. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 3–16). Dordrecht: Kluwer Academic Press. http://dx.doi.org/10.1007/1-4020-2278-6

Cortina, J., Saldanha, L., & Thompson, P.W. (1999). Multiplicative conceptions of arithmetic mean. In F. Hitt (Ed.), *Proceedings of the 21st Annual Meeting of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 466–472). Cuernavaca, Mexico: Centro de Investigación y de Estudios Avanzados.

Groth, R.E., & Bergner, J.A. (2006). Preservice elementary teachers' conceptual and procedural knowledge of mean, median, mode. *Mathematical Thinking and Learning*, *8*(1), 37–63. http://dx.doi.org/10.1207/s15327833mtl0801_3

Konold, C., & Pollatzek, A. (2004). Conceptualizing an average as a stable feature in a noisy process. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 169–200). Dordrecht: Kluwer Academic Press. http://dx.doi.org/10.1007/1-4020-2278-6_8

Lampen, C.E. (2013). *Learning to teach statistics meaningfully*. Unpublished doctoral dissertation. University of the Witwatersrand, Johannesburg, South Africa. Available from http://wiredspace.wits.ac.za/handle/10539/13349

Makar, K., & Confrey, J. (2004). Secondary teachers' statistical reasoning in comparing two groups. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* Dordrecht: Kluwer Academic Press. http://dx.doi.org/10.1007/1-4020-2278-6_15

Makar, K., & McPhee, D. (2009). Young children's exploration of average in an inquiry classroom. In R. Hunter, B. Bicknell, & T. Burgess (Eds.), *Crossing divides: Proceedings of the 32nd Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 347–354). Palmerston North: MERGA. Available from http://www.merga.net.au/node/38?year=2009

Merriam-Webster. (1913). *Webster's revised unabridged dictionary*. Springfield, MA: Merriam-Webster. Available from http://machaut.uchicago.edu/websters

Merriam-Webster. (2015). *Merriam-Webster Online dictionary*. Available from http://www.merriam-webster.com

Mokros, J., & Russell, S.J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education*, *26*, 20–39. http://dx.doi.org/10.2307/749226

Pfannkuch, M. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning*, *13*(1/2), 27–46. http://dx.doi.org/10.1080/10986065.2011.538302

Porkess, R. (2004). *Collins dictionary of Statistics*. London: HarperCollins.

Sfard, A. (2008). *Thinking as communication*. New York: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511499944

Sfard, A. (2012). Introduction: Developing mathematical discourse: Some insights from communicational research. *International Journal of Educational Research*, *51–52*, 1–9. http://dx.doi.org/10.1016/j.ijer.2011.12.013

Shaugnessy, J.M. (2007). Research on statistics learning and reasoning. In F.K. Lester (Ed.), *Second handbook of research on the teaching and learning of mathematics* (Vol. 2, pp. 957–1009). Charlotte, NC: Information Age Publishers.

Thompson, A.G., Philipp, R.A., Thompson, P.W., & Boyd, B. (1994). Calculational and conceptual orientations in teaching mathematics. In A. Coxford (Ed.), *1994 Yearbook of the NCTM* (pp. 79–92). Reston, VA: NCTM.

Thompson, P.W. (2013). In the absence of meaning. In K.R. Leatham (Ed.), *Vital directions for research in mathematics education* (pp. 57–93). New York: Springer. http://dx.doi.org/10.1007/978-1-4614-6977-3_4

Venkat, H., & Adler, J. (2012). Coherence and connections in teachers' mathematical discourses in instruction. *Pythagoras*, *33*(3), Art. #188, 8 pages. http://dx.doi.org/10.4102/pythagoras.v33i3.188

Watson, J. (2006). Statistical literacy at school. Growth and goals. In A. Schoenfeld (Ed.), *Studies in mathematical thinking and learning*. London: Lawrence Erlbaum Associates.

Watson, J., & Moritz, J. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics*, *37*, 145–168. http://dx.doi.org/10.1023/A:1003594832397

Watson, J., & Moritz, J. (2000). The longitudinal development of the understanding of average. *Mathematical Thinking and Learning*, *2*(1), 11–50. http://dx.doi.org/10.1207/S15327833MTL0202_2

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–248. http://dx.doi.org/10.2307/1403705