LEARNING TO TEACH STATISTICS MEANINGFULLY

Christine Erna Lampen

Student number 12453

A thesis submitted to the Wits School of Education, Faculty of Humanities, University of the Witwatersrand in fulfilment of the requirements for the degree of Doctor of Philosophy

Johannesburg

2012

ABSTRACT

Following international trends, statistics is a relatively new addition to the South African mathematics curriculum at school level and its implementation was fraught with problems. Since 2001 teaching statistics in the Further Education and Training Phase (Grades 10 to 12) has been optional due to lack of professional development of teachers. From 2014 teaching statistics will be compulsory. This study is therefore timely as it provides information about different discourses in discussions of an ill-structured problem in a data-rich context, as well as in discussions of the meaning of the statistical mean.

A qualitative case study of informal statistical reasoning was conducted with a group of students that attended an introductory course in descriptive statistics as part of an honours degree in mathematics education at the University of the Witwatersrand. The researcher was the course lecturer. Transcripts of the discussions in four video recorded sessions at the start of the semester long course form the bulk of the data. The discussions in the first three sessions of the course were aimed at structuring the data-context, or grasping the system dynamics of the data-context, as is required at the start of a cycle of statistical investigation. The discussion in the fourth session was about the syntactical meaning of the mean algorithm. It provides guidelines for meaningful disobjectification of the well known mean algorithm. This study provides insight into informal statistical reasoning that is currently described as idiosyncratic or verbal according to statistical reasoning models.

Discourse analysis based on Sfard's (2008) theory of Commognition was used to investigate and describe discursive patterns that constrain shifting from colloquial to informal statistical discourse. The main finding is that colloquial discourse that is aimed at decision making in a data-context is incommensurable with statistical discourse, since comparison of data in the two discourses are drawn on incommensurable scales – a qualitative evaluation scale and a quantitative descriptive scale. The problem of comparison on a qualitative scale also emerged in the discourse on the syntactical meaning of the mean algorithm, where average as a qualitative judgement conflicted with the mean as a quantitative measurement. Implications for teaching and teacher education are that the development of statistical discourse may be dependent on alienation from data-contexts and the abstraction of measurements as abstract numerical units. Word uses that confound measurements as properties of objects and measurements as abstract units are discussed. Attention to word use is vital in order to discern evaluation narratives as deed routines from exploration narratives and routines.

Keywords

Commognition
Discourse analysis
Exploration routines
Evaluation routines
Informal statistical reasoning
Informal statistical discourse
Statistical reasoning
Statistical discourse
Statistical literacy

DECLARATION

I declare that this thesis is my own unaided work. It is being submitted for the degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

Christine Erna Lampen

14th day of December in the year 2012

To my granddaughter, Isliedh Habika Herbert (aged 7) who said:
"No Ouma, we can't tell stories with numbers, stories need words."

PUBLICATIONS AND PRESENTATIONS EMANATING FROM THIS RESEARCH

Lampen, E. (2009). *The zeroth problem and statistical reasoning*. Paper presented at the The 33rd Conference of the International Group for the Psychology of Mathematics Education.

Lampen, E. (2010). *Structuring contexts for statistical treatment: Initializing statistical reasoning*. Paper presented at the Eighth International Conference on Teaching Statistics.

Lampen, E. (2011). Structuring everyday contexts for statistical treatment: The struggle for shared focus. *African Journal for Research in MST Education, 15*(1), 51-66.

ACKNOWLEDGEMENTS

I wish to thank the following:

My supervisor, Professor Karin Brodie for her patience and unfaltering respect for my working rhythm. Her keen discerning of my word use, while working with me to develop a meaningful narrative in my second language, is commendable and deeply appreciated. Her unfaltering belief that I have something worthwhile to say about teaching statistics meaningfully, both to teachers and the research community, sustained me through those dark days when I could not find my voice. I have great respect for her rigorous engagement with theories of learning and teaching and have benefitted from the challenges she put to me.

My husband, and magical genie of the past five years, Gerhard, who appeared on demand and vanished without a trace when my bidding was done. He increasingly managed to discern imminent bidding and towards the end almost obliterated the need for command. What a pity that genies do not stay forever.

My friend and technical advisor, Mike Reed, for helping me plan and set up the video filming during the course, and for hours of capturing video electronically when cameras did not have memory cards. In addition, for helping me to proof read when I could no longer see errors. My mentor Piet Human who engaged so thoughtfully and patiently with me at the initial stages when focusing was the main challenge. My friend and colleague, Pete van Jaarsveld, who stepped in at the last moment to provide a fresh and agile mind to my research narrative.

The student participants in my research who trusted me enough to allow me to analyse their narratives.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CAPS, Curriculum and Assessment Policy Statement

FET, Further Education and Training (Grades 10 – 12)

ICOTS, International Conference on Teaching Statistics

IIR, Informal Inferential Reasoning

NCS, National Curriculum Statement

PPDAC, Problem, Plan, Data, Analysis, Conclusion

SOLO, Structure of Observed Learning Outcomes

Chapter 1: Motivation, background and problem statement

1.1 Prelude

I was privileged to be part of the Minister of Education's working group that wrote the National Curriculum Statement for the Further Education and Training phase (FET)[1] between 2002 and 2004. I participated in the working group for Mathematical Literacy. At the stage of writing, statistics education had recently been introduced in curricula of developed countries like the United States of America, Australia, New Zealand and the United Kingdom. At the time, no research was available about learning and teaching statistics at school level in South Africa, and curriculum writers had no other option than to study the curricula of these countries and adapt goals and content from these for our curriculum. The situation was extremely unsatisfactory, since no member of our writing group had any experience or knowledge of teaching statistics at school or at university. Few had any training in statistics at all. The situation was slightly better among members of the Mathematics working group alongside whom we worked. Some of the members had experience with the statistics in the then Additional Mathematics curriculum, which was similar to first year university statistics. However, following curriculum reform in countries like the United States of America, the United Kingdom, Australia and New Zealand, our efforts culminated in outcome statements for statistics education that attempted to reflect the nature of statistics as a knowledge field. The statement for *Learning Outcome 4: Data Handling and Probability* read as follows: "The learner is able to collect, organize, analyze and interpret data to establish statistical and probability models to solve related problems" (Department of Education, 2003, p. 50). The Department of Education made *Learning Outcome 4: Data Handling* optional for the first implementation of the FET curriculum in schools (between 2005 and 2012), acknowledging teachers' lack of knowledge in statistics. As a result, very few teachers actually included data-handling in their teaching.

---

[1] The Further Education and Training Phase encompasses Grades 10 to 12, which are the final three years of formal school education in South Africa.

In the wake of the new mathematics curriculum, many endeavours were launched to support the introduction of statistics into the mathematics curriculum. In 2002, South Africa hosted the Sixth International Conference on Teaching Statistics (ICOTS 6), which raised interest in statistics education at school level in our country. Additionally, Statistics South Africa[2] launched initiatives to promote statistics education in schools, such as the *Math4Stats Project*, and sponsored South African involvement with the *Statistics in Schools Project*, which gathers and disseminates data sets about school children in different countries (North & Scheiber, 2008).

Despite all these efforts, a discourse about data-handling as a subject distinctly different from mathematics did not take root in the few classrooms in which it was taught. At most, calculations of mean, median and mode were clothed in context. In 2010 a new curriculum document was published to replace the National Curriculum Statement of 2003. In the new Curriculum and Assessment Policy Statement (CAPS) the term 'statistics' replaced 'data-handling' and the focus was officially placed on learning only a few basic statistical techniques. The rationale behind this unfortunate decision was to make the content to be studied more explicit to teachers and to keep teachers accountable for teaching the exact content.

1.1.1 Reflecting on a personal informal awareness of statistical thinking and reasoning

Worldwide, consensus is growing that the introduction to statistics at school level, aimed at learning techniques, is not advancing statistical literacy, thinking or reasoning (Pfannkuch & Wild, 2004). My experience as a student of Statistics 101 many years ago at university confirmed for me the inadequacy of 'traditional' statistics education pedagogy. I had to study formula after formula and rely on key words in problems in order to know when to apply a given formula; only to arrive at a numerical answer that was all but meaningless to me. Learning statistics and passing the course did not help me to reason statistically or to interpret statistics critically. When I was asked in 2005 to teach statistics to in-service teachers, I could only compare my own ways of reasoning about statistical methods and results with what I came to understand as mathematical reasoning, and with my own everyday reasoning. I thus

[2] Statistics South Africa is the official statutory statistical authority of South Africa.

set out to familiarise myself with definitions of statistical reasoning and thinking in statistics education literature, deliberating on how to teach statistics differently from mathematics. There was a large corpus of tasks and questions available in journals dedicated to statistics education, but I struggled to find guidance on how to 'talk statistics' with my students.

1.1.2 Mathematics and statistics

Reflecting on the reasoning and communication skills I developed as a mathematics teacher, and comparing the rigorous logic and proof of mathematics to the kind of reasoning that I need when I am confronted with data from real life situations, I became increasingly aware of subtle differences. One of these differences is: in a mathematics argument, I warrant the argument with the definition of the mathematical object about which I am reasoning, and I experience closure and certainty. For example, the argument about the nature of the sum of two odd numbers is settled once I define odd and even numbers structurally in relation to each other. The rest is demonstration. In statistics, hardly anything is always exactly *so*. I can warrant a data-based argument with the properties of the statistical tool I used, for example the choice of a confidence interval; but my decisions are made in context – where absolute certainty evades me. With this study I wanted to deeper explore differences between reasoning in school mathematics and school statistics.

1.1.3 Everyday reasoning and statistics

In my determination to make conceptual sense of statistics, I came to realise the genuine value of a statistical frame of mind, and of how such a frame of mind relates to one's world view and epistemology. I realised that my own intuitive statistical reasoning was based on frequency, similar to that of my teenage children claiming a statement to be true, due to the 'fact' that "everyone does it". I further became aware of the tug of personal experience and worldview on reasoning, through conversations with students in my class. One example will suffice. I once provided my statistics class with a set of data that gave the number of physicians in a sample of developing and developed countries, as well as the populations of those countries. The question to be answered was: Do we have enough doctors in South Africa, compared with other developing countries? One student immediately declared that we did not. I asked for substantiation and he replied that South African doctors leave the country, because of affirmative action. I was immediately aware that this student was not talking statistics and would not yield to comparison of numbers between countries. As I will discuss in Chapter 2, such reasoning is widespread in statistics classrooms, and is labelled *idiosyncratic* in models of statistical reasoning and thinking. Pushed for stronger evidence, the same student in the following session provided the class with numerical data about the number of doctors that had indeed left South Africa during the previous four years. He offered the total number of doctors that had left South Africa in the past 10 years as his final warrant. In his everyday reasoning, this student used data to tell a story about a cause, namely the diminishing numbers of doctors in South Africa.

Something else was at play, though. I realised that my use of the term “enough” was not a signal for comparison of numbers of doctors between countries as I intended, but rather a signal for decision making in context, enough was interpreted as enough-for-South Africa, and the answer did not need statistics, since anecdote provided the answer. I realised that everyday reasoning may not be potentially statistical, even if statistical information is available. Reasoning about data in contexts has to be preceded by a phase of alienation, that is, literally taking oneself out of the situation, in order to make statistical reasoning possible. In addition, one needs to mentally re-structure one’s spontaneous understanding of the context in order to find a way to compare between countries. It seems to me as if one has to open up the context to think about aspects such as which doctors to count (those in public service, specialists or family doctors?); what the criterion would be for enough or not; and what other factors to take into account (do countries with strong systems of clinics and public health care need as many doctors as those without?). On reflection, it became clear to me that this kind of pre-structuring of contexts is not supported by current mathematics education. This, I believe, is due to the rather closed, deterministic and exact nature of mathematics at school level. If students and learners are used to *doing mathematics* in already alienated, abstracted contexts with the simple purpose of providing a result, we can only expect them to calculate totals, or perhaps averages; rather than engage in data stories that compare numbers between reference classes.

1.1.4 Worldview and Statistics

A further aspect that plays a role in my interest in statistics education is the evolving personal realisation that much can be tolerated and understood if we realise that the world we live in is largely non-deterministic, and that variability is ubiquitous. Hacking (1990) called this the most decisive conceptual event of the twentieth century. The stable and regulated society that I grew up in around 40 years ago gave me little opportunity to experience non-determinism or variation. Life seemed rule-based and as predictable in the rules as in the consequences of breaking them. In my opinion, many learners today grow up experiencing the world mainly as non-deterministic, and often chaotic. Learners are bombarded with consumer choices, conflicting information, and with news and images of countries at war rearranging their borders and alliances. Political heroes and other role models at home lose their status and constituencies overnight as the media informs and influences opinions. Very often learners fall back on social and emotional imperatives when they have to make decisions and act in non-deterministic situations, including situations that carry personal risk (Watson, 1998). There is a need for teachers and teacher educators to learn to develop narratives about statistics that will allow learners of statistics to extend their worldviews while they learn to reason statistically about variables in a non-deterministic world. My view finds support in Lesser and Blake's (2007, p. 2) call for learners to participate in the messiness of statistical situations in order to learn the tools to develop arguments based on statistics as evidence; a skill which Lesser and Blake said adolescents, and even many adults, find difficult. Experiences that legitimise the narratives on which we base our reasoning and arguments in mathematics, or in statistics, or in everyday life, are not widespread in our school environment today. Yet, this is the kind of critical and social contribution that meaningful teaching and learning of statistics can bring to the school experience.

I undertook the study reported on here as a teacher educator in search of understanding statistics as a way of talking about the world. The participants in my study were a group of in-service mathematics teachers involved in an introductory statistics course that I was teaching.

1.2 Rationale

1.2.1 Comparison between international curriculum goals and statistics in the South African curriculum

Descriptive statistics (as opposed to theoretical statistics) has become part of the mainstream of mathematics instruction across the world (Moore, 2004; Wild, 2006), as a response to dealing with non-deterministic situations. Despite limited time allocation in curricula, international consensus is growing that the most important goals for statistics education are the development of statistical literacy, reasoning and thinking (Ben-Zvi & Garfield, 2004). Table 1 presents current writing on three strands of statistical deliberation, namely statistical literacy, reasoning and thinking as described by Ben-Zvi and Garfield (2004). The table compares concepts and skills of the three main strands as goals for statistics education. At first glance, Ben-Zvi and Garfield seemed to assign practical, procedural fluency to statistical literacy. Yet, the manual skill of organising and representing data is of little importance in a world of computer applications. A statistically literate person will have to make decisions about appropriate presentations for different kinds of data, in different kinds of investigations. Without knowledge of how to use statistical representations as tools for communication in context, statistical literacy is incomplete. According to Ben-Zvi and Garfield, an imaginary statistical *reasoner* must make sense of and fully interpret statistical results, and thereby use statistics critically. Critical use of statistics points to the need to weigh different possibilities for analysis, as well as a need for extensive general knowledge about situations that yield the data for analysis. It would also call for a level of statistical *thinking*, such as an understanding of the pervasiveness of variation and the nature of sampling. I therefore reason that statistical literacy, reasoning and thinking are not qualitatively different to the way we are used to viewing mathematical literacy (or numeracy) and mathematical reasoning and thinking. The development of statistical thinking may require interrelating procedures, concepts and ways of reasoning across the three main strands that describe the goals for statistics education. Indeed, Garfield and Ben-Zvi (2008) argue that statistical literacy is a goal that may only be reached in adulthood, and that statistical reasoning and thinking probably require more statistics education than at school level. However, the quest has to start in school.

Table 1: International goals for learning statistics, adapted from Ben-Zvi & Garfield (2004)

Main strands	Skills and Concepts
Statistical literacy Has no formal definition as yet.	Statistical literacy skills include being able to: • organize data • construct and display tables • work with different representations of data; • understand: — concepts — vocabulary — symbols — probability as a measure of uncertainty
Statistical reasoning May be defined as the way people reason with statistical ideas and make sense of statistical information. Reasoning means understanding, and being able to fully interpret statistical results.	Statistical reasoning involves making interpretations based on: • sets of data • representations of data • statistical summaries of data • connecting one concept to another (e.g. centre and spread) • combining ideas about data and chance
Statistical thinking Involves an understanding of why and how statistical investigations are conducted and the "big ideas" that underlie statistical investigations. Statistical thinkers are able to critique and evaluate methods used and results of a statistical study.	Statistical thinking include understanding: • the omnipresent nature of variation • when and how to use appropriate methods of data analysis such as numerical summaries and visual displays of data • the nature of sampling • how we make inferences from samples to populations • why designed experiments are needed in order to establish causation • how models are used to simulate random phenomena • how data are produced to estimate probabilities • how, when, and why existing inferential tools can be used to aid an investigative process • how to utilize the context of a problem in forming investigations and drawing conclusions • the entire process of statistical investigation (from question posing, to data collection, to choosing analyses, to testing assumptions)

As I indicated before the South African National Curriculum Statement (2003) was written with cognisance of the intended curricula in other parts of the world. The elaborated statement, that describes the learning outcome for Data-Handling in the National Curriculum Statement for FET (Department of Education, 2003, p. 14) follows below:

> Learning Outcome 4: Data-Handling and Probability
>
> *The learner is able to collect, organise, analyse and interpret data to establish statistical and probability models to solve related problems.*
>
> […] Measures of central tendency and spread will be explored. A basic appreciation of the difference between data that is normally distributed about a mean and data that is skewed, will be developed. Learners will become critically aware of the deliberate abuse in the way data can be represented to support a particular viewpoint. Learners will carry out practical research projects and statistical experiments […] The basic understanding of probability and chance gained at General Education and Training level will be deepened so that, for example, learners can compare the actual odds in winning popular games of chance with the odds offered by gaming houses. A basic understanding of the way the probability of everyday events can be calculated and used in prediction will be developed. Wherever possible, contexts that are investigated will focus on human rights issues, inclusivity, current matters involving conflicting views, and environmental and health issues.

The exposition of the goals for statistics in the FET band in the NCS provides a similar narrative to that of the international goals. It presents a vision of learners who are engaged in investigation and critical deliberation, using statistical objects in their activity. The curriculum was ambitious, in that it required the deep conceptual understanding that underlies statistical thinking; it did this by making critical awareness, research projects, and predictions in context, explicit and assessable goals. Internationally the translation between intended curricula and implemented curricula is known to be less than isomorphic (see for example Usiskin and Dossey, 2004). Teachers who are unsure in the wake of curriculum changes tend to turn tasks based on new curriculum demands into "something they already know how to do" (Tyack & Cuban, 1995, p. 64). In South Africa the implementation of statistics education according to the goals of the NCS did not get off the ground and in 2010 a highly specified curriculum and assessment policy statement was issued. In the Curriculum and Assessment Policy Statement (CAPS) (Department of Basic Education, 2010), statistics is allocated a total of seven weeks of study, four and a half hours per week, spread over Grades 10, 11 and 12. The mark allocation for descriptive statistics in Grade 12 assessment is, at most, 23 marks out of 150. Not surprisingly then, the CAPS, which is now the only curriculum document that South African teachers are allowed to refer to, treats statistics in an extremely limited way when compared to the international goals in Table 1. In Table 2, I present the complete references to statistics for Grades 10 to 12.

Table 2: Curriculum and Assessment Policy Statement (Grades 10 – 12) references to statistics (Department of Basic Education, 2010)

<table>
<tr><th>No. of weeks</th><th>Curriculum statement</th><th>Clarification</th></tr>
<tr><td>Grade 10:
2 weeks

(p.17)</td><td>1. Revise measures of central tendency in ungrouped data.
2. Measures of central tendency in grouped data: calculation of mean estimate; identification of modal interval and interval in which the median lies.
3. Revision of range as a measure of dispersion and extension to include percentiles, quartiles, interquartile and semi-interquartile range.
4. Five number summary and box-and-whisker diagram.
5. Use the statistical summaries (measures of central tendency and dispersion), and graphs to analyse and make meaningful comments on the context associated with the given data.</td><td>Example:
The mathematics marks of 200 grade 10 learners at a school can be summarised as follows:

<table><tr><td>Percentage obtained</td><td>No. of candidates</td></tr><tr><td>0≤ x<20</td><td>4</td></tr><tr><td>20≤ x<30</td><td>10</td></tr><tr><td>30≤ x<40</td><td>37</td></tr><tr><td>...</td><td>...</td></tr></table>
1. Calculate the approximate mean mark for the examination.
2. Identify the interval in which each of the following data items lies:

2.1. the median
2.2. the lower quartile
2.3. the upper quartile, and
2.4. the thirtieth percentile.</td></tr>
<tr><td>Grade 11:
3 weeks

(p.39)</td><td>1. Histograms
2. Frequency polygons
3. Ogives (cumulative frequency curves)
4. Variance and standard deviation of ungrouped data
5. Symmetric and skewed data
6. Identification of outliers</td><td>Comments:
• Variance and standard deviation may be calculated using calculators.
• Problems should cover topics related to health, social, economic, cultural, political and environmental issues.
• Identification of outliers should be done in the context of a scatter plot as well as box-and-whisker diagrams.</td></tr>
</table>

<table>
<tr>
<td>Grade 12:
1 week

(p.48)</td>
<td>1. Revise symmetric and skewed data.
2. Use statistical summaries, scatterplots, regression (in particular the least squares regression line) and correlation to analyse and make meaningful comments on the context associated with given bivariate data, including interpolation, extrapolation and discussions on skewness.</td>
<td>Example:
The following table summarises the number of revolutions x (per minute) and the corresponding power output y (horse power) of a diesel engine:
<table>
<tr><td>x</td><td>400</td><td>500</td><td>600</td><td>700</td><td>750</td></tr>
<tr><td>y</td><td>580</td><td>1030</td><td>1420</td><td>1880</td><td>2100</td></tr>
</table>

1. Find the least squares regression line y = a + bx
2. Use this line to estimate the power output when the engine runs at 800 rpm
3. Roughly how fast is the engine running when it has an output of 1200 horse power?</td>
</tr>
</table>

Even a superficial discursive comparison between the CAPS and international goals reveals the paucity of statistical reasoning demands in the South African curriculum statement. Whereas the international goals are a narrative on human deliberation, decision making and action, the South African curriculum emphasises only one discourse: a discourse about calculation of statistics and simple reading of graphs. The cursory requirement to "make meaningful comments on the context" is vitiated by the examples given. If the South African statistics curriculum has to shift closer to international goals in future, professional development of teachers cannot be limited to the specifics in the CAPS. My study is aligned with the curriculum goals for statistics education as presented in the National Curriculum Statement (2003) and provides much needed information about informal statistical reasoning of South African teachers and learners. This knowledge will help to engage meaningfully with novice statistical reasoners.

1.2.2 The implication of the curriculum goals of the NCS (2003) for the teaching of statistics at FET level

The NCS (2003) curriculum goals for statistics have important implications for the teaching of statistics by mathematics teachers. They imply that mathematics teachers must be able to reason statistically and understand how to foster the development of statistical reasoning through their teaching. This is a daunting task, more so because most mathematics teachers in South Africa have not had any statistics courses in their professional training, and have limited opportunities for in-service development of their own statistical reasoning. They mainly have their pedagogical knowledge of teaching mathematics as a subject to draw upon for the teaching of statistics.

Despite the endeavour to reform the practice of teaching mathematics since South Africa introduced the NCS, many teachers struggle to teach mathematics in ways that elicit meaningful (theoretical) conversation about mathematical concepts like pattern and structure, let alone modelling and discussion in context. Where teachers do teach mathematics in ways that involve discussion and group work, they find it difficult to take up learner thinking as part of the teaching process (Brodie, 1999; Brodie, Lelliott, & Davis, 2002). The implications for statistics education are clear – if teachers struggle to develop meaningful mathematical discourse they will also struggle to develop statistical discourse. There are currently no guidelines for teachers about the nature of statistical discourse or the difference between statistical reasoning and mathematical reasoning.

The difference between mathematics and statistics seems to lie mainly in the nature of the objects of analysis. Cobb and Moore (1997) explain that numbers in statistics are always numbers in context. In mathematics context obscures structure and in statistics the opposite is true. Numbers without context are meaningless. It follows that reasoning in statistics involves reasoning in context. In contrast, mathematical reasoning often involves stripping away context to talk about the numbers only. Burrill and Biehler (2011) discuss various tensions between mathematics and statistics. These include the use of standard magnitudes in mathematics as compared to error prone measurements in statistics; the expectation that variables perfectly fit function models in mathematics, while measurements are scattered around a function model in statistics; and the tension between the deductive certainty that is attainable in mathematics and inductive reasoning as the basis of faith in statistical conclusions.

One of the biggest challenges to teachers of statistics is thus to conceptualise the subject as different in nature from mathematics, and therefore to teach statistics in a different way from traditional methods of teaching mathematics that currently prevail (Burrill & Biehler, 2011; Garfield & Ben-Zvi, 2008). In doing so, teachers can expect to be confronted with learners who intuitively equate statistics education with mathematics education and expect the focus in statistics to be on exact numbers, computations, formulae and on one right answer. This challenge in the attitude of learners is reported in statistics education research. Ben-Zvi and Garfield (2004, p. 4) indicate that "[learners] are uncomfortable with the messiness of data, the different possible interpretations based on different assumptions, and the extensive use of writing and communication skills." Without direct intervention in teacher education that supports teachers to enrich the curriculum, South African teachers will continue to avoid the tension between mathematical reasoning and statistical reasoning that remains unaided by the current curriculum, and teach only the procedures of statistics.

1.2.3 The cycle of statistical as pedagogic guideline for teaching and learning statistics

Pfannkuch (2008) argues that teaching statistics meaningfully requires teaching the complete cycle of statistical inquiry. Such a holistic approach to data-handling encompasses and integrates the notions of statistical literacy, reasoning and thinking. Wild and Pfannkuch (1999) structured the process of statistical inquiry with a model of an investigative cycle and an interrogative cycle; along with the dimensions of thinking and the dispositions required for statistical inquiry. Figure 1 presents Wild and Pfannkuch's complete model, which is widely endorsed to scaffold the teaching and learning of statistics at school level.

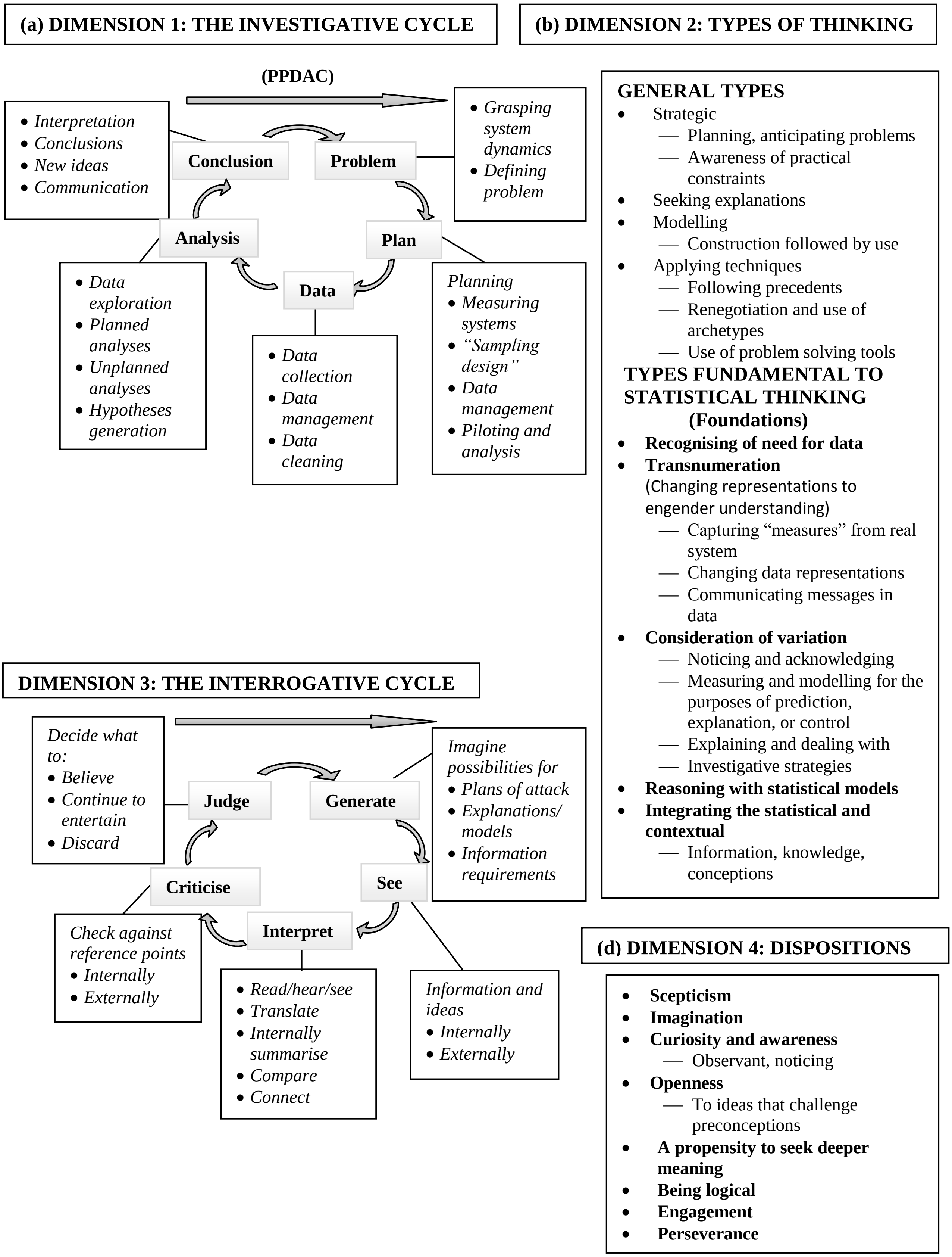


Figure 1: The dimensions of statistical inquiry (Wild & Pfannkuch, 1999, p. 226)

Of specific interest in this study is Dimension 1, the investigative cycle, as presented in Figure 1. The investigative cycle is mostly referred to by the acronym PPDAC, which serves as a reminder of the processes of problem formulation, planning, data gathering, analysis, and conclusion in the cycle. A short deliberation will confirm for the reader that the interrogative cycle and dispositions are at work throughout the investigative cycle; and the separation of the cycles is of theoretical, rather than practical importance. The investigative cycle commences with the demand to identify and define or operationalise a problem that needs a statistical answer. As indicated in the model, defining a problem for statistical analysis in turn requires learners of statistics to grasp the system dynamics of the context that holds the problem. In contrast to traditional ways of teaching mathematics, the context is problematised rather than simplified, in the PPDAC cycle. The pedagogic implication for teaching statistics according to the PPDAC cycle, is thus to introduce contextual complexity into tasks and discussions even before the participants are acquainted with statistical tools. In particular, teachers have to be able to mine data rich contexts for opportunities to develop statistical reasoning, and follow through with the complete investigative cycle through tasks and classroom discussions.

A pervasive finding in mathematics education research on teacher development is that a lack of content knowledge and conceptual understanding have direct consequences upon the quality of teaching (Hiebert & Grouws, 2007; Phillip, 2007). It stands to reason that teachers who do not reason statistically themselves will have grave difficulty ‘talking statistics’ in class. While international studies about teachers’ understanding of, and reasoning with, statistical concepts are increasing, Wessels (2011) indicates that there are few studies about statistical reasoning of South African teachers or high school learners. Gierdien (2008) investigated the intuitive probabilistic reasoning of in-service teachers. He shows how taking up the intuitive subjective probability reasoning of the teachers in thorough discussions of tasks shifted their awareness to the reasoning their learners are likely to bring to classroom discussions. To date only two post graduate studies, related to statistics education at school in South Africa, have been done. Wessels (2006) studied the types and levels of data modelling of Grade 4 to 7 learners in her doctoral research. The second study, a master’s thesis of limited scope, investigated the role of visualisation in data-handling in Grade 9 (Makina, 2009). In a recent study, Wessels and Nieuwoudt (2011) profiled the statistical knowledge, beliefs and confidence of a group of South African mathematics teachers. Their findings point to teachers having difficulty applying statistical knowledge in social contexts, and in interpreting newspaper reports. The authors conclude that there is a pressing need for professional education courses that are specifically aimed at the discussion and development of statistical reasoning.

1.3 Problem statement

My argument up to this point is based on the view that learning statistics means purposeful development of statistical reasoning through discussion of authentic data and data-contexts. My study was guided by the following main research questions:

a) What kinds of reasoning do teachers bring to a study of statistics, where discussion and argumentation about authentic data related contexts and issues are foregrounded?
b) How does teachers' reasoning develop during discussion?
c) How does reasoning about authentic data based issues afford or constrain the development of statistical reasoning?

I conducted a case study of a class of teachers who were students in an introductory statistics course at the Wits School of Education. I chose my own class as research site, since there were no other similar courses for teachers at the time in South Africa. The study involved a discursive analysis of two key group discussions in the course and it describes the reasoning that emerged.

1.4 Outline of the study

In this chapter I have discussed international goals for learning and teaching statistics and reflected on the need to develop ways to 'talk statistics' in the classroom. I reflected on the relation between teaching and learning statistics, and teaching and learning mathematics. In particular, I indicated that the role of context is important in statistical investigation. I am particularly interested in the influence of colloquial discourse on the development of statistical discourse. I use the term "colloquial discourse" in the same sense as Vygotsky (1986) used everyday reasoning to indicate spontaneous or informal discourse of the kind that participants do not formally develop in school. Figure 2 provides a Venn diagram, showing my perspective of the relationships between the discourses.

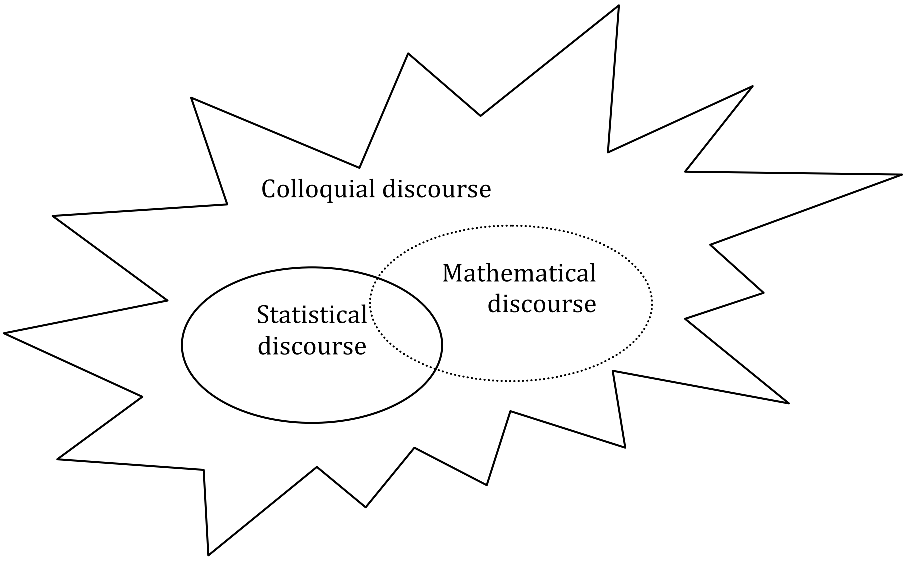


Figure 2: An initial framework for the relationship between colloquial, statistical, and mathematical discourses

I use Figure 2 to suggest that colloquial discourse is less rule-based and less structured than either statistical or mathematical discourse; yet colloquial discourse yields information about persistent ways of reasoning. Mathematical and statistical discourses have strong formal structures, yet both are related to everyday discourse – historically and through their applications. The dotted boundary of mathematical discourse serves to indicate that although I acknowledge the influence of mathematics on statistics, I do not take up a complete analysis of the relationship as part of this study.

With these relationships in mind I proceed as follows. My literature review is presented in two parts. In Chapter 2 I present a review of literature about statistical reasoning in education settings and the role of context in such reasoning. I reflect on findings that refer to tensions between everyday reasoning and statistical reasoning. In Chapter 3 I then review literature about informal, everyday reasoning. In Chapter 4 I outline the theoretical framework for this study, namely Anna Sfard's Theory of Commognition, which defines thinking as communicating, and mathematics as a discourse, and allows me to further refine my research questions. In Chapter 5 I describe my research design and methodology, based on guidelines for commognitive research. I present my analysis in two parts: Chapters 6, 7 and 8 contain the analysis of data in relation to classroom discussions of a data-based problem in an everyday context. In Chapter 6 I analyse the students' written contextual questions and in Chapter 7 I analyse their in depth discussions of the questions. The analysis in Chapter 8 concerns the same problem as in the previous two chapters, but focuses on the students' discursive structuring of relationships between contextual variables. The second part of my data analysis focuses on discussions about the meaning of the mean algorithm and is presented in Chapter 9. I discuss my conclusions the implications of my research for research and practice in Chapter 10.

My commognitive study of discussions in a statistics class is novel, and sheds new light on those aspects of statistical reasoning that are currently framed by the literature as merely idiosyncratic. Hence, this study allows me to explicate existing models of statistical reasoning, as well as provide useful guidelines for understanding the kind of talk taking place in introductory statistics classes.

Chapter 2: Literature review: Reasoning and Statistics Education

2.1 Introduction

Although data gathering by governments is as old as biblical times, statistics was recognized as a science only in 1834 (Rao, 1999). Statistics education as a field of study is only forty years old. The first international conference of teachers of statistics (ICOTS) was held in 1982 in Sheffield in the United Kingdom and the establishment of the International Association for Statistics Education followed in 1991. The impetus for drawing together statistics educators and statistics education researchers was the "increasingly strong call from practising statisticians for statistical education to focus more on statistical thinking" (Pfannkuch & Wild, 2004, p. 17). With the establishment of ICOTS the quest for reform in statistics education was taken on. The reform movement is aimed at developing statistical literacy, thinking and reasoning by engaging learners with discussion and interpretation of real-world problems and issues. Teachers' lack of experience was identified as a major obstacle from the onset. As late as 2004, Pfannkuch and Wild argued that the problem of teacher inexperience is exacerbated by the lack of a coherent body of research about statistical thinking and claimed that even experienced statisticians are challenged to find adequate pedagogies for teaching statistical thinking. Research related to statistics education has proliferated since the turn of the century, and is starting to yield information about learner reasoning as well as guidelines for pedagogy (Batanero, 2004). Three dedicated statistics education journals (*Statistics Education Research Journal, Journal of Statistics Education,* and *Teaching Statistics*) published 138 research articles between 2005 and 2009. In the same period, the International Association for Statistical Education has indexed 24 doctoral dissertations in the field (Van der Merwe & Wilkinson, 2011).[3] Analysis of the content of these publications showed that 25% were focused on teaching and learning activities and strategies. Another 20% were aimed at statistical reasoning about a specific statistical concept

[3] Twelve doctoral dissertations have been added to the index between 2010 and 2012.

(Van der Merwe & Wilkinson, 2011). Researchers have described misconceptions and limiting heuristics[4] in relation to the use of statistical tools and procedures, which seem to span both age and context. Teachers are called upon to take up learner reasoning and co-develop a shared scientific or literate statistical discourse. How teachers are to achieve such a discourse in statistics classes is not clear, and there is a need to explore and describe the development of statistical thinking in classroom discussions in order to develop theoretical guidelines for teaching. However, no studies have yet been conducted that analyse and describe the development of teachers' reasoning from colloquial reasoning in context to statistical reasoning in context. A focus on the introduction of such reasoning allows for an intersection to be drawn between everyday reasoning, mathematical reasoning and statistical reasoning. I commence with a review of research about statistical reasoning, and show that the demand of the contexts to which the data pertain evokes everyday reasoning routines. In the next chapter I review research about everyday reasoning.

2.2 Models of statistical reasoning

In statistics education research, statistical reasoning has been defined as "the way people reason with statistical ideas and make sense of statistical information … "and includes understanding and being able to explain statistical processes and being able to fully interpret statistical results" (Ben-Zvi & Garfield, 2004, p. 7). In these models, statistical information pertains to graphs or calculated statistics, such as measures of central tendency and spread of data sets. Statistical thinking is a wider construct that involves meta-cognitive skills; such as a sense of scepticism, and a feeling for when and how statistical tools can be used appropriately. Despite efforts from researchers to define statistical reasoning and statistical thinking as separate processes, the terms are often used interchangeably.

Based on the definition of statistical reasoning above, research about statistical reasoning in education has centred on the way people reason about formal statistical concepts. Various

---

[4] Misconceptions in the mathematics education literature refer to generally held, erroneous understandings of concepts. For example "multiplication makes bigger". Gigerenzer & Gaissmaier (2011, p. 454) have written that "A heuristic is a strategy that ignores part of the information, with the goal of making decisions more quickly, frugally, and/or accurately than more complex methods." Limiting heuristics lead to erroneous reasoning and decision making when compared to statistical principles or formal logic.

studies investigated reasoning about aspects such as data analysis (Ben-Zvi, 2004); variation (Reading & Shaughnessy, 2004); distribution (Bakker & Gravemeijer, 2004); co-variation (Moritz, 2004); and samples and sampling distributions (Chance, DelMas, & Garfield, 2004). From such research, a general model of statistical reasoning was developed by Garfield, as represented in Table 3.

Table 3: General Model of Statistical reasoning (Garfield, 2002, pp. 8; 11)

Model of Statistical Reasoning
Level 1. Idiosyncratic reasoning. The student knows some statistical words and symbols, uses them without fully understanding them (often incorrectly), and may scramble them with unrelated information. For example, students have learned the terms 'mean', 'median', and 'standard deviation' as summary measures; but use them incorrectly (comparing the mean to the standard deviation, or making judgements about a good mean or standard deviation).
Level 2. Verbal reasoning. The student has a verbal understanding of some concepts, but cannot apply this to actual behavior. For example, the student can select or provide a correct definition, but doesn't fully understand the concepts (for example, why the mean is greater than the median in positively-skewed distributions).
Level 3. Transitional reasoning. The student is able to correctly identify one or two dimensions of a statistical process without fully integrating these dimensions; for example that a larger sample size leads to a narrower confidence interval; or that a smaller standard error leads to a narrower confidence interval.
Level 4. Procedural reasoning. The student is able to correctly identify the dimensions of a statistical concept or process, but does not fully integrate them or understand the process. For example, the student knows that correlation does not imply causation, but cannot fully explain why.
Level 5. Integrated process reasoning. The student has a complete understanding of a statistical process, coordinates the rules and behavior. The student can explain the process in his or her own words with confidence. For example, a student can explain what a 95% confidence interval means in terms of the process of repeatedly sampling from a population.

The general model is based on the Structure of Observed Learning Outcomes (SOLO) taxonomy, a general developmental model for the assessment of learning (Biggs & Collis, 1982).

Garfield's model might have foregrounded formal statistical content and skills to the extent that reasons why participants reason in the described ways are buried. The complexity of reasoning at Levels 1 and 2 is only hinted at by the terms *idiosyncratic* reasoning and *verbal* reasoning. To cite Sfard (2008, p. 16): "The words *idiosyncratic* and *verbal* refer not so much to the inner coherence of students' thinking as to possible disparities between students' conception and the generally accepted versions of the same ideas". Similarly, the influence of the data-context on the decision making in context is under-described in Level 5, *integrated process reasoning*. The fifth level of reasoning described in Table 3 constitutes correct or ideal statistical reasoning; while the preceding levels constitute partially-correct reasoning by different criteria. However, there is no evidence in the literature that the levels in this general model of statistical reasoning describe a developmental trajectory for its development. SOLO based models like that of Garfield must be supplemented by research into the development of statistical reasoning during extended reasoning processes. Brickell, Ferry and Harper (2002, p. 67) describe properties of effective reasoning as follows:

> Effective reasoning requires the ability to develop arguments, assess the validity of the argument in generating and testing hypotheses, judge the credibility of assertions made during the problem solving process, identify possible directions for action, and think through the consequences of choosing a particular direction of action.

Comparing this description of effective reasoning to the description of statistical reasoning in Garfield's model emphasises the limited role of judgement in the model of statistical reasoning. Statistical thinking and reasoning is embedded in holistic contextual reasoning (Wild & Pfannkuch, 1999) and should also include processes and abilities such as those described by Brickell et al. above. Bakker and Derry (2011) concur that making judgements is an innate property of human reasoning, and indicate that judgements as informal inferences[5] are embedded in a web or space of reasons, which include context; the texts that

[5] Informal inference in statistics refers to decisions and judgements made on the basis of exploratory statistical information, rather than formal probabilistic models.

convey the problem; and the rules of statistics as a specific knowledge base. But Garfield's research on assessing statistical reasoning suggests that statistics instructors do not specifically teach students how to use and apply methods of reasoning. Garfield observes "instead, most instructors tend to teach [statistical] concepts and procedures, provide students with opportunities to work with data and software, and hope that [statistical] reasoning will develop as a result. However, it appears that reasoning does not actually develop in this way" (Garfield, 2002, p. 3).

A more integrated picture of statistical reasoning is possible if we consider the general model of statistical reasoning against the background of the PPDAC[6] process as represented in Chapter 1 (see Figure 1). Statistical reasoning and thinking is required in both the investigation and interrogation cycles. It stands to reason that the nature of statistical reasoning in the different cycles may be different. For example, it is plausible to argue that idiosyncratic and context-based reasoning finds relevance at the start and conclusion of the investigative cycle, when the contextual reality must be understood.

2.3 Statistical reasoning and context

An extensive body of research about the influence of context knowledge on informal inferential reasoning in statistics has followed the sixth research forum of The International Collaboration for Research on Statistical Reasoning, Thinking and Literacy in 2009. This body of research focuses on making informal judgements based on statistical exploration of context-based questions. Informal inferential reasoning (IIR) is generally promoted for its potential to support learners' general statistical reasoning (Madden, 2011). Pfannkuch argues that the statistical learning-experience context has to coordinate the data-, or task-context, the discussions between learner and teacher, and the previous statistical knowledge of the learner. This triadic framework is indispensible "for the interrogations, conversations, and constructions inherent in (IIR)" (Pfannkuch, 2011, p. 43). Yet, context is generally seen as problematic in statistical reasoning. Context has been described as a tyranny, overriding the more abstract reasoning required for statistical analysis and informal inference (DelMas,

---

[6] PPDAC is the acronym for the five stages of statistical investigation proposed by Wild & Pfannkuch (1999), namely problem formulation, planning, data gathering, analysis and conclusion.

Garfield, & Zieffler, 2006). A different perspective is provided by Langrall, Mooney and Williams' (2005) research about students' use of context knowledge in the conclusion and interpretation phase of statistical inquiry. They found that despite familiarity with the data-context the majority of secondary students in their study failed to integrate their mathematical-statistical solutions with the context and provided only the numerical results. It seems that sound contextual reasoning is necessary but not sufficient for sound reasoning about the same situation as a data-context and that statistics may also wield tyranny over context.

Problem formulation was studied as the first step of the planning stage in the statistical investigation cycle by various researchers (Konold & Higgins, 2003; Lehrer & Romberg, 1996; Pfannkuch & Horring, 2005) and hence as a link between the data-context and the purpose of statistical investigation. Arnold (2009) categorises questions according to their goals: 'posed questions' are formally structured and 'asked questions' are spontaneous and arise continuously in order to interrogate reasoning throughout the investigation. Posed questions emerge from consideration of a contextual problem, and can be further categorised as survey questions (questions asked to get data) and investigative questions (questions asked of the data). Arnold (2009) stresses the importance of keeping the connection between the context-based survey questions and the more abstract and statistically supported investigative questions. Investigative questions during data analysis are categorised as comparison questions and summary questions. She explains: "Summary questions are posed when a description of the data is needed and are usually about a single data set. Comparison questions are posed for comparing two (or more) subsets of data across a common variable, for example, male, female, young and old. Relationship questions are posed for looking at the interrelationship between two paired variables" (Arnold, 2009, p.2). The Grade 10 students in Arnold's study were more successful in posing suitable comparison questions than summary questions. Arnold does not comment on reasons for this difference, but it is plausible to suggest that contextual reasoning tends to be causal and therefore the comparison questions are likely to be an indication of seeking for explanations of the observed variability.[7]

[7] I will take up Arnold's classification of questions in Chapter 6 in order to analyse my students' contextual questions in order to understand the data-context.

Gil and Ben-Zvi (2011) argue that the role of explanations in IIR warrants more attention. Indeed, evidence and explanation is at the core of reasoning and argumentation in statistics (Lipton, 2004). Gil and Ben-Zvi studied the nature of explanations in a case study of Grade 6 learners' IIR. The researchers describe explanations of statistical inferences as "an account of the *why* and the *how* of the inference" that provide information about statistical reasoning (Gil & Ben-Zvi, 2011, p. 91). The learners' explanations changed according to different kinds of reasoning. In particular, *descriptive explanations* that use data as appropriate evidence for an inference, were observed in relation to evidential reasoning,[8] while *conflict resolution explanations* (resolving conflict between expectations and observed data) and *reasonableness explanations* (explaining why an inference is (un)reasonable), were observed to go hand in hand with abductive reasoning.[9] Context and statistical aspects were interwoven, with context commonly used as the source of evidence in reasonableness, abductive, and conflict-resolution explanations (Gil & Ben-Zvi, 2011, p. 102). The tendency to rationalise data at hand, that is, to construct abductive arguments, either by citing facts about the context or by creating plausible explanations, was also observed by Langrall and her colleagues (Langrall, Nisbet, Mooney, & Jansem, 2011). I take this as early evidence that context evokes everyday reasoning routines, which may support or thwart statistical reasoning. Madden (2011) studied high school teachers' informal inferential reasoning in a technology-rich environment. She classifies tasks as being contextually provocative; statistically provocative; technologically provocative; or combinations of these categories. Amongst other gains, contextually provocative tasks elicited rich argumentation, which assisted the teachers in her study in making conjectures and informal inferences guided by contextual considerations. Madden (2011, p. 117) lists the characteristics of contextually evocative tasks as follows:

a) Elicited controversial or curious reactions from one or more learner due to a sufficiently engaging problem scenario;
b) Provoked dialogue where informal inference tended to be guided by the problem context as opposed to statistical considerations;

---

[8] "*Evidence* is the data that supports the inference. *Evidential reasoning* is the process of arriving at the inference that justifies why the data should be regarded as appropriate evidence in support of the inference and weighing the strength of the inference" (Gil & Ben-Zvi, 2011, p. 92).
[9] "Abduction is a form of reasoning suggesting an explanatory hypothesis. It provides a contextual or theoretical causal explanation for a phenomenon or data" (Gil & Ben-Zvi, 2011, p. 92).

c) Connected to learners' prior knowledge by providing an access point for rich argumentation and dialogue that appeared to assist learners in making conjectures.

Madden's tasks were intended to provide meaningful contextual problems that would invoke informal inference. Her classification of tasks was a retrospective act, based on the experiences of the teachers in her study. Therefore, statistical and contextual understanding was involved in all tasks, as well as understanding of the technological tools at their disposal. Madden did not describe what enabled her participants to shift from "controversial or curious reactions" to statistically more acceptable ways of reasoning. However, Madden's study emphasises a real and meaningful interaction between context, statistics and technological tools. For example, context-based conjectures were suggested when statistical summaries did not convince as evidence for inferences. In the discussion of a statistically provocative task, she reports that, "one can hear context informing teachers' statements and IIR about statistical significance" (Madden, 2011, p. 120). The continual presence of the data-context during the development of statistical concepts may also interfere with the statistical reasoning process. Pfannkuch suggests that teachers may need to "deliberately encourage students to periodically dissociate from the data-context" (Pfannkuch, 2011, p. 44). It seems therefore that teaching that takes up the structuring of a data context is not an easy task, and there is little guidance for teachers to do so adequately.

Mallows (1998) argues as a practicing statistician when he stresses the importance of understanding the factors that influence situations or data-contexts as part of statistical investigation. Mallows (1998, p. 2) comments on the lack of attention given to this aspect of statistics and proposes that understanding the context is the Zeroth problem:

> *Problem (0):* Considering the relevance of the observed data [information], and other data that might be observed, to the substantive problem.

With regard to the teaching and learning of statistics, the imperative to "grasp the system dynamics" of a data-context has since been included in the PPDAC cycle (see Chapter 1). Yet, grasping the system dynamics seems to be but one of the reasoning tasks at the start of statistical investigation, at least for the novice. Garfield (2002, p. 3) summarises the problem as follows: "People may be able to apply statistical rules in one setting (for example, to random generating devices such as dice) but rarely or never to similar problems that involve social content." From the perspective of statistical reasoning in everyday contexts, it seems that statistical model structures are not transferred to social contexts, because people do not relate to their contexts as if they were marbles in a jar (Schwartz & Goldman, 1996). Research about informal reasoning also indicates the existence of a social effect that may facilitate everyday reasoning, but may prohibit understanding of the system dynamics. This research will be reviewed in Chapter 3.

In this study I argue strongly for the inclusion of contextual information and everyday reasoning in teaching statistics. However, I do not deny the importance of reasoning about statistical objects as is emphasized in the model represented in Table 3. Understanding the conceptual meaning of statistical objects like measures of centre and spread, or of ideas like variability and distribution, is vital for teachers who have to take up learner reasoning about the uses of these tools.

2.4 Reasoning about statistical concepts

In this section I will review statistics education research about understanding of variation, distribution and measures of centre and spread. These concepts are fundamental to statistics. In relation to the PPDAC cycle (see Figure 1, Chapter 1), reasoning about statistical concepts is evident in both the interrogation and interpretation cycles. I argue that access to the investigation cycle requires more than intuitive awareness of these statistical concepts. In order to enter the investigative cycle, sources of variability must be consciously ordered in relation to each other, and variable properties of objects must become objective measurements.

2.4.1 Understanding variation

Consideration of variation is posited as one of the fundamental properties of statistical reasoning (Pfannkuch, 2008). Variation causes uncertainty and creates the need for statistical tools for its description and control (Moore, 1990). While people seem to be intuitively aware of variation, the effect, or implication of variation is not generally evident in decision-making and reasoning (Watson, 1998). Reading and Shaugnessy (2004) differentiate between *variability* as an observable property of an entity and *variation* as the measure of variability. This distinction is appropriate when understanding of the context is seen as a valid element of statistical reasoning. Variation is measured and described by fitting mathematical models to measurements (data). In exploratory data analysis, variation is measured by: the range (mathematically the difference between the largest and smallest measurement); the inter-quartile range (mathematically the values that capture the middle 50% of ordered measurements); and the standard deviation (mathematically the square root of the average of the squared deviations of measurements from the mean). Co-variation between two variables that are linearly distributed is measured, for example, by the correlation coefficient and the gradient of the least squares regression line. Deviation from such mathematical models is measured as an indication of the role of unexplained variation. Reading and Shaugnessy (2004, p. 202) report that students are more concerned with finding real explanations (in context) for variation in measurements, than describing the variation. Indeed, the tendency to postulate causes for variation through abductive reasoning is a persisting impulse (Wild & Pfannkuch, 1999) and has recently been acknowledged by statistics education researchers as an aspect of statistical reasoning (Makar, Bakker, & Ben-Zvi, 2011). Students of introductory statistics may gain understanding of unexplained (rather than random) variation through contextual reasoning about the complexity of relationships between concomitant variables. Typical learning tasks in textbooks do not provide opportunities to consider sources of variability, but focus instead on statistical measures of centre, in order to reduce data. Reading and Shaugnessy (2004, p. 204) advise that "educators need to modify learning experiences so that students can comfortably move from identifying variability; to describing, representing and sifting out causes for variation; and finally, to measuring variation." Hence there is a need to make the structuring of the context in terms of possible causes of variability a learning task, and to recognise that meaningful measurement of variation is the endpoint of this process.

Descriptions and causal explanations of variation

Based on a study of understanding isolated random variation embedded in sampling tasks, Reading and Shaugnessy (2004) developed a *description hierarchy* and a *causation hierarchy* from the interview responses of six primary school and six secondary school students. The challenge of describing variation seems to lie in finding an "anchor" from which to measure variation in data, while the challenge of understanding sources of variation, was to break free from imagined involvement in the context.

The *description hierarchy* is based on increasing sophistication in responses related to notions of spread in sampling results. Less sophisticated responses focus either on middle values or on extreme values; while consideration of what happens between extremes, indicates increased sophistication. At the higher end of the hierarchy are students' discussions of deviations of measures from some value that serves as an anchor. The anchor is not necessarily a central value such as the mean or median. For example, one student reasoned that the deviation in the numbers sampled was 3, by stating, "there are three numbers on each side" of an unspecified centre. Another student reasoned about deviations from extreme values: "at least two away from the highest and the lowest" (Reading & Shaughnessy, 2004, p. 216). The most sophisticated reasoning relates variation to deviation from a central value as anchor, expressed as "average … would be around" a certain value (Reading & Shaughnessy, 2004, p. 216).

The *causation hierarchy* is developed from spontaneous discussion of causes in elaboration of descriptions of variation in the sampling tasks. Four levels are proposed, with identification of extraneous causes of variation at the lowest level. These are related to imagined action in the sampling situation. For example, low level reasoning about sampling lollies from a bowl cites the position of the lollies to be sampled as the reason behind why only certain ones were picked; or, the that the size of the person's hand determined the number of objects that were sampled. More sophisticated responses involve discussion of the frequencies of different objects in the sample space as cause of variation in the sample, as well as reasoning about proportion and likelihoods based on proportions (Reading & Shaughnessy, 2004, p. 218).

Reading and Shaugnessy (2004, p. 223) draw the following implications for teaching statistics from their research:

> First, do not be afraid to give students more challenging tasks […] since they allow more opportunity for discovering and attempting to explain the variation that occurs. Second, do not separate the study of central tendency and spread […] Third, when learning situations involve reasoning about variation, allow students to have their untrained explorations into what is happening with extreme and middle values.

Masnick, Klahr and Morris (2007) studied understanding of variation in the context of scientific experiments. In particular, they studied how variation in measurements influenced prior beliefs about factors that influence a phenomenon. They report that Grade 2 to 4 learners were not surprised by variation, but struggled to use variable measures explicitly when drawing conclusions. The learners were able to generate many potential causes for the observed variation when the data-context was familiar to them. Comparing reasoning in familiar and unfamiliar data-contexts, the researchers found that the learners had difficulty adjusting their prior beliefs if the data conflicted with their contextual beliefs, while adults were better at reviewing their beliefs in such situations. Masnick et al.'s study required participants to execute a cycle of investigation by designing experiments, taking measurements, and interpreting the measurements. A permutation of their research required reasoning about data sets that were presented to the participants; in other words, the participants did not reason about the context in order to design an experiment that yielded the data. In this situation, the researchers report that almost half the participants still developed theoretical explanations to explain the variation in the data. The participants' theoretical explanations supported their reasoning about the statistical properties of the data set. These results suggest that both data (measurements) and theory (explanations), that is, understanding of the structural or causal relationships between variables in a context, are important in "designating what constitutes signal as well as noise"[10] (Masnick, et al., 2007, p. 25).

Sánches, Borim da Silva and Coutinho (2011) summarised major studies related to students' and teachers' understanding of variation. Of importance to my study is research about understanding of variation in data contexts, rather than chance contexts, such as sampling experiments. With regard to students' understanding of variation in data contexts, they report that these studies focus on the language of variability; perception of variability in comparing

[10] 'Signal and noise' is a metaphor for the contrasts that can be used for understanding variation. Signal represents a central value like the mean of a data set, and is interpreted as a "true value" when properties of physical objects are measured in experiments, while noise represents the variation of measurements in relation to the central value. The metaphor was coined by Konold and Pollatsek (2004).

groups; the role of technological tools;[11] and understanding standard deviation. I will briefly report on research related to each of these aspects.

The meaning of variation in data contexts

Watson (2006) constructed a hierarchy based on Grade 7 and 9 students' responses to the questions: *What does variation mean*?; *Use the word variation in a sentence*; and *Give an example of something that varies.* According to the hierarchy, at the most sophisticated level, students should relate the idea of *change in states* to the concept of variation, for example, "Something that differs from its previous state"; "There is a big variation in the results"; and "The weather varies" (Watson, 2006, p. 220). Surprisingly, less than 10% of the students in the sample were able to respond at this level. The idea of *change in states* is important for measuring variation as the difference between the initial and final state, but it is not intuitive to reason about *differences* between many observed states as *changes* in state if the context is not explicitly temporal. To use an earlier example, the varying incidence of AIDS in different countries does not intuitively suggest a change in the incidence of AIDS. This may be a specific instance where restructuring of contextual understanding is required before measurement of variability becomes meaningful.

Da Silva and Coutinho (2008) mapped the understanding of variation of nine in-service teachers to a SOLO based model proposed by Garfield (2002) which provide levels of increasingly sophisticated understanding of variation. In their study, the teachers gathered data about the ages of people they interviewed. At the lowest level, that of *idiosyncratic reasoning*, the teachers gave answers that related the mean to the context, but did not include any mention of variation. At one level higher, the teachers exhibited *verbal reasoning*, indicating that they had some understanding of the application of standard deviation; such as indicating irregularity in the data. Da Silva and Coutinho (2008) classified the observation that all the values of the data set are not equal to the mean, as verbal reasoning, along with teachers' attempts at making sense of the standard deviation based on textbook information. A still higher level of reasoning, *transitional reasoning*, was evident when the teachers

---

[11] I will not review research about the use of technology, since in this study, technology played a minimal role. This is due to lack of access to technology in South African schools and previously disadvantaged communities. The bulk of my teacher students are from such communities and schools.

described variability – not only in relation to a measure of centre but also taking some other aspect of the distribution into consideration, such as extreme values or the graphical shape of the distribution. However, the authors report that teachers who reasoned transitionally showed no need to measure variation despite awareness of contextual variability. More advanced *procedural reasoning* about variation was evident when teachers reasoned with the mean and deviations from the mean, and started to develop understanding of calculating an interval around the mean. *Integrated process reasoning* is at the apex of this hierarchy, indicating that "the student has complete understanding […] and can explain the process in her or his own words with confidence" (Garfield, 2002, p. 8). Despite having gathered their own data in a familiar context, Da Silva and Coutinho (2008) report that none of the teachers in their study used integrated process reasoning about variation, while verbal reasoning (viz. using the terminology to define the concept without understanding the concept) was predominant. The researchers conclude: "being limited to the [sic] verbal reasoning about variation does not allow mathematics teachers to teach their students the meaning of measures such as standard deviation, restricting them to the teaching of algorithms" (Da Silva & Coutinho, 2008).

Research reports such as Da Silva and Coutinho's (2008) emphasise the need for research about understanding of variation in everyday contexts. Their research indicates that attempts to make sense of formal statistical measures are not integrated with an informal understanding of these measures in context.

Reasoning about variation in informal language

Bakker (2004b) investigated Grade 8 and 9 students' reasoning about variation through the consideration of shapes of graphic distributions. Bakker required the students to draw graphs of their data (with computer tools) and to reflect on the results of experimenting with the graph and the data. Bakker argues that diagrammatic reasoning enables the abstraction of concepts underlying variation like "the majority", "the average", and "the shape" as statistical objects (Bakker, 2004b, p.80). At the early stages of a task in which variability became evident through growing samples, the students used terms like *close together*, *further apart* and *spread out,* to indicate differences in spread between samples of different sizes. They used terms like *average, range* and *spread* in statistically unconventional ways, and included non-statistical terms such as *majority*, *semi-circle* and *pyramid* in their expressions of shapes of distributions. Contextual reasoning supported the students' diagrammatic reasoning. Without access to data, the students had to explain which shape of graph would *not* represent the general distribution of people's weights, based initially on their general knowledge. The everyday contextual examples that they were able to evoke of the weight of specific (though imagined) people, enabled them to choose the appropriately-shaped graphical representation.

With regard to teachers' understanding of variation in data contexts, Makar and Confrey (2005) studied teachers' variation-talk, that is, the informal terminology the teachers used to relate variation and distribution. They report that the teachers' talk was similar to that of students in other studies, and included unconventional terms like *clustered*, *spread out*, and *modal clumps*. While the use of standard statistical terms increased from the first to the second interview in their study, only two of the 17 pre-service teachers used *standard deviation* as a term to discuss the concept of spread – the majority preferring informal descriptions of variation.

Variability in comparing groups

This topic refers specifically to research about the comparison of sets of data that are represented graphically. In typical tasks of this nature, students find it difficult to compare groups of different sizes, due to an apparent lack of proportional reasoning (Watson & Moritz, 1999). Other difficulties in comparing data sets are related to the inability to understand a data set as a whole or an aggregate, with features such as shape, centre and spread (Konold & Higgins, 2003). Comparison of groups often elicit local, rather than global views of the data (Ben-Zvi & Arcavi, 2001), as is evident from the comparison of specific case values between two groups (e.g. the best performances in two groups) or the frequency of cases in different groups that correspond to a chosen measurement (e.g. the number of people taller than six feet in two groups). This may be indicative of the influence of contextual comparison of variability where the reason for the comparison gives meaning to the procedure.

Understanding deviation

Lehrer and Schauble (2002) conducted research with fourth and fifth grade students who gathered and analysed data through measurement tasks. These tasks were aimed at helping the students to understand variation in relation to measurement error. In situations where the students believed that the measured object (e.g. a flagpole or a pencil) had a true length, they related a central value in the distribution of measurements, typically the median, to the true length. Error as variation came to be understood as a relation between the data and their distance from a centre, such as the median. Distributions with less variation were taken as indicative of more precise measuring processes. An important shift from describing variation to measuring and describing deviation was achieved when the same students compared the accuracy of different measurers who all measured the length of the same pencil. The students calculated the differences from the median of each measure and created a distribution of the deviation measures. Lehrer and Schauble (2002) report that the learners had difficulty to accept the concept of a typical or representative value in the context of growing plants, where a 'true measure' was not expected. Although Lehrer and Schauble do not comment further on this contextual difference, it is plausible to suggest that the concept of variation as a change of states may have conflicted with the static differences between the observed heights of the plants. In a more recent study Lehrer and Kim (2009) pursued the measurement of variation through deviation with 11–12 year olds. The students invented measurements of variation of distributions by calculating and totalling differences between pairs of data, or data above and below the median or the mode. These invented deviation strategies may serve as informal forerunners of measures of variation such as the interquartile range or the standard deviation.

The concepts of variation, distribution, mean, and deviation from the mean must be integrated in order to develop an understanding of the standard deviation. Studies about the formal understanding of the standard deviation were mostly done with high school and university level participants. A study by DelMas and Liu (2005) explored understanding of standard deviation among college students in a computer environment, using context-free problems. The students in their study varied in the level of integration of the supporting concepts, and specifically in relating the standard deviation conceptually to the mean as a balance point. Similarly, Mathews and Clark (2007) found that college students who performed well in an introductory statistics course that dealt with standard deviation expressed limited understanding of the concept of standard deviation in relation to the mean and other features of the distribution.

The close interplay between context and data in statistical reasoning, evident from the discussion on understanding variation, is further supported by research on the concept of data as distributions with shapes.

2.4.2 Understanding distribution

If variability is the reason for the existence of statistics, distribution is the signature of variation. A data set does not necessarily represent to novice statistics students an aggregate unit with statistical properties. Instead, novice students "tend to conceive a data set as a collection of individual values" (Bakker & Gravemeijer, 2004, p. 147). Lacking the concept of data as a distribution, students who reason locally rather than globally about a set of data refer to individual cases, rather than the distribution as a whole (Ben-Zvi & Arcavi, 2001). Additional abstraction from the context is needed to conceive of data as a distribution. Bakker and Gravemeijer (2004, p. 148) explain:

> An underlying problem is that middle grade-students generally do not see "five feet" as a value of the variable "height", but as a personal characteristic of, say, Katie. In addition to this view, students should learn to disconnect the measurement value from the object or person measured, and to consider data against a background of possible measurement values.

Bakker and Gravemeijer (2004) conducted design research with seventh grade students in a series of lessons supported by the use of data-handling software. Emphasising the complex nature of the concept of distribution, Bakker and Gravemeijer indicate that developing an understanding of distribution entails an upward perspective, as well as a downward perspective. An upward perspective starts with data as individual values, subsequently abstracting properties of data sets, to finally viewing the data as an aggregated distribution. A downward perspective views individual values in relation to the distribution. Bakker and Gravemeijer maintain that expert statisticians easily invoke the appropriate perspective, either upward or downward. In particular, the researchers report that thinking about everyday judgements like "good but not reliable" in the absence of data "create[s] the need for a conceptual unity that helps in imagining a collection of data with a certain property. The notion of distribution serves that purpose" (Bakker & Gravemeijer, 2004, p. 155). Bakker and Gravemeijer describe a developmental path for learning to understand distribution. One step away from reasoning with data as individual points, participants in their design research divided unimodal distributions into three groups. The groups correspond to the everyday comparative notions of *low*, *average* and *high* values. One more step away, participants thought of "small, average, tall *and* in between", while average was conceived of as a majority group in the middle, rather than an individual value (Bakker & Gravemeijer, 2004, p. 159). Finally, when participants started to reason with shape properties of distributions, such as 'bumps', the researchers judged that they had constructed the concept of distribution. An implication for teaching from the research reported here, is that meaningful and conceptual learning about distribution is time consuming, and places exceptional demands on teachers to learn to "orchestrate" productive class discussions (Bakker & Gravemeijer, 2004, p. 167).

2.4.3 Understanding the statistical mean

The mean as a measure of the centre of a distribution is a key concept in statistics. Averaging in the sense of calculating a *mean* pervades the structure of more complicated statistical models. So, for example, the standard deviation is in essence a mean average. However, the everyday notion of average is not the same as the statistical concept *mean*. Average is an everyday concept that relates to the statistical measures of mean, median, as well as mode. In its everyday use, the goal that a person has in mind assigns further meaning to average, meanings such as fair share; typical value; signal in noise; or data reduction in order to reduce complexity (Konold & Pollatzek, 2004). In all these meanings, the mean is embedded as a summary value. When we think of average as a fair share, Konold and Pollatsek argue, we do not think of the computed value in relation to the actual data values, nor are we aware of how the actual values are distributed – the individual values are completely subsumed by the mean value in our reasoning. Indeed, a pervasive finding in statistics education research is that novice participants routinely describe variable measures with an average value, without mentioning how the actual data values are spread in relation to the average value.

Average as a typical value, rather than a fair share, is more closely related to the notion of distribution. Konold and Pollatsek (2004) report that younger students favour the mode as a typical value, since it satisfies the intuitive notions of typical as located in the middle, as well as typical as being the most frequent value. Makar and McPhee (2009) studied young students' conceptions of typical-as-average when they were solving an ill-structured problem, and report that the students considered typical as a *reasonable range* of values, or as the most common data in the class. Studying a dataset, these young students debated whether most common would be one number, or an interval of values.

The notion of 'signal in noise' is evoked when measurement deviations are perceived as errors and a true value is hidden in each measurement. The average is then perceived as closest to the true value. The idea of looking for a *true value* actually problematises the nature of the mean as a measurement. Children in primary school often think that the mean or the median must necessarily be chosen from one of the actual measurements (Mokros & Russell, 1995; Strauss & Bichler, 1988). Similarly, when pre-service elementary teachers provided free-response definitions for mean, median and mode, they made explicit that the median need not be the same value as an actual data point; but they failed to state this for the mean as well (Groth & Bergner, 2006). Bakker (2004a) reports that, in contrast, students in his study chose non-actual measurements to represent a dataset. Bakker ascribes this to estimating the mean from a graph of measurement values, rather than working with measurements given as numbers.

Measures of centre such as the mean are useful to compare data sets, but various researchers report that students do not use the mean for this purpose – they rather revert to comparing clumps of data or frequencies. An example of frequency comparison is provided by Konold and Pollatsek, stating that "… to decide if males were taller than females, they might inspect the sample for all individuals who were 6 feet tall, and argue that males were taller because there were more males than females of that height" (2004a, p. 181). This "slicing technique" according to some contextually meaningful value seems widespread, and does not take into account the overall number of cases in the groups that are compared. Neither does it indicate that a typical value, a true value, or a fair share value, has common sense meaning in such comparison problems. At most, slicing is a way to reduce complexity, while frequency holds the key to comparison. Of importance for my study is the propensity to choose a contextually salient value as a point of comparison. For example, six feet is seen as *tall*. Tall seems to be qualitatively defined on an experiential scale, rather than an abstracted measurement scale related to the data set.

Although most people know how to calculate the mean, Mokros and Russell claim that “the mathematical relationship itself remains opaque” (1995, p. 22). Makar (2004) argues that the statistical relationship between a distribution as an object and the mean as a measure of the object is also opaque. Based on in-depth interviews where children from Grade 4 to Grade 8 had to construct and interpret means and weighted means in contextual problems, Mokros and Russell (1995) classify understanding of the mean as either non-representative or representative of a distribution of data. Non-representative understanding is evident from viewing average as “mode” and using the algorithm with limited strategies to determine the reasonableness of solutions. Understanding of the representativeness of the mean was evident from strategies where average was used as a value that would be reasonable (where a student would use everyday experiences of the context to judge the reasonability); strategies where the average was taken as a midpoint; and strategies using average as “mathematical point of balance”. A much larger scale study was conducted by Watson and Moritz (2000) with regard to the development of the concept of average. They administered four survey items to 2250 students from Grades 3 to 11. The items were designed to access everyday understandings of average and to assess application and calculation of the mean and the median in straightforward settings. In their analysis, Watson and Moritz employed the Structure of Observed Learning Outcomes (SOLO) (Biggs & Collis, 1982) to classify student responses hierarchically. The authors report that response levels increase with grade level:

a) *Prestructural responses* did not include any clear concept of average. For example, when asked, “If someone said you were average, what would it mean?” a Grade 5 student replied that it meant, “you are not the best friend for me.”
b) *Unistructural responses* consisted of a single relevant aspect from the domain of the task set. Another Grade 5 student in answering the aforementioned question said that it meant, “that you were okay.”
c) *Multi-structural responses* included two or more aspects of the domain of the task, usually presented in sequence. For example, a Grade 5 student responded to the question with “It is if you have some numbers, you add them together and then divide by how many there are,” reflecting the algorithm for

the arithmetic mean. A Grade 9 student responded, “Not really good and not really bad; in between,” reflecting the idea inherent in the median.

d) *Relational responses* exhibited an integrated understanding of the relations involved in the task set. Students responding at this level, for example, knew that the median was the middle value of a data set and could find it after ordering the values given. When asked why the median might be used to report house prices rather than the mean, a Grade 9 student appreciated the merits of each for representation in the context by saying, “because it shows a fair representation of the prices. If the average was used, a particularly cheap or expensive house would muck up the fair representation” (Watson & Moritz, 2000, p. 15).

The authors also asked students what the meaning of average is in the phrase: “the average wage-earner finally can afford to buy the average home”, and conclude that “[the] results indicate that, in these contexts, a majority of students conceptualized average in terms of ‘middle,’ with the idea of ‘the most frequent’ used by fewer students and the algorithmic idea of ‘mean’ seldom employed” (Watson & Moritz, 2000, p. 15).

The same researchers conducted two follow-up studies of 134 Tasmanian and South Australian students’ understanding of the mean, conducting in-depth video-taped interviews with the students. The first follow-up study was conducted with 22 of the original South Australian students, three years after their first participation when they were in Grades 3, 5, 7 and 9. The second follow-up study, involving 21 of the original Tasmanian students was done four years after their first participation in Grades 3, 6 and 9. The interview protocol for average consisted of the four questions. The first question probed free responses to the meaning of average and explanations of the word average in the statement “Australian primary school students watch an average of 3 hours of TV per day” (Watson & Moritz, 2000, p.19). The second question asked the participants to explain how the average hours of watching television could have been obtained. The third question queried the meaning of the numerical value of the average as a number with a decimal fraction part in the statement “On average, Australian families have 2.3 children” and followed up with a question about the missing value that would yield the average of 2.3 children. The last question concerned weighted averages. The researchers interpreted the students’ verbal expressions of the meaning of average in relation to the mean, median and mode. For example, “a response that average means ‘the same as most others’ is associated with the mode, and “not good, not bad, but in between” is associated with the median” (Watson & Moritz, 2000, p.20). They categorised the responses on six levels, according to the SOLO model. The characteristics of the levels are summarised in Figure 3. It is significant that colloquial ideas around the meaning of average as either *middle* or *most* persist through all the levels.

Characteristics of Six Levels of Students' Understanding for the Concept of Average
Preaverage (P)

Use no term for average, even in a colloquial sense
Tell imaginative stories about the context
Often are not asked complex questions by their interviewer

Single Colloquial Usage for Average (U)

Often use colloquial terms for average, such as 'normal' or 'okay'
Often use imaginative ideas related to the context to support response
Sometimes refer to "add up" colloquially but not in a calculation sense
Make no progress on complex questions

Multiple Structures for Average (M)

Use at least one, often two or three, [colloquial] ideas—including most, middle, and the add-and-divide algorithm for the mean—to describe average in straightforward situations
Rarely use more than one of these ideas in complex questions and make little progress toward solutions
Sometimes acknowledge conflict between incorrect calculations of mean and idea of mode

Representation With Average (R)

Refer to add-and-divide algorithm for the mean to describe average in straightforward situations, often also with ideas of most or middle
Often realize association of decimal form with algorithm for mean
Often express some idea related to the representative nature of average (e.g., prediction, estimation, or representing whole data set)
Often refer to most to describe data distributions compatible with mean or offer mode as alternative average concept
Know the mean but do not successfully apply it in complex contexts; make partial progress on more complex contexts with prompting, for example, have sense of weighted mean not being exactly in the middle, but lack precision
Often use visual features in preference to the mean to compare data sets presented in graphs, although may use mean when prompted

Application of Average in One Complex Task (A1)

Refer to add-and-divide algorithm for the mean to describe average in straightforward situations, often also with ideas of most or middle
Often realize association of decimal form with algorithm for mean
Often express some idea related to the representative nature of average
Apply understanding of the mean to determine total (Part 3), or apply weighted mean algorithm directly (Part 4), but not both
Rarely refer to most to describe data distributions compatible with mean

Application of Average in Two Complex Tasks (A2)

Refer to add-and-divide algorithm for the mean to describe average in straightforward situations, often also with ideas of most or middle
Often realize association of decimal form with algorithm for mean
Often express some idea related to the representative nature of average
Apply understanding of the mean to determine total (Part 3)
Solve weighted mean problem by calculation (Part 4); when inappropriate calculations of weighted mean yield an unusual result, revert to proportional reasoning to solve problem
Often refer to most to describe data distributions compatible with mean or offer mode as alternative average concept
Often use the mean to compare data sets presented in graphs

Figure 3: Characteristics of six levels of students' understanding for the concept of average (Watson & Moritz, 2000, p. 22)

Across the age groups the trend to understand average as *middle* or *most* persists as well. Often students calculated the mean when asked what the average is, and added a descriptive phrase related to the context, which involves *most* or *middle*. Watson and Moritz describe this trend as indicative of "representative understanding", that is, an indication that the mean is seen as representative of the data set (Watson & Moritz, 2000, p. 46). They relate comments from students involving *most* to the mode concept, and those referring to *the middle*, to the median concept. The researchers report that comments related to balancing (which would be related to the mean concept) hardly ever occurred, with comments about the add-and-divide algorithm as the only elaboration of the mean. However, the researchers hail the variety of ideas associated with average as good news for teachers, as they provide a foundation for classroom discussions. Findings from their study indicate that these eclectic ideas do not include balancing and evening out, and the researchers provide no further ideas of how to map these sophisticated understandings of the mean onto everyday concepts for the benefit of teachers.

Summarising students' responses to the question "What is the average?", Bakker reports that 13 out of the 26 students in his study mentioned the mean algorithm or part of it, and ten mentioned other aspects such as 'most', 'about', 'roughly', 'in between', 'a bit in balance', and 'midpoint'. In the context of other interview questions, the students also used expressions such as "large amount" or "half is below, half is above" when looking for the average (Bakker, 2004b, p. 97). These open-ended descriptions of average yielded the same kind of interpretations as recorded by Watson and Moritz (2000). Bakker (2004b, p. 96) indicates that the responses from his students must be interpreted against the information "that the Dutch word for average, *gemiddelde*, refers both to the informal meaning of average and the statistical meaning of arithmetic mean, but it does not function as a collective noun for mean, median, and mode, as the term average sometimes does in English."

The various interpretations of the mean reported in research with school and college students, and teachers, reveal a practical understanding of the concept *mean* as a contextual average. Numerical calculations that require manipulation of the mean algorithm, such as finding missing values or dealing with weighted averages, seem contextually meaningful and attainable at school level. In contrast more abstract meanings of the mean, for example that of a balance point, seem to be difficult to develop. Evening out strategies[12] to estimate the mean have potential to relate the concepts of the mean as a fair share and the mean as a balance point. Bakker (2004a) identified evening out as an intuitive strategy when learners estimate the mean on value bar graphs. There are no studies that formalise evening out strategies to relate them to the mean algorithm.

It seems to me the mean algorithm is accepted in statistics education without problematising its syntax. Five theoretical conceptualisations of the mean and recommendations of how they follow conceptually on from one another are proposed by Waiter, Lamontagne and Chartier (2011). The first two conceptualisations are presented as metaphors and the third as a mathematical motivation for the mean.[13]

a) The *Socialist* conceptualization: For the purpose of introducing the concept (to social science students), the mean is assigned human qualities in order to formulate meaning. This conceptualization is, in essence, that of assigning equal value to each case to achieve a sense that the total is shared by all – hence the metaphorical use of the term “socialist”;
b) The *Fulcrum* conceptualization: This engineering metaphor highlights the property that “the mean is equivalent to the balancing point of deviations in a distribution”; and
c) The *Algebraic form of the Least Squares* conceptualization: once students have experience with the fulcrum conceptualization, and with mean deviations, the algebraic proof that the mean is the value that minimizes the sum of the squared deviations provides a central value that is representative of the data set (Watier, et al., 2011, p. 3).

I find no reason in the first two conceptualisations for *why* the mean must be considered as a balance point or an equal-for-all value. The algebraic conception also does not indicate the ontogeny of the mean, it only proves a property, namely that the mean minimizes the sum of

[12] Evening out strategies refer to informal methods to allocate parts of differences between data points to the actual data points, so that all the data points end up with equal values.
[13] The fourth and fifth conceptualisations, which are not discussed here, are: The Geometric Form of the Least Squares Conceptualization: “From this point of view, the mean is equivalent to the coordinates of a new origin in n-dimensional space that minimizes the length of the hypotenuse formed from combining vectors of independent observations” (Watier, et al., 2011, p. 7); The Vector Conceptualization, based on analytic geometry and generally too advanced for a general undergraduate course in introductory statistics.

squared deviations from itself. To the uninitiated this sounds like circular reasoning. Finally, without motivating the conceptual trajectory to link one conceptualisation to the other, Watier et al.'s claim that the conceptualisations are conceptually organised, does not convince me. In Chapter 9 I will take up an alternative algebraic argument, not as a proof but as a way to define the statistical mean as an object rather than a procedure.

2.5 Summary

The role of context in statistical reasoning is acknowledged as a key aspect of informal inferential reasoning. The effect of context is seen as problematic in the development of statistical reasoning, and Pfannkuch (2011) indicates that deliberate back-grounding of the data-context may be periodically necessary when learning statistical concepts. I argued that such back-grounding may be achieved through conscious structuring of a data-context in terms of concomitant factors. Pfannkuch confirms the paucity of research about the relationship between context and reasoning:

> Finding out about the data-context typically takes a great deal of time and work, thus students must first understand the problem, gradually abstract out elements of the real situation, then simultaneously maintain a contextual and abstract view of the data, and finally produce findings that make sense with the real-world situation…There appears to be little research, however, specifically focused on the role contextual knowledge plays in learners' reasoning" (Pfannkuch, 2011, p. 31).

My study provides information about the process of gradually abstracting out elements of the real situation, as Pfannkuch indicates above.

The bulk of research about statistical reasoning at school level focuses on reasoning about specific concepts like variation, distribution, mean and sampling and the "Dimensions of statistical Inquiry" model (Figure 1, Chapter 1) has been developed to integrate these with cycles of statistical reasoning. SOLO-based models of concept development is generally used by statistics education researchers to describe hierarchic levels of reasoning ranging from idiosyncratic to integrated process reasoning. I argued that the lower levels, which describe reasoning as idiosyncratic or verbal, prestructural or unistructural, do not provide information about why participants reason in these ways. A closer look at the short transcripts that are used to exemplify lower level reasoning reveals a concern with personal contextual issues, both when statistical findings have to be interpreted in context and when participants give free responses about their understanding of concepts like the mean. While there is consensus that understanding the data-context is necessary for statistical reasoning, the relationship between context and reasoning remains underresearched. Research shows that measurements are not always abstracted as numbers, and instead remain properties of the data objects and questions are often aimed at qualitative inquiry about specific cases. I argue that the inclusion of "grasping the system-dynamics" as one of the facets of the PPDAC cycle implies that data-contexts must be consciously structured by learners as a reasoning task. The research reviewed provides no coherent information about ways in which teachers can support learners' reasoning in order to abstract relevant aspects from the complexity of a data-context. Instead many teachers' reasoning about data in contexts and about statistical concepts are placed at similar low levels as that of school level learners.

Studies with younger children and teachers alike show that informal language to describe variation hold potential for concept development. I reviewed studies about informal reasoning in informal language about variability and variation, deviation and distribution. These concepts are integrated and at the heart of statistical reasoning. While participants in the various studies are generally aware of variability in the data-context, they are not reasoning about measures of the variability, but tend to reason with contextually salient data values, like 'the best' value. Such word use indicates qualitative, rather than quantitative, statistical comparison. Reasoning about distribution is shown to develop from reasoning about such individual points in a data set, through division of the collection of data points typically in three groups – representing small, medium, and large – until 'distribution' is understood as a shape property of a set of data represented on a measurement line. Meaningful distributional reasoning coordinates central values and measures of the spread of the data. Lehrer and his colleagues (Lehrer & Kim, 2009; Lehrer, Kim, & Schauble, 2007; Lehrer & Schauble, 2002) showed that primary school learners can develop their own measures of variation in contexts of repeated measurement. Finally, I reviewed research about understanding of the mean as a measure of centre of a data set. The semantics of the mean is the focus of the body of research with no studies about participants' understanding of the syntax of the mean. The term mean is usually replaced by the term average by researchers as well as participants, and interpreted as 'in between'. Average is usually related to the median and the mode, despite participants being able to calculate the mean correctly. Six SOLO-based levels of understanding the mean were described, ranging from no awareness of typicality; through colloquial understanding and calculation; to integrated understanding of the mean as a typical value that can be used to represent a data set. The influence of the data-context on reasoning about the mean as a representative value emerged in the research of Lehrer and Schauble (2002), since the participants did not invent measures of centre in a context where a true measurement was not expected. In my study I will report on the understanding of the semantics as well as the syntax of the mean that emerged in a complex discussion where average was not endorsed as an adequate concept of the statistical mean.

The importance of considering the implications of everyday reasoning patterns for statistical reasoning is supported by research about explanations in informal inferential reasoning (Gil & Ben-Zvi, 2011). In the next chapter, I review research about informal reasoning and informal statistical reasoning in out-of-school settings. My goal is to understand how people perceive of typicality and variation when they are not learning statistics and how they reason colloquially to solve problems where there is uncertainty in the context.

Chapter 3: Literature review: Colloquial reasoning, and informal statistical reasoning

3.1 Introduction

In this study, I take the point of view that the arguments that the students in my study construct, as well as their analysis and evaluation of each other's arguments in the statistics class, will be influenced by practical life experience, through differing degrees of familiarity with the contexts of the data, as well as by their experience of mathematical reasoning. Statistical reasoning is different from mathematical reasoning in that it is concerned with describing uncertainty and variation, rather than the ideal objects of mathematics. The data-context provides the uncertainty and variation and while contextual information temporarily recedes to the background during the process of statistical problem-solving, it cannot be stripped away as in mathematics (Cobb & Moore, 1997). At the early stages of statistical investigation context knowledge is the primary driving force and reasoning to abstract aspects of the context is needed to create opportunities for the use of statistical knowledge (Wild and Pfannkuch, 1999). Context knowledge develops from contextual reasoning, therefore, from a discursive perspective, the way we talk about statistical situations in everyday discourse influences the way we talk statistics in learning and teaching situations.

In this chapter I review research about informal reasoning. I differentiate between informal statistical reasoning, and colloquial reasoning about non-statistical problems. I show that colloquial reasoning, as well as informal statistical reasoning, is influenced by the contexts of the problems in patterned ways. This literature review is important in order to understand the strengths and limitations of colloquial reasoning as a base from which to support the development of statistical reasoning (O'Toole, 2006). Currently, such informal discourse in statistics classrooms is described as idiosyncratic reasoning without explanation why such reasoning is pervasive in introductory courses.

Reasoning, argumentation and thinking have many different definitions, depending on researchers' particular interests in the purposes and the forms of these activities. I will use the term reasoning as an umbrella, to include all three. I view reasoning as a communicational activity. Whether we reason with ourselves or with others, we are communicating beliefs about relationships between claims and the evidence on which we base our claims. A review of the literature of informal reasoning is in itself an argument with claims and evidence.

Reasoning takes different forms, most notably deductive and inductive forms. Deductive reasoning takes general premises to be true, and makes claims about particular instances. For example, if $P \rightarrow Q$ is a true statement and P is true, we deduce that Q is true. Formal deductive reasoning rules prescribe how to deal with inversions, negations and the conditionals of claims. Inductive reasoning cannot claim that a rule or its premise is universally true, but it compares and collects information about many particular instances to make general, truth-like claims. If $P_1, P_2, \ldots, P_n \rightarrow Q$, we infer that $P_{n+1} \rightarrow Q$. Depending on the evidence collected, an inductive claim can be made with variable levels of confidence, excluding absolute confidence. Klauer and Phye (2008) distinguish between inductive reasoning and inductive inference, in order to facilitate an operational definition of inductive reasoning. The difference lies in the end-goals of reasoning and inferring: inductive reasoning is aimed at detecting generalisations, rules, or regularities in order to construct types or categories; while inductive inference extends generalisations to the totality of objects or relations that belong to a category. In essence, inductive reasoning is about comparison, and statistical reasoning is inductive. Our observations and contextual experiences serve as samples from which we reason and make inductive inferences. Formal statistical reasoning is inductive since we reason from a random sample of observations back to the population of possible observations. Klauer and Phye provide the following schematic definition of inductive reasoning in Figure 4:

Inductive reasoning consists of detecting regularities and irregularities by finding out

$$\overset{A}{\left\{\begin{array}{l} a_1 \text{ similarity} \\ a_2 \text{ difference} \\ a_3 \text{ similarity} \\ \text{\& difference} \end{array}\right\}} \text{ of } \overset{B}{\left\{\begin{array}{l} b_1 \text{ attributes} \\ \\ b_2 \text{ relations} \end{array}\right\}} \text{ with } \overset{C}{\left\{\begin{array}{l} c_1 \text{ verbal} \\ c_2 \text{ pictorial} \\ c_3 \text{ geometrical} \\ c_4 \text{ numerical} \\ c_5 \text{ other} \end{array}\right\}} \text{ material}$$

Figure 4: Definition of inductive reasoning (Klauer & Phye, 2008, p. 87)

Klauer and Phye's definition of inductive reasoning reads as if similarities or differences and attributes and relationships exist "out there". Yet, it is not my experience that the material of our statistical reasoning is out there, nor that we *find out* similarities or differences of attributes and relations. I claim that we construct the materials, comparisons and categories discursively, and that the categories we construct and compare are variable and diffuse. Categories shift and change as we formulate different questions about the same materials. Nor is it my experience that statistical inference is theoretically separable from statistical reasoning in the same way as Klauer and Phye distinction between inductive reasoning and inference draws attention to the reasoning involved in establishing contextual categories that serve as reference classes for statistical comparison. In order to make statistical inferences we need to be cognisant of suitable reference classes, but in order to construct reference classes we make inferences about attributes and relationships between potential category members. In the review that follows, I will not distinguish between reasoning and inference, but I will discuss how the similarities and differences we construct between attributes and relations in different everyday situations influence informal reasoning in statistical contexts.

3.2 Aspects of everyday contexts that influence informal statistical reasoning

The question of how inductive inference can be logically justified at all, whether from small numbers or large numbers of observations, has plagued philosophers since the sixteenth century. Mill (1843/1974) explained the problem as follows: "Why is a single instance, in some cases, sufficient for a complete induction, while in others myriads of concurring instances, without a single exception known or presumed, go such a little way towards establishing a universal proposition?” (Nisbett, Krantz, & Kunda, 1993, p. 28). This is also the problem of informal statistical reasoning. On the one hand statistical statements are often interpreted as universal propositions by statistical novices and translated into over-generalisations. On the other hand personal experience is known to carry more weight than statistics in everyday reasoning. The statistical investigation process is characterised as shuttling between the context sphere and the statistical sphere (Wild & Pfannkuch, 1999), and may involve a to-and-fro between inductive and abductive reasoning, or the making of hypotheses (Gil & Ben-Zvi, 2011). We may abductively proffer a plausible relation between a claim and information in order to formulate an investigative question, and then set out to reason inductively to support or refute the proffered rules, generalisations or regularities. We gather data about attributes or relations of a collection of individual cases to use as evidence in probability statements about similar objects or situations. The formal tools of statistics are mathematically derived and allow some measure of deductive certainty in our analysis, but in the end our data based inferences remain inductive. According to Nisbett (1993) any inductive reasoning, however informal, must satisfy some statistical principles to be correct. These principles determine the confidence with which we make generalisations, predictions, and establish causal relationships in our everyday communication. Nisbett, Krantz and Kunda (1993) indicate the most important statistical principle as the coordination of awareness of variability and the law of large numbers. They explain:

> Concepts should be discerned and applied with more confidence when they apply to a narrow range of clearly defined objects than when they apply to a broad range of diverse and loosely defined objects that can be confused with objects to which the concepts do not apply. Generalisations should be more confident when they are based on a large number of instances, when the instances are an unbiased sample, and when the instances in question concern events with low variability rather than high variability. Predictions should be more confident when there is high correlation between the dimensions for which information is available and the dimensions about which the prediction is made, and failing such a correlation, predictions should rely on base rate or prior distribution for the events to be predicted (Nisbett, et al., 1993, p. 15).

An abundance of research, mostly by psychologists (Kahneman, Slovic, & Tversky, 1982) exists on incorrect statistical reasoning, showing that misconceptions persist even among professionals (mostly other than professional statisticians). Even people who know statistical methods often fail to base contextual decisions that involve statistical information on the need for many observations (Garfield, 2002). Inevitably, such incorrect inductive reasoning is the result of violating any one of three statistical principles described by Nisbett, Krantz and Kunda (1993), namely (a) clarity of the sample space or reference class; (b) awareness of variability; and (c) recognition of sources of variation. Based on these principles for appropriate reasoning in statistical contexts, I propose the following working definition:

> Informal statistical reasoning is informal reasoning that takes into consideration the causal as well as random components of contexts; and the constitution of a suitable contextual aggregate (reference class) for comparison and induction; and bases inferences on a large numbers of observations.

I will now elaborate each principle in turn, with some examples from my own observations.

3.2.1 Clarity of reference classes

The term *sample space* belongs to the formal discourse of inferential statistics. In descriptive statistics, an equivalent concept is that of a reference class – a contextual aggregate of cases that have some variable attribute to compare. In order to be included in the contextual aggregate, these cases have to be similar in relevant attributes other than the target attribute, for our inferences to have strength. Informally, we understand that we have to compare apples to apples. In the context of randomising devices like coins, dice or spinners, the sample space for a single trial is clearly delineated (e.g. head or tails for a coin) and repeatability is easily imagined. In such situations it is easy to see what knowledge is relevant (e.g. it does not matter what denomination of coin we use, but it does matter how the toss is executed). However in everyday observational contexts, the reference class is obscured by the complexity of contributing factors and it is often hard to imagine sampling as repeated trials. Thinking of repetition by placing the same person in different situations, or by placing different people in the same situation, does not yield equivalent reference classes. For example, what must be repeated to answer the question: "Is a person more likely to contract malaria or AIDS?" Must we consider many people contracting malaria but not AIDS? Must we consider many people contracting AIDS but not malaria? Must we consider all people and count how many got either or both diseases? Should we consider only people in Africa or even South Africa? So people tend to fall back on the representative heuristic[14] and reason about frequency observations from their own experience. The term ‘person’ seems unproblematic, until one reflects on the complexity of the example question. Surely, if we consider individual people, all have different likelihoods of contracting either of the diseases. Similarly, different aggregates of people have plausibly different chances of contracting the diseases: those that live in a malaria area versus those that visit a malaria area only sporadically. However general or abstract a ‘person’ we choose to consider, a suitable reference class has to be constructed as an inferential database, and this reference class places the ‘person’ in a concrete context, complete with agency to make causal decisions. Everyday reasoning that acknowledges the problem of constructing a sample space or a contextual aggregate of cases is informal statistical reasoning.

Research about how people construct categories shows patterned contextual reasoning. For example, people reason differently about categories of natural objects than artefacts. The basic difference between these categories is that natural objects are not made by humans for a purpose as artefacts are. In addition, the basis of our inductive inference is influenced by the roles we perceive objects to play in our lives and when we reason about social categories.

Awareness of kinds as a precursor of reference class

Reasoning about an object’s kind is the basis of our everyday inductive inferences about the object, as well as about its unobserved properties. Category relationships are suggested by labels or names given to objects; the visual (form) properties of objects; and the functions an object can perform, or can be used to perform. For example, despite close similarity in form, a cell phone and a television remote control device belong to different categories, on the strength of their different functions. Very young children reason quite easily about kinds through labels (names) attached to objects and the form and function of objects. Even children of pre-school age make inferences about unobserved properties of objects, based on

[14] The representativeness heuristic is at work when people make inferences based on their subjective familiarity with an event. They assign a higher probability to an event that they are familiar with. For example, judging AIDS as a more likely cause of death than malaria, because AIDS related deaths are a familiar observation in our country, is based on the representativeness heuristic. In my example, I link the representativeness heuristic to the complexity of determining a suitable reference class.

their kind or category membership, rather than using perceptual clues (Schulz, Bonawitz, & Standing, 2008). At an even more basic level, Rost (2011) showed that that three year old children learn words best in the presence of variable objects of the same kind, when the supporting objects highlight invariant elements of the referent category as well as those elements that are allowed to vary. Anggoro, Gentner and Klibanoff (2005) showed that children's learning to construct relational kinds is promoted by a combination of learning relational terminology and object comparison. Thus, awareness of kinds and categorising seems naturally linked with learning language and with learning about objects in the world. However, construction of relational or role-governed categories, such as gift, parent, or barrier, is not as simple. Even moving from *dog* and *snake* as feature-based categories to *pets* as a role-governed category, provides obstacles for children younger than nine years old. Gentner and Kurtz (2005) indicate that children initially often treat role-governed categories as feature-based categories in learning: "For example, a brother may be described as a boy about 12 years old, rather than any male (however young or old) who is someone's sibling" (Gentner & Kurtz, 2005, p. 162). Through the work of Luria (1976) we have access to examples of reasoning where school-based learning, or scientific concepts had no influence, because the participants in his research were illiterate and unschooled. I cite two examples that illustrate that opportunities for comparison, as well as relational language were available, yet the participant did not construct the expected kind. Firstly, with regard to categorisation, which is crucial for statistical comparison, Luria showed that naïve categorisation is based not on objective properties of objects, but on subjective understanding of their function. For example, a saw, a hatchet and a log were judged as "the same types of things", because "they work together" or "they are needed together" (Luria, 1976, p. 69). Secondly, categorisation of objects based on their features was hampered by persistent emphasis of the differences between natural objects than similarities between them. The illiterate participants were not willing to categorise a chicken and a dog as animals, due to the observation of the differences between them, for example that a dog has four legs and a chicken two. When the researcher suggested that both were animals, they eventually conceded (Luria, 1976, p. 81).

Natural kinds and artefactual kinds are approached differently when people reason inductively. Whereas basing an inductive inference on a small number of observations was earlier criticised as a misconception, in some contexts the requirement to observe large numbers of cases before generalising falls away, specifically when reasoning about natural kinds. Gelman (1988) used the example of a silver dollar to explain the difference between natural kinds and artefactual kinds: “As long as the substance or structure remains the same, a particular natural entity retains its identity across varied forms (e.g. a silver dollar is still silver when melted), whereas the artefact does not (e.g., a silver dollar is no longer a dollar when melted)” (Gelman, 1988, p. 70). Hence, we tend to generalise with more confidence from limited data about silver, than we do from limited data about coins. The question is whether and how people distinguish between natural and artefactual kinds. Gelman indicate that the distinction between natural kind and artefact is not consistent before children are about eight years old, and artefactual category membership is often decided by the function of the artefact, due to the fact that artefacts are after all made for a purpose. From Gelman’s research it seems that inductive reasoning about artefacts may evoke diverse inferences, since they suggest role-governed categories rather than feature-based categories. So, for example, a huge variety of objects belong to the kind *clothes* based on the role definition ‘you can wear it on your body’, for example shoes, socks, shirts of various kinds and even a cloth. However, since jewellery or glasses are not regarded as clothing, it illustrates that this functional similarity does not serve to include everything one can wear on one’s body as clothes and that artefactual kinds have indistinct boundaries. Clarification of the boundaries depends on our ideas about the shared function, but also on our ideas of typicality and representativeness.

Typical and ideal exemplars

Kinds or categories are constructed on the basis of similarity in some attribute, despite wide differences between other attributes of members of the kind. In order to form categories as reference classes for statistical reasoning, we have to acknowledge the variation, but abstract similarity between objects. Rein, Goldwater and Markman (2010) distinguish between judgements of typicality in feature-based categories and role-governed categories. A good member of the feature-based category *birds* is one that is close to some prototype bird in terms of the intrinsic properties attributed to birds. Role-governed categories such as *pets* are formed mostly on the basis of extrinsic properties, that is, properties of connections between objects bound by the concept *pet*. For example, pets are connected by the extrinsic property of ‘pets belong to people’. Membership of a role-governed category is judged according to the closeness of a candidate to an ideal member, rather than one with prototypical or central tendency (i.e. average) features. If an ideal pet is perceived to be cuddly, a cat is more likely to be included in the category *pet*, than a snake. Rein et al. (2010) illustrate the role of an ideal in the constitution of role-governed categories by example of the category *diet food*: "The central tendency exemplar of a diet food is one with an average number of calories and a mediocre taste. In contrast, the best example of a diet food is one with zero calories and a great taste, even though this exemplar may not even exist" (Rein, et al., 2010, p. 378). While typical or ideal features matter to establish membership of feature-based or role-governed categories, Rein et al. indicate that generalisations to the broader, or super-ordinate categories, are made on the basis of central tendency features, irrespective of the type of category. This finding makes sense, since an ideal is ideal because it is special in some way, rather than mundane or general, and it is unreasonable to generalise from an ideal example. The implication of the findings of Rein et al.'s is that consideration of ideal cases may aid in the process of constructing reference classes of at least role-governed cases, and that reference to ideal cases may provide information about the way objects are categorised into reference classes.

The influence of category formation on statistical reasoning is also evident from the conjunction fallacy. The well known Linda problem provides an example. Kahneman, Slovic and Tversky (1982) provided a description of a woman, Linda, and asked their participants to judge the relative likelihood of two statements: Linda is a bank teller (A) and Linda is a bank teller and is active in the feminist movement (A&B). The description read as follows:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Statistically the probability of A is greater than the probability of A & B, since not all bank tellers are feminists. Yet, many people fail to reason in this way, and commit the conjunction fallacy. According to Sloman, Over, Slovak and Stibel (2003) most people judge A & B to be more likely, because they consider Linda's properties, rather than the abstract relationship between the two categories, namely the set of bank tellers include the set of feminist bank tellers. Their reasoning is based on perceived typical features of a feminist. I pondered the Linda problem from the perspective of category formation. It seems plausible that the question to judge the likelihood of Linda being a bank teller or a feminist can be interpreted as a question about membership of a category. The category of bank tellers is a candidate for a role-governed category that relates Linda as a bank teller to the reasoner as a client. If this is the case Linda's membership to the category is likely to be judged by comparison to an ideal bank teller. Yet, the information does not paint any ideal features of a bank teller, such as accuracy, mathematical knowledge, or people skills. The information rather describes a typical feminist, and therefore Linda is more likely to belong to the category of feminists. According to Sloman and colleagues (2003) people normally form categories from a perspective that provides access to the category's internal structure – they naturally take an inside view and possibly consider prototypical instances in order to decide if a particular case belongs to a category. Sloman et al.'s explanation does not differentiate between different kinds of categories.

Category formation bears on the clarity of reference classes and eventually on sampling concepts. Typicality and representativeness are related to statistical measures of the centre of a data set. While nontechnical dictionaries tend to define mean, median or mode in terms of their calculations, they routinely relate the statistical terms to everyday meanings of average. In turn, synonyms for average are norm, normal, common, par, standard, usual, typical or representative.[15] Hence, the way we informally construct ideas of typicality is relevant to informal statistical reasoning as it may foreshadow ideas about average in statistical reasoning. As demonstrated in the Linda example, from a perspective of clarity of the reference class, our ideas of typicality or ideals in relation to categories can also lead us to reason in statistically incorrect ways. The implication for informal statistical reasoning lies in the use we make of ideal cases. References to ideal cases (indicative of local views of data) are productive to compare cases for the purpose of inclusion in role-governed categories, but basing inferences on ideal cases would be inappropriate and not considered statistical reasoning.

3.2.2 Awareness of variability

The terms typical and representative only have meaning in relation to variation. Antonyms of typical are unusual, unrepresentative and uncharacteristic (Opposite Dictionary).[16] If there is no variation, there is no need to decide on a typical exemplar or a representative measure. In

---

[15] Merriam-Webster online dictionary, accessed on 15 October 2012.
[16] The Opposite Dictionary website, accessed on 12 November 2012.

contexts that are perceived to be homogeneous, people find it easier to conceive of a typical or representative object than in contexts that are highly variable.

Research about informal statistical reasoning suggests that people's intuitions about the homogeneity or variability of properties of some reference class play a role in their willingness to make inferences, whether based on a small number or large number of observations. People believe that when a collection is homogeneous with regard to a property, a higher percentage of the population will be similar to the sample, even when the sample is small. Hence, they are willing to make an inference based on small numbers. When people believe a situation is heterogeneous, they are uneasy about generalising from a small number of observations (Nisbett, et al., 1993). Natural objects, as opposed to artefacts, are usually homogeneous enough in their intrinsic attributes to allow sensible generalisation from small numbers (Gelman, 1988). For example, one will not confuse a dog with a cat even if both are atypical examples, and knowing one dog is sufficient evidence to identify a second animal as a dog. Similarly, people generalise with more confidence when the generalisation is based on observable features of objects, rather than on relations between objects. Generalisation from small numbers of observations, even of natural object categories, can be problematic if the objects are constituted in a role-governed reference class: it would be incorrect to deduce on the basis of having a dog as a pet that all dogs are pets.

In problems with a social context, the relationship of the person with the social group she has to make inferences about, influences perceptions of heterogeneity or homogeneity. In general, reasoning about a group that one belongs to, tends to be based on greater awareness of the heterogeneity in the group. People tend to view an out-group as more homogeneous than the in-group (Quattrone & Jones, 1980) and are more willing to generalise on the bases of limited observations about an out-group than about an in-group. It seems that if the boundaries of a reference class is clear (such as for an in-group) familiarity with the members of the reference class allows one to reason with awareness of variability, while unfamiliarity elicits reasoning about similarities rather than differences. It may be that the difference is related to the need to establish the boundaries of the out-group, before attention can shift to variation within the category. Quattrone and Jones (1980) guided their research participants to reflect on similarities of members of an in-group, and then investigated to what extent their inferential reasoning about the behaviour of a single member of the group was influenced by these contextual central tendencies. They found that forced contemplation of central tendencies (how members of an in-group are similar) caused their participants to be more willing to generalise for the in-group. Nisbett and colleagues (1993) concur that the extent to which people base inductive inferences on small numbers or large numbers "may in part be a function of arbitrary encoding and retrieval factors that accidentally emphasise either the homogeneity or the heterogeneity of events" (Nisbett, et al., 1993, p. 46).

3.2.3 Recognition of sources of variation

Reasoning based on a deterministic understanding of events is generally contrasted with reasoning based on the recognition of non-deterministic variation. Yet, most everyday events reflect a complex interaction between unexplained, chance variation and cause-effect relationships that can be explained. Wild and Pfannkuch (1999) argue that probabilistic thinking should not be seen as an alternative to deterministic thinking, "but something to be grafted on top of the natural thinking modes that directly address the primary problem" (Wild & Pfannkuch, 1999, p. 238). Nisbett, Krantz and Kunda (1993) give examples of everyday observations, where people had no difficulty recognising random or chance components. They indicate that most people have no problem understanding that a good player can to a large extent control the trajectory when kicking a rugby football, while the bounce of the ball remains random. Similarly, students can recognise the unpredictability of academic test performance, by considering their own repeated performance. They understand that one can, to some extent, control one's own test performance by effort; but that a random element remains. However, in situations where the random component is not as salient, people tend to reason on the basis of their direct experiences, failing to view their own experience as a sample consisting of one observation. Typically, people tend to avoid a restaurant where they had one bad experience, falling back on the faulty heuristic 'law of small numbers'. In experiments where Nisbett and his colleagues (1993) manipulated the problem to enable the recognition of chance factors, people did reason in statistically appropriate ways. One example suffices. Two scenarios were presented to a group of students who had to make a decision on behalf of a protagonist. In the first scenario, the protagonist David L. had to choose between two colleges that he wanted to attend. He asked the opinion of like-minded older friends at the two colleges, but then decided to visit both colleges himself. He had a pleasant experience at the college about which his friends complained most. The participants in the study had to decide between the colleges on behalf of David L., and 74% decided to choose on the basis of the personal visit. In the second scenario, David L. systematically drew up a long list, for both colleges, of courses that interested him; places to see; and people to speak to. He then randomly selected from the list the activities to pursue on visits to both colleges. With this scenario, non-statistical reasoning dropped to 56%. The participants were more likely to comment on the adequacy of the sample of friends and the sample of one visit than in the first scenario, where there was no probabilistic cue.

Expert knowledge of a data-context

In Nisbett et al.'s (1993) study of informal statistical reasoning, participants with expertise in a particular context, for example sport, preferred statistically-based explanations in a problem from the particular context; while non-experts preferred deterministic, cause-effect explanations. The researchers argued that context expertise raises awareness of the causal factors at work in a context, but also of the remaining uncertainty. However, experts appropriately chose to base their reasoning on the causal factors where it made sense, and in such situations did not reason statistically. Evidently, expertise is a double edged sword, since "expert knowledge … may encourage unreflective assumptions that the future will resemble the past and that populations will resemble samples, and substitutes either statistical reasoning or reasoning in accordance with well-established causal rules" (Nisbett, 1993, p. 42). Indeed, recent studies confirm this observation. Hall, Ariss and Todorov (2007) refer to the "illusion of knowledge", and showed that increasing event-specific information lead to less accurate, but more overconfident predictions.

Subjective-objective perspectives on the data-context

Nisbett et al. (1993) identified different tendencies in informal inductive reasoning, which they ascribe to a subjective/objective distinction in the nature of their task-contexts. They presented their participants with two classes of problems. The first class covered problems that could be dealt with by objective[17] means, such as abilities; achievements; and physical illness. The second class dealt with problems about personal preferences among objects;

[17] Nisbett et al. do not define the meaning of "objective", but I assume the term refers to situations with observable attributes, especially attributes that can be measured in some way.

assessment of leadership potential; and judgements about the need for sexual fidelity in relationships. They found that objective problems were more likely to evoke statistical reasoning (57% of participants provided statistical responses to objective problems against 26% statistical responses to subjective problems), since the sample spaces and units of measurements are relatively clear, and the role of chance relatively transparent. Although it seems plausible that statistical answers are not necessarily appropriate for the kinds of subjective situations used in their problems, Nisbett et al. maintain that the high degrees of uncertainty in these subjective problems would make statistical reasoning appropriate, at least for deliberation, if not for decision making. I am interested in examples of non-statistical responses to problems that were seen as objective by Nisbett and his colleagues. It is in these responses that a social effect, this time constraining statistical reasoning, seems evident. I will discuss one example.

An objective problem posed to participants was to determine the best playoff system between football (Super Bowl) and baseball (World Series). The participants were presented with the following claim and substantiation, and had to write open-ended answers in support or defence:

> Charlie says that the Super Bowl is the best way if determining the world champion because, according to him, the seven games of the World Series are all played in the home cities of the two teams, whereas the Super Bowl is usually played in a neutral city. Since you want all factors not related to the game to be equal for a championship, then the Super Bowl is the better way to determine the world championship. Which procedure do you think is a better way to determine the world champion…? Why? (Nisbett, et al., 1993, p. 43).

Informal statistical answers to this task took into consideration the role of chance and the need for a larger sample, for example: “World Series is better. Anyone can get lucky for one game, but it is harder to be lucky for four. Besides, being home or away is part of the game, you don’t play on neutral ground during the season.” Non-statistical answers were laden with human causal influence, for example: “Super Bowl, because of neutral ground and also a one shot deal – either you make it or break it – one chance. The pressure is on to perform the team’s best” (Nisbett, et al., 1993, p. 43). The participant turned the objective problem into a subjective one. The subjective preference for a particular team to win influences the reasoning, and I can sense that the winning team for this participant would be *the best team* in a very real sense.

To research reasoning on explicitly subjective problems, participants had to provide open-ended responses to a question like the following one:

> It is the first week of the winter term. Henry signed up for five classes, but plans to take only four. Three of these classes he knows he wants, so he must decide between the other two. Both are on subjects interesting him. The student course evaluations indicate that Course A is better taught. However, he attended the first meeting of both classes this week, and found Course B's session more enjoyable. Which class should he take?

An example of a non-statistical answer was "He's got to choose for himself", while a statistical example indicated awareness of the limitations of a sample of one observation: "You can't tell from one time – thus a survey that is over a longer range is better. Although Henry's idea of a good class could be different from most students" (Nisbett, et al., 1993, p. 89).

The law of large numbers loses relevance when people reason about personal preference or situations where practical action is required, but it becomes relevant when events are multifaceted and complex, where people start to doubt the relevance of their own experience (Nisbett et al., 1993).

In the following sections I review research about colloquial reasoning. Here the question is what constitutes good everyday reasoning, rather than good (informal) statistical reasoning. I will consider the implications of colloquial reasoning patterns for informal statistical reasoning. A refrain throughout the literature on informal reasoning is the tendency to initially reason from a personal perspective, setting up kinds and categories, examples and experiences according to subjective needs and criteria. We seem to imagine ourselves actively immersed in a proposed context, replete with our own references and personal causal beliefs, and reasoning with two main goals: to establish the prudence of actions; and to determine the truth of claims. Informal reasoning literature distinguishes between such practical and theoretical reasoning and the effect of evidence and explanation in the informal inferences and decisions people make. From a statistical reasoning perspective Fielding-

Wells (2010) draws attention to the difficulties of relating problem, evidence and solution in the planning stages of statistical inquiry.

3.3 Practical and theoretical reasoning

Walton (1990, p. 405) distinguished between practical and theoretical reasoning as follows:

> *Practical reasoning* is a kind of goal-directed reasoning that seeks out a prudential line of conduct for an agent in a particular situation. *Theoretical (discursive) reasoning* seeks evidence that counts for or against the truth of a proposition.

Hence, in theoretical reasoning (also referred to as truth-testing reasoning), premises are assumptions that point to a result. Support for the relationship differs in the extent to which the evidence called on can be employed deductively, inductively or plausibly (as opinion). Practical reasoning (also referred to as deontic reasoning) is "characteristically based on uncertainty or incomplete knowledge of a particular (changing) situation. Premises describe goals and knowledge. The conclusion is an imperative" (Walton, 1990, p. 205). In particular, practical reasoning is a search for necessary and sufficient conditions to achieve the goal of the reasoning task. Although reasoning usually has a forward direction, from premise to result, reasoning backward, from an observed effect to possible causes is a practical reasoning move that allows reasoners to creatively propose new cause-effect or premise-conclusion relationships. This process of abductive reasoning is then followed by truth-testing reasoning.

By these definitions, statistical reasoning comprises both theoretical and practical reasoning – theoretical reasoning by to test proposed relationships between variables, and practical reasoning to apply statistical results in contexts. However, Walton's proposal that the conclusion of practical reasoning is an imperative, cannot be applied uncritically to statistical reasoning. For example, arguing the question "Should government pay child support grants to teenage mothers?" is, by Walton's definition a practical reasoning task. The question arises from a practical conflict around the need (premise) to provide care for vulnerable children; knowledge of the context is uncertain, changing and incomplete; and a decision to act is the goal of the reasoning. But as it stands, the question can only be answered by plausible means. Any stronger support for a decision necessitates abductive reasoning in order to establish a premise-conclusion statement that can be supported or refuted by evidence. It is the reasoning processes by which we set up relevant premise-conclusion relationships that are of interest in a study of statistical reasoning at the start of the investigative cycle. The child support question can be reasoned statistically only if we test the plausible truth of statements such as "since child support was instituted there has been less malnutrition among babies" or "child support for teenage mothers leads to more teenage pregnancies" by gathering evidence for the necessity and sufficiency of each premise-conclusion pair. A statistical reasoning approach (based on clear reference classes, awareness of variability and identification of sources of variability) will provide more information about premise as well as conclusion, but constitutes a descriptive, tentative, rather than imperative conclusion. In essence, the statistical reasoning task hinges on identification of measurable premise-conclusion claims, and of what we are willing to take as evidence in support or refutation of abductive claims.

3.4 Evidence in informal reasoning

Means and Voss (1996) emphasise the role of evidence in their definition of informal reasoning: "…we regard informal reasoning as a goal-dependent process that involves generating or evaluating (or both) evidence pertaining to a claim or conclusion" (Means & Voss, 1996, p. 140). In arguments informal reasoning is evaluated in terms of three criteria: (a) the acceptability of the supporting evidence; (b) the relevance, or support of the evidence for the stated relationship between the claim and conclusion; and (c) whether the proposed evidence can withstand evidence for the contradiction of the claim. Means and Voss argue that people's beliefs influence their evaluation of evidence, hence people whose beliefs are not based on recognition of variation, chance factors, and appropriate reference groups, may not accept statistical data as evidence for or against a claim.

Informal reasoning accepts many kinds of evidence that are not acceptable in scientific knowledge domains like statistics. For example, citing authority or expert opinion as evidence is a widely used strategy in rhetoric, but is not acceptable in statistical arguments.[18] During informal reasoning, narrative or anecdotal evidence is routinely offered. Such anecdotes are usually based on a limited number of cases, and are not necessarily typical of a situation; such as in the example of choosing a college on behalf of David L., mentioned

[18] This statement has a caveat: the authority or expertise of the entity that provided statistical information should be unquestionable, judged by the methods used.

above. In the David L. example, evidence for the best college was taken as David L.'s experience during a single visit. It is important to identify appropriate types of informal evidence available in the informal reasoning repertoire for the development of statistical reasoning.

3.4.1 Types of informal evidence

Kuhn (1991) describes a range of evidence types, which varies from anecdotal to correlated-change evidence, all uttered in informal language during informal reasoning. Kuhn found that participants were not forthcoming with evidence for their causal claims. Evidence only followed explicit prompts by the interviewer, asking "How do you know that this is the cause?" or an explicit call to justify, even to the explicit call for "evidence", pronounced with emphasis, and requesting a statement or action to "prove" the assertion (Kuhn, 1991, p. 44). Although none of the participants were unable to provide evidence, Kuhn stated that the majority of responses failed to provide genuine evidence. The most prevalent response was pseudo-evidence, while a third, smaller category of non-evidence also emerged. Kuhn's criteria for evidence are as follows:

a) *Genuine evidence*. "The criteria adopted for genuine evidence … are simply that it (*a*) be distinguishable from description of the causal sequence itself and (*b*) bear on its correctness." The antecedent and the outcome are differentiated by the participant, rather than embedded in a narrative scenario as in the case of pseudo-evidence (Kuhn, 1991, p. 45).
b) *Pseudo-evidence* "…takes the form of a scenario, or script, depicting how the phenomenon might occur…description, either in the form of one or two specific instances or in more general summary form, serves to show how the events depicted might plausibly lead to the outcome … evidence by illustration" (Kuhn, 1991, p. 65).
c) *Non-evidence* "covers a range of responses in which subjects (*a*) imply that evidence is unnecessary or irrelevant, (*b*) make assertions not connected to a causal theory, or (*c*) cite the phenomenon itself as evidence regarding its cause" (Kuhn, 1991, p. 82).

Comparing the examples of pseudo-evidence and weak forms of genuine evidence cited by Kuhn, with the examples of non-statistical (deterministic) reasoning provided by Nisbett (1993), confirms that they are indicative of the immersion of the reasoner in the data-context, often to the extent of explaining how the reasoner would act in a similar situation. For example, a participant who tried to provide evidence for her claim that lack of family support causes prisoners to fall into crime again, acted out the following imagined conversation: "…he knocks on the door, and she says, "Who are you? He'll say, "Well I'm your husband." She will say, "Well, my husband wasn't in prison" (Kuhn, 1991, p. 69). Inside-views and subjective considerations are the order of the day as participants drifted away from the causal sequence and narrated experiential and life-like stories, engulfed by the complexity of the context. Alternatively, participants shifted from providing evidence for the cause, to the consequences of the outcome, evident from a failure to distinguish between "cause and remedy" (Kuhn, 1991, p. 73). Within a discussion, pseudo-evidence can serve its function as giving "a plausible account of a course of events (e.g.) that might lead a prisoner to return to crime" (Kuhn, 1991, p. 70). Kuhn confirms the role of good pseudo-evidence in everyday reasoning in order to heighten our interest in testing an abductively constructed causal theory. This function is important, given the restrictions on our resources to test every conceivable theory.

Genuine evidence offered by Kuhn's participants was classified according to strength. Weak forms of genuine evidence were external evidence that introduced counter-factual arguments to the stated claim, and indirect evidence such as analogy. *Co-variation evidence* was the strongest type of genuine evidence and the most common co-variation evidence "show[ed] some reliance on the presence of co-variation between alleged causal antecedent and outcome as support for the theory that the antecedent causes the outcome" (Kuhn, 1991, p. 45). Kuhn categorises three types of co-variation evidence, in order from weaker to stronger:

3.4.2 Correspondence

The weakest form of co-variation evidence "does no more than note a correspondence, or co-occurrence, of antecedent and outcome" (Kuhn, 1991, p. 46). Two of Kuhn's examples are relevant to informal statistical reasoning. In the first example the participant proposed a causal relationship between repeated criminal action[19] and unemployment, and offered the following evidence:

> (Prompt: How do you know that this is the cause?) Probably from the reading I've done, the associations I've made in my mind, in reading the *Daily News* accounts of people who are involved in crime, and it seems that they tend to be unemployed. They tend to be lower-income people" (Kuhn, 1991, p. 46).

This participant seems to have constructed the reference class as "people who are involved in crime" rather than the reference class of repeated offenders. Awareness of variation is evident from the use of the phrase "tend to". The discourse in this example indicates an aggregate view through reference to many observations by the use of the plural "people who are involved in crime" and "they tend to be". The noted correspondence between crime and unemployment may serve to establish a typical feature of people who are involved in crime. The use of the plural suggests that the frequency of observations of correspondence will give weight to the evidence (most or all will "prove" the causality) and that typicality is judged by frequency. In the absence of reasoning that compares repeated offenders in subcategories (e.g. those with and without jobs) frequency provides a notion of central tendency. Similarly, the second example[20] shows awareness of a reference class of "kids with parents who force them to do well", rather than children who fail in school. Again frequency of observation of the correlated events lends weight to the evidence:

> (Prompt: How do you know that this is the cause?) Lots of times parents who force their kids to do well, you know, force them to study, lots of times those kids will be the ones horsing around because they're angry, and it makes them want to go against their parents and not do it [do well in school]" (Kuhn, 1991, p. 47).

Kuhn notes the absence of explicit quantitative terms for the observed correspondence and codes the response in the second example that cites "lots of times" as closer to the ideal of quantitative comparison than the response that cites moderate modal terms like "probably" and "tend to be" in conjunction with frequency like in the first example. From a statistical reasoning perspective the tentative nature of the evidence reflected in the term "tend to" is an indication of appropriate consideration of variation and uncertainty in informal statistical

[19] The initial question to elicit proposals of causal relationships was: "What causes prisoners to return to crime after they're released" (Kuhn, 1991, p.16).
[20] The initial question to elicit proposals of causal relationships was: "What causes children to fail in school?" (Kuhn, 1991, p. 16).

inference (Makar, et al., 2011). From the perspective of theoretical reasoning correspondence evidence is informal statistical evidence, since it calls for observation of many instances of correspondence. However, in both examples the reference classes are inappropriately constituted.

Kuhn's examples of more sophisticated correspondence evidence involve proposals of a study of some kind, such as "a study of cases of students…who drop out of school…and sees where they have family problems" or "look at kids that are failing in school and see if the parent or parents are doing their job" (Kuhn, 1991, p. 46). In these examples high frequency of correspondence again provides inductive strength to the argument. Clearly aggregates of *school failures* or *people involved in crime* or *kids with parents that force them* are role-governed categories. I have shown earlier that it is difficult to conceive of a typical member of such categories, since members have few common features, although they share a single relationship. It may well be that frequency evidence is evoked by this type of category.

3.4.3 Explicit Co-variation

Explicit co-variation evidence[21] differs from correspondence evidence in the use of comparison and quantification: "Instances that represent one level of the antecedent are compared with those that represent another; the comparison is with respect to incidence of the outcome" (Kuhn, 1991, p. 47). Participants who offered explicit co-variation evidence compared subgroups of two reference classes, for example, those who are jobless and involved in crime, and those who have jobs and are involved in crime. Evidence for the causal relationship between job status and crime would be the higher frequency of jobless people involved in crime. Participants who gave explicit co-variation evidence, used frequency comparison phrases like "on the whole" and "as a whole", or referred to percentages. Evidence that explicitly referred to the need for large scale observation (e.g. "take a survey") is coded as more sophisticated than an unspecified number of correspondences. The use of explicit co-variation evidence is therefore informal statistical reasoning, since it is based on intuitive understanding of the law of large numbers, takes into

[21] Kuhn (1991, p.47) uses "Covariation evidence" both as a kind of genuine evidence and as a category of the kind. I use her description of the category, namely "in this category the idea of covariation becomes explicit" to name the category "Explicit co-variation" in order to avoid confusion.

account meaningful reference classes and shows awareness of variation. The strongest form of evidence described in Kuhn's study is correlated change as evidence.

3.4.4 Correlated change

Kuhn writes that "a stronger form of evidence for a causal relationship, one in which change in the antecedent co-occurs with change in the outcome. Co-occurrence of change increases the likelihood that change in one factor is responsible for the change in the other" (Kuhn, 1991, p. 50). Correlated change as evidence is the kind of evidence statistics requires, especially in regression analysis.[22] The underlying reasoning process is to affect change in one variable and then observe whether there is change in another variable as a result. Judging by the dearth of this kind of evidence in Kuhn's study, it is clearly an indication of sophisticated reasoning.

Yet, Kuhn also cites examples of correlated change reasoning that refers to personal experience and involves a single case: "I had a son [who did poorly in social subjects]…I was trying to show him…he started doing better. So to me that is proof enough…it seems to me it's very obvious" (Kuhn, 1991, p. 51). However sophisticated correlated change evidence may be in informal reasoning, when it rests exclusively on personal experience, or single case comparison, such evidence is not indicative of informal statistical reasoning. Statistical reasoning requires an aggregate view of cases.

Embedding her examples in the surrounding discourse between experimenter and participant, Kuhn explains that participants' responses to a request for evidence often spanned different categories. Most commonly participants at first offered pseudo-evidence, but then proceeded to some form of genuine evidence when the interviewer probed further. When participants offered non-evidence, they did not proceed beyond pseudo-evidence regardless of prompting. According to Kuhn, education level correlated significantly with the kind of evidence given across contexts in her study. Co-variation evidence was the most common type of genuine evidence, although less than half of all the participants gave genuine evidence.

---

[22] I am not claiming that correlated change evidence per se allows us to infer causality in Statistics. Kuhn (1991, p. 50) was also cautious when putting it that: "Co-occurrence of change *increases the likelihood* that change in one factor is responsible for the change in the other." [emphasis mine] I am claiming that correlated change is at the root of statistical inference.

Of importance is the relationship between the ability of participants to conceive of different causes for a stated outcome and their ability to offer genuine evidence. Kuhn (1991, p. 93) states “Subjects whose theories are of the multiple cause type are about equally likely to offer genuine evidence versus not. Subjects whose theories consist of a single causal sequence, in contrast, are more likely to offer pseudo-evidence that provides only an elaborative description of the theory. This pattern holds for each of the three topics and is statistically significant in each case”. This finding is in line with Nisbett’s finding that explication of the sources of variability supports informal statistical reasoning, and that expert knowledge of a context provides better understanding of the causal factors at work as well as of the remaining uncertainty. Kuhn found that co-variation evidence was more likely when participants had personal knowledge of the context. The importance of awareness of multiple sources of variation, or of multiple causal factors, necessitates a discussion of the role of explanation in informal reasoning. Causal theories are “inferences to the best explanation”, (Groarke, Fall 2012) often derived through abductive reasoning.

3.5 Explanations in informal reasoning

Glassner, Weinstock and Neuman (2005, p. 105) provide a succinct example of the role of explanations compared to evidence in argumentation:

> When someone introduces a claim (e.g. ‘classical music increases cows’ production of milk’), there are two indispensable basic questions that can be asked in order to justify the claim: ‘How do you know?’ and ‘Why is it so?’ (Donaldson, 1986). The first one questions the extent of the truth of the claim and should be answered by presenting relevant evidence, which can be used as an indication of the truth of a claim (e.g. ‘After hearing classical music there was more milk in the tanks’). The second one questions the causes of a claim and should be answered by presenting relevant theoretical explanations (e.g. ‘Classical music relaxes the cows and relaxed cows produce more milk’). A satisfactory justification of a claim should consist of both.

This example is an appropriate heuristic for the relationship between theoretical and empirical justification in statistical reasoning. The question ‘How do you know?’ can be answered with statistical evidence,[23] while the question 'Why is it so?' cannot, since it calls for an explanation. However, an answer to the 'Why is it so?' question can lead to another 'How do you know?' question; there had to be a primary 'Why is it so?' question that produced the 'How do you know?' question in this example. Asking "How do you know?" can help to formulate and test explanations. In complex real-world data contexts, there are many possible answers to any 'Why is it so?' question and reasoning about the concomitant factors is necessary to grasp the system dynamics in order to choose the theory with the best explanatory power.

3.5.1 Mechanism-based explanations

According to Ahn, Kalish, Medin and Gelman (1995), when given the choice, people prefer mechanism-based explanations above evidence based on co-variation.[24] Similarly, Donaldson (1986) reports that children choose an explanation to answer a 'How do you know?' question, rather than evidence. Kuhn's (1991) research with adolescents and adults shows comparable results, where real evidence was provided by less than half of the participants, while the rest provided pseudo-evidence that hinted at mechanisms at work. Mechanism-based explanations refer to the necessity of the proposed cause. Necessity is established by consideration of counterfactuals, through reasoning 'What if a particular factor had been different?' Cummins (1995, p. 647) confirms that informal causal deduction "is exquisitely sensitive to … *alternative causes* and *disabling conditions*." The search for explanations or reasons to believe or *not* to believe a statement, suggests routines of everyday reasoning that may influence the development of statistical reasoning. Cummins proposes that alternative causes and disabling conditions influence causal deduction, because they affect beliefs about necessary and sufficient conditions for a given factor to cause an observed effect. An awareness of alternative causes vitiates the argument that a given cause is necessary for the given effect. An awareness of disabling conditions casts doubt on the sufficiency of the given cause to bring about the given effect. Deliberation of alternative causes and disabling

[23] I do not imply that descriptive statistics can answer the causal question satisfactorily.
[24] Far from implying that one is better than the other, I agree with Ahn et al., (1995, p. 341) who writes: "However, rather than competing, the two approaches have complementary roles. For example, there might be contexts in which people use co-variation analysis to identify a causal factor and those in which they will seek to identify a mechanism for the effect. Mechanism information can be used to distinguish between true causality and spurious correlation…, and also has several fundamental implications such as projection and beliefs about conditions that might lead to change."

conditions should have a positive influence on the "grasping of the system dynamics" that is required at the commencement of the investigative (PPDAC)[25] cycle. In the following section, I will briefly illustrate Cummins's theory with an example from my study.

3.5.2 Alternative causes and disabling conditions in informal reasoning

Definition of terms:

> *An alternative cause* is simply a cause (other than the one cited in the causal rule under consideration) that is capable of evoking the effect cited in the rule.
> *A disabling condition* is an event that could prevent an effect from occurring in the presence of a viable cause (Cummins, 1995, p. 647).

Consider the question of hypothesising how different variable factors contribute to variation in the prices of used cars. A relevant formal truth-logical statement is, "If a car is old it should be cheap." Formal logic allows the forward entailment, but not the inverse: "If a car is cheap it must be old." Informal deductive reasoning easily hesitates to accept the truth of both statements. Cummins argues that the mechanism of this hesitation is a consideration of alternative causes and disabling conditions. To illustrate, consider the formal causal claim:

If the car is old it is cheap.

Premise: The car is old.

Conclusion: Therefore the car is cheap.

Reasoning with alternative causes:

In looking for alternative causes, we are in effect asking: What makes the car cheap?

Claim: OLD → CHEAP

BUT

ALTERNATIVE CAUSES		SAME EFFECT
small model	will be	**cheap**
no extra's	will be	**cheap**
bad condition	will be	**cheap**

[25] See Figure 1, Chapter 1.

high km reading	will be	**cheap**
cheap make	will be	**cheap**

Hence, old is not necessary, other factors can also cause cheap.

This interrogation can continue for any of the alternative causal claims offered in a discussion, for example, arguing with disabling conditions:
In looking for disabling conditions we ask in effect: Why are there old cars that are not cheap?

Claim: OLD → CHEAP

BUT

CAUSE	Disabling conditions		Not Effect
old car	well looked after	(should) be	**less cheap/more expensive**
old car	with extras	(should) be	**less cheap/more expensive**
old car	with low km	(should) be	**less cheap/more expensive**
old car	of a different type	(should) be	**less cheap/more expensive**

Hence, old is not sufficient to cause cheap, one can find old cars that are more expensive.
As before, any other claim of a causal relation to price can be interrogated.

If people informally judge the truth of causal theories by consideration of other possible causes for an effect than that stated, as well as factors that could prevent an effect from occurring, it is clear that such reasoning can aid understanding of the data-context, and influence decisions about data-gathering and willingness to accept results based on statistical investigation. Cummins showed experimentally that consideration of alternative causes and disabling conditions overrides the semantic structure of individually-generated deductive arguments.

In addition, Cummins found that people are more likely to draw conclusions and accept all arguments (*modus ponens*, *modus tollens* as well as *denying the antecedent* and *accepting the consequent*) in unfamiliar situations, where they are not able to think of alternative causes and disabling conditions. When the scenarios are familiar, people are less likely to draw bi-directional causal inferences. Cummins further indicates that factors that are believed to be "true causes" override empirical counter-examples. For example, drunken driving is seen as a true cause of road accidents even if there are many alternative causes and disabling conditions. I observed in my teaching that when an argument induced shame or embarrassment, people are likely to strongly accept or reject claims and cease to search out alternative causes and disabling conditions.

I want to return to the intuitive and pervasive personalisation of reasoning, which I mentioned earlier. Statistical reasoning and immersed, personalised reasoning are at opposite ends of a continuum. Discursively viewed, contextual information such as claims, evidence or explanations, is text, whether written or spoken. Such texts are indications of how participants structure data-contexts. Hence, I turned to the literature on text comprehension to gain an understanding of how people make sense of texts, and came upon the notion of *the immersed experiencer*. This concept helps to explain personalised reasoning in statistical contexts.

3.5.3 The immersed experiencer

Drawing on cognitive research showing that words activate brain regions in a similar way to actual experiences, Zwaan (2004) introduces the notion of an *Immersed Experiencer Framework* (IEF) in which language is embodied and text re-produces experiential memories together with their referents as mental representations. Although such mental representations are diffuse and weakly organised at first, they yield more and different information than that presented in the text. Zwaan and Kaschak (2009, p. 368) argue that comprehension of language relies on our ability to imagine actions:

> In a very literal sense, the comprehension of a sentence about removing the pie from the oven relies on much the same machinery that would be involved in actually carrying out the action … A number of recent studies have pointed to the conclusion that the ability to mentally simulate actions (and their consequences) is crucial to our ability to plan and execute actions, and to understand the actions of others.

Kintsch (1994) proposes an informal reasoning model of knowledge, which holds that surface level understanding of the syntactic elements of text (both written and spoken) may facilitate recall, but not understanding. To enable deeper level understanding, people construct a propositional network, called a textbase, which is evident from the importance they give to semantic elements. Finally, the textbase is elaborated by and integrated with general knowledge or world knowledge in a personal situation model (Kintsch, 1994, p. 295). According to Kintsch, *why*-questions cannot be answered without reference to such a personal situation model, neither can relations like *because* and *therefore* be distinguished. Hence, personalisation of reasoning in data-contexts can be understood as the construction of a situation model in order to understand the problem.

DiSessa (1993) includes a similar notion of a primitive network of intuitive abstraction from direct experience, rather than language, in his theory of the development of physics concepts. He proposes the formation of a *causal net* of *p-prims* (phenomenological primitives) that are primitive causal lines, for example, *if you push an object in a given direction it will move in that direction.* Through a process of learning, these nets are increasingly structured so that clusters of *p-prims* are triggered in characteristic situations. According to DiSessa, conceptual change occurs through addition of new *p-prims*, by forming new relations between *p-prims* and by changing priority relationships between *p-prims*, hence a restructuring of the causal net.

Zwaan's and DiSessa's theories both refer to the structuring of networks of relationships between causes and effects in order to achieve conceptual change. The mechanism of such restructuring may very well be aided by conscious consideration of alternative causes and disabling conditions for statements to be true, or observations to be explained. DiSessa's theory in particular indicates that not all clusters of *p-prims* are appropriate for reasoning in a specific knowledge field. While the immersed experiencer may take the intuitive first step towards statistical reasoning, one learns to consciously consider the relevant and appropriate experiential or perceptual information to construct reference classes and sources of variability. In discursive terms, the applicability conditions of statistical reasoning must be learned.

Immersion in a data-context seems to be egocentric, aimed at establishing a personal role in the suggested context. My informal observations of people's reasoning suggest that the first naïve reasoning task is how the problem or claim affects the reasoner personally. I have asked countless people in varied situations the same question that I analysed in this study: "What is a reasonable price for a used car?" A typical gut reaction to the question is a statement about an action in relationship to used cars, for example: ‘I don’t want to buy a car’; or ‘Don’t buy a used car’; or ‘I would buy a Toyota’. A plausible explanation is that my unsuspecting participants were orienting themselves in an imagined context, before they seemed to be able to proceed to argue the actual question. In my example, the orientation often took the form of an imagined action, here, *to buy* a used car, which is consistent with Zwaan and Kazak’s theory that language comprehension evokes imaginary actions.

3.5.4 The social reasoning effect

Extrapolating the notion of immersion and the existence of nodes of perceptual primes to naïve causal reasoning, draws attention to *the social reasoning effect*, reported by various researchers of informal deductive reasoning. Cummins (1995) reports that the social reasoning effect is displayed by children as young as three years old, and that it persists in adults’ informal causal reasoning. She explains that “they reason flawlessly about social rules (such as permissions and obligations) but fail miserably on formally identical problems that have non-social content” (Cummins, 1995, p. 646). Hence it seems that causal reasoning about text that suggests a social role to the immersed experiencer, such as that of a custodian of permissions or obligations, is easier than reasoning about text that is not about social relationships.

Contrary to the mediating effect in informal causal reasoning, Nisbett, Krantz and Kunda (1993) maintain that errors in informal statistical reasoning are particularly pertinent when people reason about social behavior. In such situations, clear base rate information has little effect on persistent deterministic reasoning that is based on assumed dispositions and traits of the people in the sample. In a study by Nisbett and Borgida (1975), participants were told of another study in which those participants heard someone in a nearby room having a fit of some kind, and were observed in terms of their willingness to help the person in distress. Nisbett and Borgida’s participants consistently judged the personalities of the helpers, and those who refrained, despite information about the general or typical behaviour of bystanders in similar cases. It seems that helping was interpreted as a social obligation and not-helping was judged as bad. Reflecting on other examples of non-statistical reasoning provided by Nisbett, Krantz and Kunda (1993) in relation to the social reasoning effect, suggests an interesting hypothesis: perceptions about social obligations evoke deterministic reasoning.

Consider for example, the question about choosing the best playoff model between the Super Bowl and World Series. The deterministic answer is: "Super Bowl, because of neutral ground and also a one shot deal - either you make it or break it - one chance. The pressure is on to perform the team's best" (Nisbett, et al., 1993, p. 44). If a sports team has a (social) obligation towards its supporters to win by doing their best, rather than being subject to chance factors, this answer is sensible. Similarly, if a restaurant has an obligation to serve good food, a disappointed customer may judge that the obligation has been violated and constitutes a social transgression. Another explanation is offered by Schwartz and Goldman (1996, p. S100):

> …people are predisposed to treat contexts involving people, such as the Linda example, differently than contexts involving marbles in an urn. This predisposition exists because of the intuitively-based understanding that certain properties of people 'cause' their actions, opinions and decisions. This leads to a tendency to think about sampling people in causal terms rather than in chance or random terms. No such 'causal' understanding exists for marbles in an urn so people reason about sampling marbles in chance terms.

The notion of an immersed experiencer and a social effect on reasoning provides an alternative explanation for statistics students' typical interest individual cases "for instance, Who else is like me? Who is the tallest? Who has the most?" (Konold & Higgins, 2003, p. 203). Even when the question specifically asks students to compare groups, novice statistical reasoners initially compare individuals in the groups. I observed similar reasoning in my statistics course. For example, every year students in my statistics class do an online speed reading test and have to compare the typical reading speed of female and male students in the class. Many students conclude that the group with the highest maximum reading speed reads faster, regardless of other features pertaining to the distributions of reading speeds. In addition, they routinely want to verify exactly who the fastest reader is. From the perspective of an immersed experiencer, such comparisons make sense if a high reading speed is perceived as an obligation – education students are supposed to read well – and the reasoner's intuitive task is to position herself and her class mates in relation to the ideal case that complied with the obligation, namely the fastest reader.

Kuhn (1991) argues that competent argumentation, that is, the contemplation of different plausible causes related to a given effect, as well as finding evidence to support proposed cause-effect relationships, is not a given. Competent reasoning is described by Cummins (1995) as being able to reflect on one's own reasoning and exercise control over one's beliefs. Hence, intuitive beliefs are a given at the start of an informal reasoning process, evoked by the con-text of the data. It seems that our intuitive beliefs are mostly of a causal nature, and that social behaviour is intuitively judged in terms of compliance or violation of some assumed obligation. The suggestion is that these intuitive beliefs are reflected on and restructured only as a second-order process:

> In the absence of this ability [to reflect on and control intuitive beliefs], one's beliefs are utilized as a basis for organizing and interpreting experience, but only by means of this second-order, reflective thinking ability can one think about, evaluate, and hence be in a situation to justify these beliefs. Only in the latter case does one exercise control over one's beliefs (Kuhn, 1991, p. 14).

Mercier and Sperber's work on social argumentation confirms the existence of an initial stage of intuitive inferencing (Mercier & Sperber, 2008) during argumentation. Reflection and restructuring of beliefs are supported by the social deliberation of conflicting beliefs and opinions and requests for justification in argumentation. Their research shows that even specialists are prone to reasoning errors when reasoning on their own. Such second-order deliberation should be evoked during opportunities for learning to reason statistically. Indeed, Fong, Krantz and Nisbett (1993) showed with correlational studies, as well as experimental studies, that training which makes the notion of sample space, chance factors and the law of large numbers salient, markedly improve both the frequency and quality of statistical reasoning in contexts.

3.6 Summary

In this chapter I used personal anecdotes and reflections while I principally argued that personalisation of reasoning may inhibit the development of statistical reasoning. This may seem ironic to the reader. I acknowledge that my own reasoning about the development of statistical reasoning at this stage serves to reiterate my conviction about the ideas to be researched.

I reviewed literature about informal reasoning that is relevant to the development of statistical reasoning. Particularly, I focused on the seminal research done by Nisbett and his colleagues (Nisbett et al., 1993) about the way in which informal inductive, statistical reasoning provides guidelines for the structuring of data-contexts at the start of the investigative cycle of a statistical process. The authors indicate that such structuring of the system relations should ensure that the sample space is clear and that sources of variation (both causal and chance) should be identified. Based on the requirement to clarify the sample space, I reviewed literature that provides information on how people reason informally to form categories. Categories based on intrinsic properties of objects are formed more readily than role-governed categories, and are generalised with more certainty. Role-governed categories, especially those governed by a role in relation to people (such as pets), are constructed by means of consideration of an example's closeness to an ideal; while feature-based categories are added to by comparison with a typical member. A typical member shares most of the features of variable members of the category. Hence, in the process of grasping the system relations of a data-context, the kinds and categories that must be formed to constitute a suitable reference class is expected to be influenced by features, functions and relations of the possible cases.

A further necessary requirement for good informal statistical reasoning is basing inferences on large numbers of observations. Beliefs about homogeneity within a category support intuitive generalisation from small numbers of observations; which may be frugal for practical decision making, but inhibit statistical reasoning. In social situations, such beliefs of homogeneity are based on being a member of the group oneself. Nisbett et al. (1993) showed that specific manipulation of information is needed to draw attention to variability of in-groups and typicality of out-groups, in order to support the appropriate use of the law of large numbers. Objective problems, those based clearly on observable features, such as achievements, abilities, and physical illness, evoked informal formal statistical reasoning to a greater extent than subjective problems. Subjective problems tend to be answered in deterministic ways.

I also drew on Walton's distinction between practical and theoretical reasoning to argue that the purpose of reasoning influences the kinds of evidence or explanations that are accepted. Theoretical or truth-testing reasoning requires evidence to test the truth of a proposition, while practical reasoning is goal-directed and aimed at prudent action in a situation. During informal reasoning, inductive reasoning and abductive reasoning are combined to conceive of possible relationships between variables, and seek evidence to test such hypotheses. I argued that informal statistical evidence must satisfy the necessary conditions for sound informal statistical reasoning, namely a clear reference group, the recognition of variation and chance factors, and must be based on large numbers of observations. Kuhn's (1991) study on evidence in argumentation implies that co-variation evidence, which is appropriate evidence for informal statistical reasoning, is not readily provided by participants. Such evidence is dependent on reflection of one's intuitive beliefs and requires a second-order reasoning process. Personal beliefs also influence theoretical reasoning, and specifically beliefs about social obligations have a negative influence on statistical reasoning. In contrast, Cummins (1995) showed that reasoning about social obligations supports naïve truth testing reasoning. Typically, reasoners search for violations of the obligation. In general, the search for alternative causes and disabling conditions seems to be an intuitive routine in practical or deontic reasoning. Explanations in informal reasoning are used to answer "Why is it so?" questions (implicitly or explicitly posed), and are typically mechanism-based. Kuhn's study about argumentation indicates a tendency to substitute mechanism-based explanations for evidence. Hence, active searching for alternative causes and disabling conditions during informal reasoning about a data-context has potential to generate a range of possible inferential rules that can be tested statistically. This process would support clarification of the reference group, awareness of variation and causal factors, and reasoning with large numbers of cases.

Lastly, I have drawn on text-comprehension literature that posits the existence of mental models of immersion, embodiment and enactment in data-contexts. These intuitive attempts at text comprehension provide alternative explanations for subjective reasoning and consideration of individual cases in informal statistical reasoning. In the rest of my study, I will analyse with reasoning discourse, rather than make inferences about mental models. Drawing together research on statistical reasoning (as reported on in Chapter 2) and colloquial reasoning, I hypothesise that informal statistical reasoning is not merely idiosyncratic, but has routines influenced by colloquial reasoning processes, such as sensitivity to the social reasoning effect and preference for explanations above evidence in data-based contexts.

In the next chapter, I will develop the theoretical framework for this study, namely that thinking is communication, and consequently, that learning is evident from shifts from colloquial discourse towards the literate or theoretical discourse of a particular knowledge field.

Chapter 4: Theoretical framework: Learning as communication

4.1 Introduction

I argued in Chapter 1 that the nature of statistics demands acknowledgement of the central role of everyday data-rich contexts for the development of statistical reasoning. I showed that the role of informal language and informal representations in the development of statistical reasoning is now widely recognised in statistics education research. This legitimates the study of the discourse of informal statistical reasoning. Although the utterances of participants were provided as evidence for claims, the studies I reviewed in Chapters 2 and 3 were not based on discourse analysis. Using the terminology of the relevant fields, I reported on transcripts of utterances that are related to 'factors' that influence informal statistical reasoning (e.g. clarity of the sample space or reference class), reasoning patterns such as searching for alternative causes and disabling conditions when solving socially-based logic problems, and the 'existence' of personal situation models in participants' minds. Similarly, in the review of statistics education literature, I reported on descriptions of levels of statistical reasoning according to the SOLO taxonomy, where utterances are mapped on developmental levels. However, the studies reported in my literature review do not provide a coherent and adequately operationalised theoretical lens with which a researcher can describe and explain how and when reasoning develops from informal to formal statistical reasoning. Discourse analysis of informal reasoning will allow me to hypothesise developing routines (patterned ways of reasoning) and substantiation rules (what makes a narrative true), as statistical reasoning develops.

In this chapter I will summarise Sfard's theory of commognition (2008) as the theoretical framework for my study. Commognition is a theory for discourse analysis, which I chose for its focus on learning as communication, along with its analytical tools for identifying and describing shifts in discourses toward literate (statistical) discourse. As my study has focused mainly on informal reasoning about aspects of data-contexts that impact a grasp of the system dynamics at the start of the investigative cycle (PPDAC), I expected the reasoning of my participants to fall short of formal statistical reasoning. Therefore, I needed to understand the affordances as well as the constraints of their discourse, as we proceeded with learning how to ‘do’ statistical analysis and how to reason statistically.

4.2 Commognition: Thinking as communicating

The term “commognition” coined by Sfard (2008) “encompasses *thinking* (individual cognition) and (interpersonal) *communicating*.” The basic tenet of commognition is that the collective predecessor of human thinking is communication. As a combination of the words *communication* and cognition, the word stresses the fact that these two procedures are different (intrapersonal and interpersonal) manifestations of the same phenomenon” (Sfard, 2008, p. 296). Within commognitive theory, learning is defined as increased participation in a particular discourse, mathematics is defined as a form of communication, and mathematical thinking as a form of discourse. [26]

4.2.1 Discourses

First discourse must be defined:

> The different types of communication, and thus of commognition, that draw some individuals together while excluding some others, will be called *discourses* (Sfard, 2008, p. 91).

Sfard distinguishes between colloquial discourses and literate discourses. Colloquial discourse is that in which we spontaneously engage in our everyday lives, in order to conduct our affairs successfully. Literate discourse is specifically created for scientific communication rather than practical action, and draws together discursants from various scientific communities. Colloquial discourses are the focus of research in fields such as informal reasoning, moral reasoning and/or practical reasoning. Literate discourses have formal and informal versions. Informal literate discourse is used to communicate results of a literate discourse to the colloquial discourse community. The abundance of popular science publications is evidence of such an informal bridging discourse. While all discourses, formal

[26] I will use statistics and statistical discourse in the same way as Sfard uses mathematics and mathematical discourse; not claiming that they are the same commognitive entities, but acknowledging that statistics is a scientific discourse just as is mathematics.

or informal, literate or colloquial, are mediated by metaphors, informal literate discourse is mediated by metaphors that attempt to unpack, embody and enact concepts for the lay person. My own understanding of the concept *valence of atoms* was greatly aided by the metaphor referring to the seating and standing spaces at a rock concert, which I once came across in a popular science text. Despite the efforts of popular science writers to make science discourse available to the wider community, many people remain excluded even if they are willing to participate. Popular science writers are not newcomers to the literate discourses they are communicating about, but they are communicating with newcomers in many instances. In Figure 5 I have drawn a broken arrow between colloquial and informal literate discourse to embody the quest of my study, which is to describe shifts from colloquial discourse to informal statistical discourse:

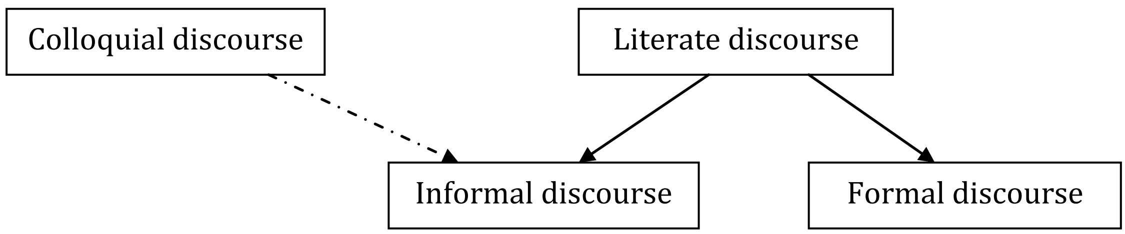


Figure 5: Delineation of discourses

According to Sfard, discourses change continuously, because individual discursants are independent agents, and mismatches in communication is a given. Efforts to correct mismatches lead to changes in a discourse. The goal of education is to effect intentional changes in discourse, to engage newcomer participants in discourses that increasingly approach a formal discourse, and in the case of mathematics, formal discourse is almost unrecognisable to those outside of the mathematical community. A commognitive perspective problematises the relationship between colloquial discourse and informal statistical discourse. One of the differences between mathematics and statistics is that statistics is communicated to a much wider community than that formed by participants in formal statistical discourse. In addition, the objects of statistics are context-mathematics blends. They are not *just* numbers or shapes that can be acted on in isolation of the primary objects from which they were derived.[27] Numbers in statistics are always signifying relations between simple or compound discursive properties of primary objects. For example, 'a standard deviation of 4 units', which is the mathematical result of a calculation, may be a large number or a small number depending on the context; and denotes how actual measurements of a property of an aggregate of objects are distributed. Discourse that doesn't take into account the extra-discursive context is not statistical discourse, but rather mathematical discourse. So for example, Gigerenzer (2004, p. 578) berates the ritual use of "mindless" statistics, such as the habitual use of a 5% significance level for rejecting a null hypothesis. This severance of statistical discourse from contextual discourse is criticised as bad practice. Gigerenzer further argues that predictions based on simple heuristics that people use in their everyday colloquial discourse, can be more accurate than predictions based on statistical rituals. Far from being useless, heuristics like 'less is more'[28] appropriately prevent people from seeking more information than necessary in uncertain situations; but conflict with formal statistical discourse. The aim of this study is to describe change from colloquial discourse to informal statistical discourse, rather than to 'unpack' formal statistical discourse in the way that, for example, popular science writers do. The motivation for intentional change of discourse is people's belief that "the improvement in a discourse may lead to an improvement in practical activity" (Sfard, 2008, p. 117). In the case of statistics, Rao (1999) argues that statistical discourse improves practical decision-making in everyday contexts. He insists that statistical discourse is not a philosophy, but a new way of thinking. It is a pervasive awareness that useful knowledge is the combination of uncertain knowledge and knowledge of the extent of uncertainty.

Commognitively defined, thinking is discourse with oneself and learning is the individualisation of a desired discourse. Reasoning concerns the relationships between sentences, and is therefore a meta-discourse. According to Sfard, discourses are identified by their objects, patterned ways of doing (routines) and their endorsement rules (what counts as true in the discourse). In the following sections, I will summarise the commognitive perspective on the origin and nature of these discourse constitutive entities.

---

[27] The context of primary objects is backgrounded during stages in the statistical thinking cycle, but never completely left behind.
[28] The 'less is more' heuristic is seen as a statistical misconception – the law of small numbers (Tversky & Kahneman, 1971).

4.2.2 Objects and concepts

In commognitive theory, words and other signifiers are discursive objects, and concepts are words or other signifiers and their discursive uses. Thus, mathematical (or statistical) concepts are discursive objects and their discursive uses. I will first summarise the commognitive perspective on objects, particularly mathematical (and by implication statistical) objects, and then explain how concepts are viewed in this theory.

Objects

Primary objects (p-objects) exist independently from our discourse about them. They are the concrete ‘things’ we see and touch, and the sounds we can hear (e.g. a piece of music is a p-object). When we name primary objects, we create simple discursive objects. The act of naming seems completely natural in human discourse, yet shows categorical differences as I have described in Chapter 3. For young children, names for rule-governed categories such as ‘brother’ do not signify objective relational properties, but rather subjective properties. Sfard argues that processes of saming, encapsulation and reification are required to create nouns as new discursive objects. Encapsulation allows us to call a variety of specific animals by the same name, for example ‘cat’, and use the singular to refer to many instances; e.g.: a cat does not have a master. Saming and encapsulation continue in our discourse on discursive entities: in mathematical discourse we create compound discursive objects like *the basic operations* based on saming and encapsulation of the discursive objects of addition, subtraction, division and multiplication; and later still we talk about subtraction and addition as ‘inverse’ operations. Reification is evident from the use of nouns or noun phrases, where previously we used verbs to talk about processes. Thus, reification serves to “eliminate the temporal dimension of phenomena” and condenses discourse, resulting in “thriftiness” (Sfard, 2008, p. 63). An example from informal mathematical discourse is the initial role of number words as signifiers for action on objects, rather than as nouns. When asked how many objects there are, young children typically count the objects and present the complete counting sequence as the answer, rather than last number in the sequence. Hence, it is a temporal process. As their discourse on quantity and number shifts toward formal mathematical discourse, the number names are used as adjectives (seven marbles) before the number word is used as a proper noun (this is *seven*). Once number words are reified as nouns, they can be acted on to create new verb phrases like *five plus two is seven*. Another iteration of objectification is needed to use the noun phrase *seven is the sum of five and two*. And again, once the new discursive object *the sum* is created, we can act on it discursively. For example we can compare the objects: *the sum of five and two is smaller than the sum of three and two*. The objects of formal mathematical discourse are highly abstracted discursive entities, which were created for the purpose of communication through iterations of the processes of saming, encapsulation and reification[29]. Concurrently, when noun phrases replace verb phrases in our discourse on mathematical objects, the agency of human actors fades into the background and the objects become alienated, eventually gaining the same kind of mind-independent status that tangible objects have. This is clear from the way mathematical objects figures in our use of language. For example, when we say “this *is* a triangle” instead of “we shall call a shape a triangle if only if it has three straight sides”, the triangle seems to exist independent of our discourse, independent of what we chose to call “it” (Sfard, 2008, p. 57).

As I have shown in Chapter 2, the use of the nouns *clump* or *hump* in informal discourse about statistical distributions is hailed as a sign of the development of the concept of *shape* of distributions (Bakker, 2004a, Cobb & McClain, 2004, Makar, 2004). Commognitive analysis confirms that *words like clump* or hump are discursive objects mediated by the shape of a graphic distribution. These words are routinely used in learning situations as early metaphors for the abstract statistical notion of distribution. The use of the noun *average* has been shown to obscure differences between the objects or processes that were encapsulated and reified, specifically the difference between the median and the mean. Although *clump*, *average* and *mean* are all nouns, they belong to different object categories. *Mean* is an abstract discursive object whose construction involved reification and that belongs to literate statistical discourse. The object status of *average* is less clear. Discursants who use average as an adjective sometimes refer to concrete discursive objects (e.g. a person of average height) and at other times to abstract discursive objects (e.g. the average number of children per family is 2.3 children). Discursants that use average as a noun may refer to a number that is a middle value in the sense of the median or the mode, or in the sense of a central interval of numbers. In colloquial discourse abstract statistical objects like the mean are often renamed with an accompanying change in discursive status. For example, the *book price* of a used car is a

[29] Saming: giving one name to objects previously not considered as being the same; Encapsulation: telling stories about samed objects in the singular rather than the plural; Reification: replacing talk about processes with talk about objects.

statistical summary that hides a multitude of discursive decisions. In everyday discourse the object *book price* signifies different discursive objects, such as the value of the car or at best an industry standard for comparison of prices. According to Sfard (2008, p. 172), abstract discursive objects such as mathematical and statistical concepts are “complex hierarchical systems of partially exchangeable symbolic artefacts.” The conflation of *mean* with *average* and *average* with various concepts of *middle*, reported in Chapter 2, might be explained as obscured relations between partially exchangeable symbolic artefacts.

Concepts

Sfard draws on Vygotsky’s notion that the word *is* the concept and Wittgenstein’s notion that word meanings are evident from their use, to define *concept* in commognitive terms: “*Conceptual commognition*” is defined as “what we encounter whenever a commognitive actor re-acts with *the same word or symbol* to *an entire class of phenomena*” (Sfard, 2008, p. 112). Hence, a statistical concept (like distribution, mean and standard deviation) is a discursive object, defined by Sfard (2008, p. 296) as a “word or other signifier together with its discursive use.” Statistical concepts are not created anew by school-level learners. Statistical words and other statistical signifiers with their uses already exist in the formal discourse which is endorsed by the scientific statistics community. In addition, many statistical concepts are part of colloquial discourse in communities where statistics are widely reported (Nisbett, et al., 1993). Hence, conceptual commognition in statistics implies that colloquial, informal statistical words and other informal signifiers must be used in ways endorsed by the statistical community or be replaced by formal statistical words. Formal statistical symbols in turn, must be underpinned by their appropriate uses in colloquial discourse. Sfard explains how the processes of discourse development preserve relations between new and subsumed components:

> Note that all three constructions that create a new object S – saming, encapsulating and reifying – turn the component d-objects into realizations[30] of S. Indeed, according to the definitions of these three constructions, whatever endorsed narrative is now created on S, this narrative is a translation of a narrative on its component subject… The

[30] A formal definition of realisations follows in the next section. At this stage, realisation should be understood to refer to its literal sense, that of making real or observable.

discourse on S is thus isomorphic to certain closed subdiscourses about component objects (Sfard, 2008, p. 172).

Hence, the shift from informal statistical discourse (as a sub-discourse of informal discourse about component objects) to formal statistical discourse requires translation of narratives that structures isomorphism between the two discourses. What is endorsed in formal statistical discourse must be endorsed in informal statistical discourse. Reflecting on my literature review, I can now frame informal statistical reasoning as a discourse about collections of objects. However, I have shown in Chapters 2 and 3 that not all narratives about collections of objects belong to informal statistical discourse, and therefore that informal statistical discourse may require a translation from colloquial discourse on collections of objects. This translation from colloquial discourse to informal statistical discourse will be evident when the narratives take into account the need for large numbers of observations, suitable reference classes, and sources of variability; even without explicit data at hand. A further translation toward formal statistical reasoning would be evident when the narratives are about new abstract discursive objects, such as measurements of attributes of the concrete discursive objects. Formal statistical discourse will be evident from narratives about, yet again, new abstract discursive objects like distributions, measures of centre, and measures of dispersion. As the discourse shifts, the earlier objects become realisations of the new objects, so that formal talk about distribution is realised by talk about distributions of measurements and in turn, talk about distributions of measurements is realised by the variable objects in suitable reference group from which the measurements were collected. Sfard stresses that discourses do not shift uni-directionally, and that individualisation of a new discourse (learning) involves processes “grounded mainly in attempts to realize new signifiers to which one is exposed, while participating in the discourse with more experienced interlocuters”, thus “attempt(ing) to connect a new signifier to familiar objects” (Sfard, 2008, p. 191). I will now explain the role of realisations in commognition.

Realisations

Realisations are important when teaching and learning are understood as processes of communication aimed at shifts from informal to formal scientific discourses. Teachers and learners involved in a discourse have access only to each other’s realisations of the objects or concepts of the discourse. Realisations are also used purposefully as discourse mediators. Sfard defined realisations as follows:

> *Realization of the signifier S* is a perceptually accessible thing S' so that every endorsed narrative about S can be translated according to well-defined rules into an endorsed narrative about S' (Sfard, 2008, p. 154).

A signifier can have many realisations that together say more about the signifier than any could say on its own. Sfard proposes that the concept is the complete organisation of its realisations:

> *The (discursive) object signified by S* (or simply *object* S) in a given discourse on S is the realization tree[31] of S within this discourse (Sfard, 2008, p. 166).

Realisations in mathematical discourse can be visual or vocal and serve to mediate the discourse. Visual realisations can be verbal (written words or algebraic symbols), iconic (graphs), concrete (e.g. manipulatives) or gestural. Visual realisation in particular is important, since visually accessible objects can be interrogated more easily than purely vocal realisations; and hence can be endorsed more immediately. An example of the productive role of visual realisations to mediate mathematical discourse is the use of graphs to construct narratives about the plausibility of the solution of an algebraic equation. An example from statistical discourse would be the use of a graph to decide if the mean or median is a suitable measure of central tendency for a distribution, or if a linear regression is a suitable model for the relationship between two variables.

Visual mediators are not always concrete things that we can physically identify. Visual images (visual imaginations) which are evoked by signifiers and then described verbally, vocally, iconically, or gesturally, are often the primary or leading realisations that mediate informal discourses. According to Sfard (2008) visual mediators in colloquial mathematical can also be imaginations of concrete objects. Imaginations of concrete objects have the advantage of enabling hypothetical reasoning, viz. reasoning about objects that are not perceptually available during the reasoning process. Formal mathematical discourse is visually mediated by symbols or icons created specifically for the purpose of communication

[31] Sfard (2008) provides tree-like diagrams of relationships between realisations. She described realisation trees as hierarchical structures, but in personal communication (October 2012) indicated that realisations are not necessarily hierarchical, and more often than not, constitute a map of relations.

in the discourse. Once again, these mediators can be imagined and we can act on imaginary objects and 'graph in the head' while we talk about the act. It is here that the notion of a situation model and an immersed experiencer is applicable to statistics as a discourse. In Chapter 3 I argued that the data-context serves as a text that evokes complex images, which hold more information about the context than is provided explicitly. I related the already personal nature of situation models to research about the almost complete immersion of the interlocutor (Zwaan, 2004; Zwaan & Kaschak, 2009). I now argue further that situation models and immersion in the data-context provide access to realisations of a visual and enacted nature, which mediate discourse at the start of a statistical investigation. I will return to the role of realisations in discourse in the next chapter, as realisations are the primary data of a commognitive researcher.

Commognitive theory holds that in addition to their objects, we identify discourses by their patterned ways of doing, or their routines, to which I will now turn.

4.2.3 Discursive routines

Discourses are conducted according to routines, or identifiable discursive patterns, that allow discursants to know immediately that they are involved in a statistical discourse and not, for example, a grammatical discourse. The identification and recycling of routines enables people to act appropriately to a discourse in new situations. Routines are defined by a set of meta-rules that constrain the applicability conditions (when the routine is appropriate); the procedures (how a routine works); and the closing conditions (when the goal of the routine is reached) (Sfard, 2008, p. 221). These meta-rules are not cast in stone, but they are deeply ingrained conventions that enable and constrain routines.[32] Identification of the routines used in mathematisation illuminates the reasons and goals of the discursants. Paraphrasing Sfard (2008), routines hold information about the reasons behind why we mathematise. Sfard argues that while the discursive actions of two people may follow identical procedures, (the how of a routine), they may be following different routines, "set apart by their applicability

[32] An example of a meta-rule that produces a familiar classroom routine is the meta-rule "when the teacher asks another learner the same question, it means my answer is wrong", or, the meta-rule "if given an algebraic equation, I must solve for *x*" which prevented fourth year students in a colleague's class from considering the truth value of an equation of the form $\sqrt{expression\ containing\ x} = -(another\ expression\ containing\ x)$ before plunging into an extended solving procedure.

and closing conditions" (p. 221). Hence, the same procedures are followed with different goals and in differently construed situations. Sfard theorises about three different discursive routines, namely *deeds*, *explorations* and *rituals*. Through the lens of these routines it becomes clear as to why some learners in mathematics classrooms have no recourse to evaluate their own mathematical productions, while others cannot proceed without explicit and often narrow examples from the teacher.

Deed and Ritual Routines

In general, deeds are aimed at physical changes to reality and rituals are performed to achieve social coherence or solidarity with other discursants, through doing what they require. Deeds and rituals can be performed with symbols as well, and are present in the discourse of mathematics teaching and learning.

Sfard draws on the inherent circularity in the development of discourses to argue that novice discursants initially have no other access to a new discourse than by imitating more knowledgeable discursants, thereby acting out rituals. The problem in teaching and learning is that inappropriate routines are employed that may not be identified by the results of their actions alone. As Sfard (2008, p. 238) says, "A sequence of actions that for one person is an implementation of exploration, for another person may be an implementation of a deed." Deeds are not just physically acted out, but can also be acted out symbolically. A deed regularly performed in high school mathematics classes is: *take the number to the other side and change the sign,* which is usually given during *algebraic manipulation*. Hence, the process is manipulation and the closing condition of the manipulation is to achieve a change in the object. It is not unusual for learners to choose to execute rituals in the mathematics class by following a teacher's example, instead of aiming to make sense. Such rituals are extremely rigid and "highly situated and associated with prompts, which are very specific and thus extremely restricting" (Sfard, 2008, p. 243). A statistical ritual reported by Gigerenzer (2004) is the routine and uncritical use of a 5% significance level for rejecting a null hypothesis.

Deeds and rituals are not desired mathematical routines, although they are abundant in mathematics education discourse. The desired mathematical routines are explorations, aimed at creating narratives about relations between mathematical objects. Statistics by its nature is an invitation to explore uncertain situations and create narratives about relationships between variables. Therefore the desired discursive routines are also explorations, which should include narratives about the origin and construction of statistical objects as well as their use to tell stories about everyday contexts.

Exploration routines

Exploration routines can be divided into construction, substantiation and recall subroutines.

Construction subroutines end with new discursive objects. In scholarly mathematics, construction routines build on previously endorsed narratives by means of three meta-discursive rules: deduction, induction and abduction. Only deductively correct narratives are substantiated on the basis of their structure. Narratives constructed through induction and abduction require additional substantiation to become part of the body of endorsed narratives called mathematics.

Substantiation or endorsement of narratives is interpreted differently in discourses. For mathematicians, a narrative is substantiated when it cannot be logically refuted and endorsed when it becomes part of a theory. For people who use mathematics in everyday life, endorsement is achieved when they are satisfied that the narrative reflects the true state of the situation and is safe to use in making decisions, or when the narrative cannot be refuted by empirical observation. Endorsement is thus about the perceived truth of narratives and depends on people's epistemological views. The key difference between mathematics and statistics may be in their substantiation routines. Mathematics is a *truth discourse,* while statistics is a *likelihood discourse*. Contrary to what is the case in mathematics, in statistics substantiation and endorsement is not based on the absence of empirical counter-examples such as refutations. I routinely encounter students in my introductory statistics course who are perplexed to find counter-examples in a data-set when they have postulated a pattern for which the everyday logic is sound. For example, based on experience students will make the following highly plausible statement: as its kilometre reading increases, the price of a car decreases. When they are then confronted with a data table, they expect no counter-examples amongst the data, although they are clear in their everyday reasoning that exceptional circumstances may yield counter-examples. Using tentative language like 'tend to' rather than higher modality statements, is an indication of statistical thinking (Makar, et al., 2011). Substantiation in statistics does not occur on the level of empirical observation, but in comparison to mathematically-derived stochastic models. In descriptive statistics, an empirical observation serves as a counter-example if it deviates further from the norm for the aggregate than can be accepted as standard. Such observations then warrant deeper investigation. Important for statistics teaching and learning, Sfard indicates that in colloquial mathematics discourses, such as we are likely to find in schools, the intradiscursive procedures of abstract mathematical meta-discourse (i.e. the deductive structures that are required for endorsement), are often bypassed in favour of constructing narratives on the basis of what people know from direct, everyday experience (Sfard, 2008, p. 230).[33] Reflecting on the work of Nisbett and his colleagues (1993), as well as Kuhn (1991), which I reviewed in Chapter 3, I expect such bypassing to be even more evident in statistics classrooms, because of the strong context base of statistical questions. Therefore, for the learning of statistics, meaningful integration of intradiscursive procedures with contextual exploration procedures is important.

Recall procedures are used to reconstruct, or summon to the working memory, previously constructed routines. According to Sfard recall procedures reveal much of how a narrative was originally constructed and substantiated by a person. For example, in order to perform a multiplication like 7 x 6, many learners revert to skip counting (repeated addition) as a recall procedure. This reveals a lack of integration with other multiplication facts and procedures they may know, like doubling 7 x 3. The ideal is to construct narratives which integrate and organise the discursive field. An example from statistics would be integration of concepts like the mean, with its full range of realisations, so that calculation is not the only recourse to the mean in exploratory statistics discourse. From the viewpoint of context in statistics, recall of informal concepts like *average*, *normal* or *extreme* in everyday discourse should also be integrated with their statistical realisations.

[33] An example from a Grade 1 classroom I visited can be related as follows: Problem: Four cats, eight mice, how many mice for each cat? Answer from a learner: If you are big, you eat more, so the daddy cat gets 4 mice, the mommy cat gets 2 mice, and the sister and brother gets one mouse each.

4.2.4 Development of discourses

It follows that interlocutors will encounter problems to understand one another when they use different discourses, such as colloquial discourse versus scientific discourse. If the discourses are disjointed, the learning task of shifting toward scientific discourse is particularly hard. For example, mathematical discourse about infinity by interlocutors from communities where the word infinity has no colloquial uses seems superficial, despite close similarity in structure to formal mathematical discourse (Kim, Ferrini-Mundy, & Sfard, 2012). Sfard (2008, p. 172) emphasises the continuity between new signifiers and their realisations: "…whatever endorsed narrative is now created on [the new signifier] S, this narrative is a translation of a narrative on its component subobject."

Commognitive theory holds that learning entails developing a new discourse with appropriate routines, and this typically happens through interaction with someone or some text that is already conversant in the new discourse. Sfard draws on the principle of continuity of discourses to advise as follows:

> The best, perhaps the only, workable way to develop a new discourse is by gradual transformation of a discourse in which the child is already conversant. One way to preserve the discursive continuity is to "grow" new routines in conjunction with familiar deeds that the new routines are supposed to enhance (Sfard, 2008, p. 259).

Sfard warns that it would be a misinterpretation of the principle of continuity of discourses to insist on teaching mathematics from everyday contexts, under the premise that exploration discourses must necessarily be developed from practical deeds. Discourse development is a process of meta-level learning through the evolution of meta-rules, and is often facilitated by commognitive conflict.

4.2.5 Commognitive conflict and ontological collapse

Commognitive conflict spurs on discourse development, whereas ontological collapse prohibits it. Sfard describes commognitive conflict as follows:

> Such conflict appears when one encounters a discourse incommensurable with one's own – when familiar routines are confronted with other

> people's alternative ways of implementing the same discursive tasks, grounded in different metarules" (Sfard, 2008, p. 260).

Cognitive conflict has been employed as a pedagogical tool in various fields of study, including mathematics. For example, when two learners compare their different solutions for a problem, cognitive conflict arises when such different solutions cannot both be true or correct. Hence, the reasoning in one of the solutions must be mathematically incorrect and the learners must engage in an effort to resolve the conflict. Cognitive conflict refers to the (formally endorsed) logic inherent in the mathematics, while commognitive conflict refers to the routines employed by the interlocuters who refer to it. Sfard (2008, p. 161) explains:

> … the encounter between interlocutors who use the same mathematical signifiers (words or written symbols) in different ways or perform the same mathematical tasks according to differing rules – is a common phenomenon. For example, one mathematist may use such words as *number* or *function* in the objectified way, that is, as if it signified another, basically intangible entity; whereas others may treat these signifiers as objects-in-themselves.

When interlocutors experience commognitive conflict, they have to engage in an effort to establish the object of discussion; they have to come to focus on the same 'thing' (Sfard, 2000). Commognitive conflict may also arise if the signifiers belong to different discourses, such as colloquial and informal statistical discourse. In my introductory statistics classes, I have observed the tendency for students to look for substantiation outside statistical discourse in personal empirical observation. The implication is that we use different substantiation rules and thus, we don't talk about the same truth. I have also frequently experienced that my students and I are not talking about the same thing when we talk about a concepts like average, normal, typical or reasonable, which are likely instances of commognitive conflict. Yet, commognitive conflict can be productive: Sfard argues that the possibility to shift from an informal discourse to a formal discourse lies in the endeavour to dissolve commognitive conflict between interlocutors.

Ontological collapse is evident when:

> …we "flatten" the discursive hierarchy so that the consecutive discursive layers become like a series of transparent window panes through which all the objects – discursive (words, expressions) and extradiscursive

> (independently existing material objects) – seem to belong to the same ontological category of "things in the world," with their mutual relations being similarly "objective" and mind-independent (Sfard, 2008, p. 57).

Ontological collapse is a likely result of objectification (the discursive creation of new objects) in the development of formal discourses, and may have negative consequences for learning. The consequences of ontological collapse are threefold: it "(a) may produce *illusory dilemmas*; (b) can result in *phony dichotomies* leading to tautologies disguised as causal explanations; and (c) is likely to lead us to *consequential omissions,* blinding us to potentially significant phenomena that cannot be described in the ontologically flattened terms" (Sfard, 2008, p. 57). Using Sfard's examples for each of these consequences as templates, I reflected on my own experiences as a mathematics educator. Since geometric objects are derived by the saming of material objects in many primary school classrooms, and even in colloquial discourse, they are obvious candidates for ontological collapse. Hence, illusory dilemmas are evident when the material properties of a triangular object are transferred to the mathematical object. First year students in my geometry class were taken aback when I challenged their conflation of the area of a triangle with the physical, touchable plane of the triangular object. This conflation hindered their imagination of the result of continuous change in the size of one angle of a triangle, since pieces are cut off a triangular object when one wants to change an angle. Phony dichotomies, such as refusing to accept a square as a rectangle on the basis of the physical properties of prototypical material objects, are not unusual. Tautologies offered as causal explanations by my students included: 'a square cannot be a rectangle because a rectangle is long.' Lastly, the pervasive phenomenon that young children and school learners do not accept three-sided shapes as triangles if they are judged to be 'too pointy', or 'incorrectly' oriented on the page (Clements & Battista, 1992), can be described as an instance of consequential omission, where differences between members of a category were completely negated in the process of saming. Kuhn's (1991) descriptions of non-evidence and Nisbett's (1993) descriptions of extreme deterministic reasoning may be instances of ontological collapse, since the participants seemed to flatten information within their personal experiential fields and by means of their personal beliefs.

4.3 Summary

In this chapter I presented key aspects relevant to my study of the rich theory of commognition, which theorises thinking as communication. I have interpreted commognitive theory for statistical reasoning by reflecting upon my own experience as a lecturer of mathematics and statistics, as well as upon my literature reviews in chapters two and three. I proposed that statistical reasoning is a likelihood discourse that takes into account the need for: large numbers of observations; variation and sources of variation; as well suitable reference classes. Sfard's theory that thinking and learning are acts of communication allows me to investigate the objects and the routines of informal statistical discourse. In particular, a commognitive perspective problematises the relationship between colloquial discourse and informal statistical discourse. Commognitive theory predicts commognitive conflict between interlocutors for whom the objects of the discourse signifies practical, everyday decision making; and between those for whom statistical discourse is about alienated and abstract discursive objects. Ontological collapse is also likely to occur when colloquial discourse and literate discourse are flattened. Commognitive theory further predicts that although the utterances and procedures of students engaged in statistical discourse may look the same, the students may be performing the procedures with different goals, namely as deeds, explorations or rituals.

In Chapter 1 I was obliged to frame my research questions in language from more familiar, although less operationalised discourses, about reasoning and concept development. I will now reformulate the same questions in the language of commognition. My argumentation up to this point leads to the problem statement: How do teachers learn to reason statistically about authentic data-related contexts in a statistics course that foregrounds discussion? My study is guided by the following main research questions:

a) What are the objects and the narratives of the discourse(s) that are realised during informal reasoning and discussions of statistical objects?
b) What shifts are evident between colloquial discourse and literate statistical discourse?
c) What are constraining and productive narratives in the shift towards statistical discourse?

Commognitive theory provides the following research tools, which will be defined and discussed in the next chapter:

- Building realisation trees of statistical objects
- Disobjectification of literate statistics discourse
- Description of narratives
- Identification of routines (deeds, explorations and rituals)
- Identification of discourse mediators
- Identification of instances of productive commognitive conflict
- Identification of instances of ontological collapse

In the following chapter I will present my research design.

Chapter 5: Research design

5.1 Introduction

This study is located in research about learning statistics, reviewed in Chapters 2 and 3, in classrooms that promote working with data-contexts and that involve learners in data-handling and discussion. Qualitative research methods have been employed extensively in endeavours to describe and exemplify statistical reasoning and thinking. Repetitive features of reasoning about data-contexts and statistical objects are structured into models of statistical reasoning that suggest hierarchical development within different cycles of statistical processes. I described the two most enduring constructs in Chapters 1 and 2. They are the dimensions of statistical inquiry model[34] and the general model of statistical reasoning[35] (Garfield, 2002), based on the Biggs and Collis model (1982) of structures of observed learning outcomes (SOLO).

As indicated in the literature review, teachers tend to reason in similar ways to learners, and there is little in the research findings to guide them to structure data-contexts. The few large scale studies about levels of understanding of statistical concepts (Watson &Moritz, 2000) in school settings confirm the findings of smaller qualitative studies, where very few participants provide responses that can be placed at the higher levels of the SOLO based models; generally described as multi-structural, and integrated or relational understanding.

From a commognitive perspective, these findings indicate a breach between discourses. Participants' talk may, on the one hand, consist of statistical terminology and they may be able to conduct the required procedures, but they do not relate such terminology to the context of the problem, indicating a breach between formal and informal statistical discourse. On the other hand, participants may not proceed beyond colloquial talk about practical and

---

[34] See Figure 1 in Chapter 1.
[35] See Table 3 in Chapter 2.

social issues, indicating a breach between colloquial discourse and informal statistical discourse. With my focus on the nature of discourse, I decided on a qualitative study. A large scale study was not possible, since statistics education had hardly commenced in South Africa in 2007 (Wessels, 2011). By 2007, I had taught statistics courses to undergraduate as well as honours level education students at the University of the Witwatersrand in South Africa, and I had experienced reasoning among my students that I could not simply classify as idiosyncratic. I had to take up such reasoning as legitimate contributions to the classroom discussions, and attempt to find ways to shift, rather than delegitimise the discourse. A qualitative study would allow me to probe deeper into my students' reasoning. Therefore, I decided to frame as a case the group of students who attended the introductory statistics course that I taught in 2008, and treat the discourse of this group as the unit of analysis.

5.1.1 The form, purpose and function of my research

Case study research typically enables the functions of description, comparison and explanation (Plomp, 2006), and can contribute to the advancement of knowledge by comparison of findings to existing models of the phenomenon under investigation. My primary research function (Plomp, 2006) is *to describe* discourses realised in classroom discussions in my introductory statistics course for teachers. The function of my research is motivated by the absence of previous research in statistics education from a commognitive perspective. As a secondary function, I will make conjectures from my findings to explain why idiosyncratic statistical reasoning occurs.

Stake (2003) distinguished between intrinsic and instrumental purposes of case studies. An intrinsic study aims to understand the phenomenon for its own sake, while an instrumental study researches a phenomenon in relation to its use elsewhere. Clearly the distinction between intrinsic and instrumental case studies is useful for focusing a study, but I cannot imagine research in education as narrowly restricted to intrinsic purposes. Furthermore, according to Sfard, research is never completely intrinsic, but "discourse produced with the intention of creating endorsed narratives with which we can mediate and enhance our deeds" (2008, p. 301). I made the choice between intrinsic and instrumental purposes only after the first viewing of the videos of the course. My attention was piqued by how difficult it seemed in retrospect to persuade the students to background anecdotes and opinions and use the data at hand to answer the question. I realised that Pfannkuch's (2011, p. 44) recommendation that participants should be encouraged to "periodically dissociate from the data-context" is in vain if I did not understand why their discourses are confined to their experiences. I would not be in a position to study the implications for teaching before I understood the *development of statistical discourse* as a phenomenon. Therefore, although rich data such as video material and written contributions that cover a complete course undoubtedly will yield data for an instrumental study, the scope of the current study prevented me from analysing my data in terms of teaching moves or quality of teaching.

As indicated in the preceding chapter, this study will provide deeper insight into discourse described as idiosyncratic or verbal statistical reasoning by existing models. However, Stake (2003, p. 149) warns that comparison may compete with "thick description" of the case. I understand this as a warning not to restrict the analysis of the phenomenon to providing support or refutation for existing models, but to be sensitive to the wealth of information in a complex endeavour such as a classroom discussion.

5.1.2 In search of a research framework

A good research framework provides the basis for conceptualizing and designing research studies and gives meaning to data. However good frameworks are not a given in the research of complex phenomena (Lester, 2005). I experienced a long and intense iterative process in search of a suitable framework. Having decided to focus on the development of statistical reasoning, I had the choice of making use of the frameworks based on the SOLO[36] taxonomy used extensively in statistics education research. However, I judged that using such a framework would allow me only to describe endpoints in the learning process, and would require me to focus on individual reasoning rather than collective reasoning. Moreover, taxonomy based frameworks would not enable me to acknowledge my complicity in the development of statistical reasoning in the course I taught and researched. I decided early on in my research that I wanted to focus on the dynamics of the discursive development process. Looking for a research framework that would lend structure without obscuring the messy process of learning through discussion and argumentation led me initially to

---

[36] Structure of Observed Learning Outcomes (Biggs & Collis, 1982).

Toulmin’s (1958) structure of argumentation, which is used extensively in science education research (Erduran & Pilar Jiménez-Aleixandre, 2008). Although useful to structure components of arguments, this framework did not help me to understand why certain discursive patterns seem to develop in the classroom discourse or how such patterns shift. Toulmin’s structure also became unwieldy and restrictive in terms of how parts of arguments are taken up or abandoned as the group discussion unfolded. The next obvious option to consider for analysis of classroom discussions was the wider field of discourse analysis. My focus on the role of discourse in concept development and reasoning, rather than linguistic, political or critical dimensions of discourse analysis, led me to Sfard’s (2008) theory of thinking as communication. Sfard’s theory is based on socio-constructivist and participationist perspectives, most cogently Vygotsky’s tenet that words are concepts that are appropriated through social interaction. Although Sfard’s theory is elucidated by means of reference to interaction between a few interlocuters at a time (e.g. a mother with two girls, or two learners discussing a task), it poses no restrictions to discursive situations and holds potential for the analysis of even less structured discussions than those in her examples.

Sfard argues that finding answers to the questions of *why* and *on what grounds* discourse participants act as they do, requires a fundamental change in researchers’ thinking about thought in the domain of their research. She shows that commognitive analysis yields answers to quandaries of mathematical thinking that cannot be explained from a Piagetian perspective on what learners can or cannot do, since “for the Piagetian investigator, the conversation that preceded [the mathematical action] would be dismissable as mere noise” (Sfard, 2008, p. 9). While in my view it is unavoidable for a researcher to regard some information as “noise”, I agree with Sfard that mathematical or statistical understanding does not reside exclusively in overt mathematical actions. Learning as shift in discourse is a complex phenomenon that cannot be reduced to any of its components, such as the use of specific learning materials, or the role of language or power relationships between discursants. Yet, I am aware that the data I analysed can be analysed with a different purpose. For example, I have included my own contributions to the discussions in an identifiable manner, but in my analysis I have backgrounded issues of power relations between students and lecturer.

5.1.3 Commognitive tools for analysis

In order to analyse statistical reasoning as a commognitive process, the discourse of the participants in my study is the unit of my analysis (Sfard, 2012). I identified the following four analytical tools:

Mapping realisations in search of discursive objects

The discursive object (e.g. function or mean) is the realisation tree of this object within a discourse. It implies that the discursive object exists among its realisations rather than in any single realisation. Hence, the realisation tree constitutes what the discursants are talking about. Sfard organises the mathematical realisations of discursants into realisation trees: "hierarchically organized set of all the realizations of the given signifier, together with the realizations of these realizations…, and so forth" (Sfard, 2008, p. 301). A realisation is at the same time a signifier, and hence a realisation tree illustrates "the downward movement from a new signifier to its realizations in the existing object" as they emerge from the collective of discursants (Sfard, 2008, p. 192). Realisations are personal constructs (they belong to the discursant who makes them), and are likely to be highly situated, but they are interpreted and organised by the researcher. The richness of a participants' realisation tree for a given signifier (e.g. function or mean), and the stability of the tree in different problem-situations provide information about the nature and quality of a participants' discourse. Yet, it is not possible to specify all the elements of a person's realisation tree as a research task. Realisations are evoked by tasks or other discursive actions and these are never exhaustive. Realisation trees can be transversed upward (from existing objects to new ones) as well as downward (from new signifiers to their realizations in existing objects) (Sfard, 2008, p. 192).

Describing narratives to identify object-level rules

Narratives are "series of utterances, spoken or written, that are framed as descriptions of objects, of relations between objects, or processes with or by objects, and are subject to endorsement or rejection, that is to be labelled as "true" or "false" (Sfard, 2008, p. 300). Specifically, object-level rules reside in narratives on the objects of the discourse. Colloquial discourse and informal statistical discourse can be distinguished by their narratives on objects and relationships between objects in relation to variation and reference classes.

Identifying discursive routines in search of purposes of communication

Routines are "the all-important regularities to be found in any discourse" (Sfard, 2008, p. 195). In Chapter 4 I explained the different types of routines, namely deeds, explorations and rituals. According to Sfard (2008, p. 208) "one goal of the commognitive researcher is to make routine-defining meta-rules explicit." These meta-rules define the 'how' and 'why' of routines, and are constructs of the researcher. The researcher aims to identify deviations from the common uses of routines. Misconceptions can be recast as commognitive routines that are used in inappropriate situations. Routines are defined by their purposes, with discursive deeds aimed at changes in discursive objects; rituals aimed at social acceptance; and explorations aimed at creating endorsed narratives about discursive objects. Routines constrain discourses and are indispensible in the development of a discourse.

Identifying word uses in search of objectification and individualisation

Considering that in commognition a concept is operationalised as a word and its uses, the way discursants use words tells the story of what and how they understand the discursive object. Objectification and reification is evident when words are used as nouns or in noun phrases, as opposed to verbs or in verb phrases that include human agents. Sfard describes a four-stage model that charts the development of word use during learning. When a discursant does not know what a new word means, the word is initially used passively, as the discursant reacts in routine ways to the word as uttered by other discursants, without using the word in her own utterances. As an example I will use personal observations of learners' uses of the word *gradient* in mathematics classes. Passive use is evident when learners do not use the word gradient in explanations, but rather gesture slope or refer to 'm', which is the conventional symbol used for the gradient of a linear function. Gradually the word is used actively, but restricted to familiar routines (such as when learners talk about gradient only as the value of 'm' the coefficient of *x*). Then the word use becomes phrase-driven and the entire phrase serves to constitute the discursive object (for example, 'the gradient is when you divide five [vertical difference between two points on a graph] by four' [horizontal movement between the same two points on a graph]). Finally, when the word use is object driven, the word alone serves as signifier of its realisation tree, and we can deduce that the discursant has individualized the new word. Object-driven word use is alienated, as in this example of the use of the word gradient: 'the gradient is the ratio of vertical change and horizontal change between two points on a graph or in a table of values.'

5.2 Empirical setting

The Marang Centre at Wits School of Education[37] offers a B.Sc Honours degree in Mathematics and Science Education. Part time students complete the course in two years, during which time they attend two contact sessions per week, each three hours long, from 4pm to 7pm. In the first year of study, they complete courses in Mathematical Reasoning and Algebra. In the second year they complete a Geometry course and an Introductory Statistics course. The students also complete a Research Report in their second year. Although the degree can also be completed fulltime in one year, the majority of students are in-service Mathematics teachers and study part-time. In order to be eligible for the Honours programme, students must achieve a 65% minimum mark in Mathematics in the second year of an undergraduate degree, or in the fourth year of an Education diploma. The students who enrol for the Honour's degree have varied backgrounds that mirror the socio-economic disparities and historical iniquities of South Africa. The programme is offered in English, which is a second language for the majority of the students, although they do teach Mathematics in English. My research took place as I taught a semester-long introductory Statistics course to the in-service teachers who enrolled for the Honours degree in Mathematics Education during 2007. English is also my second language. Nine of the thirteen students that attended my statistics course took the degree part-time and taught Mathematics during the day. All students had to attend one three hour session per week for eleven weeks, which is a total of 33 hours of contact time. In addition, they were assigned preparation and homework exercises that would take up about another ten hours per week.

5.2.1 Social and socio-political issues

Since my research focus would be on students' reasoning in a social setting, the situation was complex. Firstly, language was foregrounded in all aspects of the research. My role as participant in the classroom discussions included rephrasing both my own and the students' awkward utterances when I noticed in the moment that words might not be interpreted in the standard ways of first language speakers. Secondly, the course itself was language rich, since statistical findings tell a story, and must be reported in language and numbers. Thirdly,

[37] The School of Education is in the Faculty of Humanities of the University of the Witwatersrand in Johannesburg, South Africa.

teaching statistics places language demands on lecturers and teachers who want to engage their learners in discussions of data and statistical findings, and the student participants are encouraged to become such teachers. Finally, when interlocutors are not participating in their home languages, the researcher must make even more effort to alternate between insider and outsider perspectives on her own word use.

Although it would not become a focus of my study, race and power issues were likely to influence communication in the classroom. My experience was that among the students, power to participate in the discussions lay mostly in the ability to verbalise ideas in English. It seemed plausible that students with previous participation in statistical discourse through other courses or through teaching statistics, would be in a better position to verbalise their statistical ideas along with first language speakers, who simply find the right words faster. In addition, issues of power in the wider social context of South Africa also include race, gender and seniority, and were likely to influence participation in the course. I therefore had to pay careful attention to the grouping of the students. Forming groups along ethnic or language lines would have sent a message of racial discrimination. Basing groups on previous experience of statistics might have silenced whole groups with less or no experience during classroom discussions. In groups with pronounced age gaps between members, commognitive conflict was likely to be suppressed through fear of disrespect to elders. I had to coordinate all these issues in order to make small groups where all participants would be most likely to engage. I could avoid the problem by allowing the students to form their own groups, or reconstitute the groups throughout the course. However, my experience of teaching heterogeneous classes is that students tend to group themselves along ethnic and language lines, with the result that second language speakers have long deliberations among themselves in their main languages, but contribute less to classroom discussions. I judged that reconstitution of the small groups would not be wise either, since participation in discussions requires trust between interlocutors. I would take the responsibility as lecturer to negotiate a safe environment for classroom discussions, but in the small groups, trust develops through the interaction between the group members as they take up the negotiated norms. Therefore, I reasoned that the groups should remain stable throughout the course.

The principle that finally settled my selection of groups, was my belief that research settings should reflect rather than avoid the complexity of our sites of practice as teachers. In South

Africa, first language English speaking teachers have to learn to accommodate the realization of ideas and concepts in a way that allow access to non-English speaking interlocuters. Similarly, if they consent to English as language of learning and teaching, non-English speaking teachers have to learn to realise in a way that allows access to English speaking interlocuters. Hence, I decided to control age first, and then gender, and the group members would have to find their ways to communicate despite language and racial differences. Each group had discursants with previous experience of teaching Statistics, as well as participants with experience on previous statistics courses, and novices. This was acceptable from a participationist epistemology perspective, as I would not be the only source of endorsement for those without previous experience. However, I anticipated that I would have to strongly negotiate the norms of discussion to support participation of the novices.

5.2.2 Constitution of small groups

The thirteen students were grouped as follows:

Group A consisted of four mature students who were experienced teachers, one white female, one black female and two black male students. All had taught descriptive statistics before at high school level, but only two students had done a course in statistics as part of their degrees some years ago. Except for KH, all students in this group had home languages other than English.

> **KH:** Female. She completed Statistics 1 as part of B.Sc degree 30 years ago, and taught statistics at Grade 10 level in an Advanced Mathematics programme.
>
> **GK:** Female. She had not attended any previous statistics courses, and taught Statistics to Grades 8, 9 and 10.
>
> **RK:** Male. He completed two undergraduate statistics courses as part of a B.Sc degree in the past, and had experience of teaching Statistics, but was not teaching at the time of the course.
>
> **SM:** Male. He had not done any previous statistics courses, and taught Statistics at Grade 10 level.

Group B consisted of four young students, two White females, one Black female and one Black male student. The Black male and female students are young teachers (with about three years of experience). These students had limited experience of teaching statistics at high school level. Except GG, all students in this group had home languages other than English.

SDS: Female. She completed an introductory statistics course at Wits School of Education in 2006 as part of her undergraduate studies, and attended the honours course as a full time student. She had no experience of teaching Statistics.

NM: Female. She completed an introductory statistics course at Wits School of Education as part of her undergraduate studies in 2004 and taught Statistics at FET level.

GG: Female. She had not done any previous statistics courses, was a full time student, and had no experience of teaching Statistics.

MM: Male. He completed an introductory statistics course at Wits School of Education as part of his undergraduate studies in 2004 and taught Statistics at FET level.

Group C consisted of two White female students (RC and DH), two Black female students (CZ and S) and one Black male student (LT). All were mature and experienced teachers, except for one White female student who had not taught yet. Both White female students had taken previous courses in statistics as part of their undergraduate degrees. In particular, one of them did courses in inferential statistics during a career in the business sector, and was teaching Inferential Statistics in an advanced placement curriculum. I was of the opinion that this group was least balanced of all in terms of language and power relations, and decided not to video-tape them during small group interactions. I collected their written data along with the rest of the class and videotaped their contributions to whole class discussions. LT's attendance and work submission was erratic and incomplete and I decided to eliminate his written work from the data.

5.2.3 The video recording set-up

I had two video cameras set up to record the small-group interactions of Groups A and B. The cameras were positioned close enough to the groups and high enough to capture writing and book work. The lenses of the cameras were focused on the students' hands and tables, rather than on their faces. I kept track of where particular students sat around the tables on a seating plan, but the students generally sat in the same positions relative to each other. The second camera, which was focused on Group B during group discussions, was also used to pan to the front of the class and zoom in on the whiteboard during class discussions in order to capture writing on the board. On the tables of groups A and B were boundary microphones, which were sensitive enough to capture whispering. During whole class discussions, the table microphone was removed from the second camera, in order to capture sound from the whole room with the built in microphone of the camera. The first camera with its boundary microphone remained focused on Group A. Figure 6 shows a top view of the classroom set-up and the positioning of the students. I used a data projector to project graphs onto the whiteboard so that we could draw on the graphs during discussions. The data projector and my laptop were positioned on the table.

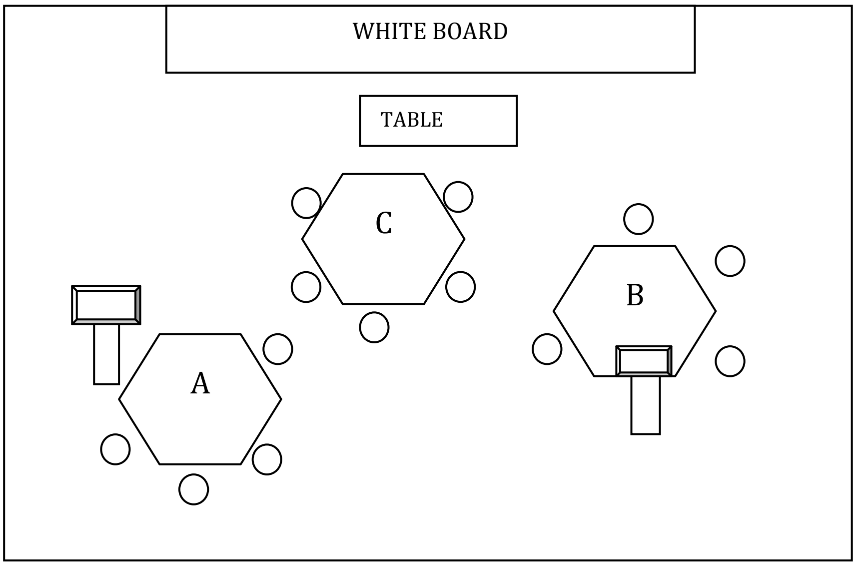


Figure 6: Classroom set up during data capturing.

The cameras were focused at the start of each session, a sound check was done and when the filming started the cameras were synchronised with a loud clap. During the session the cameras were unmanned, and the videotapes were exchanged during the coffee break ninety minutes into the session. The synchronisation enabled me to line up time codes between the two cameras, which in turn allowed me to switch between the discussions of groups A and B at specific points during data analysis.

5.3 My role as lecturer and researcher

Ball argued that the teacher-researcher “must be able to view the teaching, the students, and the learning in the context of, but also apart from, their own efforts and desires” (Ball, 2000,

p. 392). This is made possible by the use of a strong theoretical lens and collaboration with colleagues, who are able to challenge assumptions and propose alternative interpretations. Indeed, I found it a challenging task to achieve this distance in my research. Theoretical lenses that required me to focus on individual contributions in classroom discussions required my own contributions to be foregrounded and analysed. For example, when I tried out Toulmin's argument structure as theoretical framework, I was confronted with questions about my motivations to contribute in certain ways at specific moments. In addition, I judged the contributions of the group as a result of my own contributions, and this lead to more hypothetical questions: would they have responded differently if I had contributed differently? Although such a study may have merit, my research focus was the development of discourse in the authentic setting of the course I taught, rather than the quality of my own teaching. Commognition as a theoretical lens allowed me to include my own utterances in the discussion as data in the same manner as the utterances of my students.

Silverman (2006) contrasts natural talk that derives from studying video and audiotapes or written texts, with researcher-provoked data obtained in interviews and focus group discussions, but Peräkylä and Ruusuvuoru (2011) argue that differences between natural talk and researcher-provoked data form a continuum, rather than a dichotomy. My contributions in class were that of a lecturer-participant rather than as a researcher and I did not conduct interviews with the students during the sessions. So, as far as learning and teaching is a natural endeavour in our society, I have categorised my data as closer to natural talk than to researcher-provoked talk. Hence, while I fully acknowledge my complicity in the discussions, my data is "[video records of and] transcriptions of actual occurrences in their actual sequence" (Sacks, 1984, p. 525).

My dual role also had implications for the practical decisions a lecturer has to make in the moment. My responsibility towards the students in my class ethically obliged me to choose teaching above research if a conflict of interest should arise during this research. During the course, I experienced conflict between my desire to withdraw from discussions in order to observe and my obligation to participate as I believed I should in my position as lecturer. I chose to participate each time. Similarly, there were instances when, as a researcher, I would have liked to probe deeper when students tended to keep their conversations private, but in these instances I chose to withdraw.

5.4 Ethical concerns

This study was done in compliance with the ethical code of the University of the Witwatersrand. Ethics clearance was granted under the protocol number 2008ECE57. As required by the ethical processes at Wits, each student received a written information letter in which I informed them of my desire to research the course. I explained in writing that I was interested in capturing their discussions and using their written work to learn more about the development of statistical reasoning. They were assured in writing that they could choose in which aspects of the research, if any, they wanted to participate; and that they could withdraw their consent at any time. If they were not willing to be videotaped, I would set up the cameras in such a way that their visages were avoided. They could refuse to be videotaped and still give consent for the audio track including their contributions to be used; and they could choose whether or not to make their written work available. The information letter was handed out at the end of the previous semester, to allow them time to contemplate their decisions. I did not set up the video cameras in the first session, in order to avoid coercion.

Ongoing ethical concerns were inevitable, due to using a degree course that I taught as research site. I was aware that despite giving written consent, the students might experience conflict between their roles as research participants and as students. I therefore invited a colleague to the class during the first session, who agreed to act as advisor to the students for the duration of the course. He guided them to appoint a class representative that they trusted, to convey matters to him on their behalf. This was necessary since the part-time students were not on campus regularly enough to allow them to easily consult with an advisor. He also provided all students with his e-mail contact details and his telephone number, and invited them to contact him if they experienced problems during the course. After this discussion, I distributed consent letters, which I asked the students to complete and place in a closed box on leaving. All the students gave written consent to participate fully in the research.

There was no occasion during the course on which they contacted the advisor, but on two occasions individual students took up concerns with me personally. In the second contact session, RC, a student in the group that was not video-taped, felt that I had my back to their group too often. I took up this concern and immediately moved the cameras so as to be able to easily move between the tables, and made a point of moving to the front of the class when addressing the whole class in order to be facing all three groups. The second occasion involved GG, who was concerned about assessment. She felt that the course demanded too little written work, since discussion took up most of the contact time and written work was only given as homework. I then built into the contact time regular opportunities to go through their written solutions to tasks. Her concern prompted me to discuss with the class the main assessment criterion, which was explicitly stated in their course outlines, namely to communicate statistical investigation processes and findings in written and spoken English. We discussed the nature of statistics as providing evidence for or against claims, and as findings in context that must be communicated. The course outline also indicated that their final assessment would take the form of a statistical investigation of a situation of their choice, from which the findings would be presented verbally with the support of power point slides. Based on the assessment criterion, verbal communication in this course held a similar status to written work.

On two occasions, issues about invasion of privacy arose. I noticed in one instance that a student deliberately obscured her written work from the camera, and on another occasion, two students deliberately switched to their vernacular and dropped their voices during group work. I respected their need for privacy, and moved away from their table, refraining from asking them to make their work or discussion available to the video recorder. These were the only two occasions during the course of which I was aware. In general, the students seemed unaware of the cameras, and even jocularly referred to the research component of the course.

Other issues that arose were not directly related to the research aspect of the course. For example, KH discreetly voiced her dissatisfaction with being assigned to her designated group, as she preferred to work with the same student as in previous courses. They had similar educational backgrounds and both taught statistics to high school learners in advanced placement classes. She argued that the group she was assigned to did not contribute to her learning of statistics, since they had less pre-knowledge of statistics than she had. I asked her to focus on listening hermeneutically[38] to the contributions in her group, in order to develop

---

[38] Hermeneutics is the science of interpretation. Davis (1997) described hermeneutic listening as listening that negotiates understanding between discursants, and is attentive to the historical and social underpinnings of utterances. Gadamer used hermeneutics to characterise conversation as a process of coming to an understanding: "Thus it belongs to every true conversation that each person opens himself (sic) to the other, truly accepts his

understanding of difficulties in learning statistics. I usually control group formation in my courses in order to deliberately confront teachers as students with thinking and reasoning at different levels than their own, and would not change this practice in this case as in others.

5.5 The design component of the course

When thinking and learning are perceived as communication, it follows that for learning and thinking to occur, there must be something worthwhile to communicate about. Choosing problems for discussion was thus an important design task. I chose ill-structured tasks to introduce discussions, with the purpose of eliciting meta-level discussions about the data-contexts and the statistical objects that were studied. Ill-structured problems were followed up with well-structured problems that allowed for object-level learning of statistical procedures. I decided on this pedagogy because the participants knew many of the procedures, such as how to calculate the mean or how to determine the median of a data set. Shin, Jonassen and McGee (2003) drew attention to the difference between solving well-structured and ill-structured problems. Where well-structured problems (for example: 'find the mean of a given data set') have well defined solution procedures and correct, convergent answers, ill-structured problems are characterized by uncertainty about the appropriate concepts and rules needed to solve the problem. The importance of discussion and argumentation in solving ill-structured problems becomes apparent when we consider the processes used to solve them. Processes include: "(a) articulating problem space and contextual constraints; (b) identifying and clarifying alternative opinions, positions and perspectives of stakeholders; (c) generating possible problem solutions; (d) assessing the viability of alternative solutions by constructing arguments and articulating personal beliefs; (e) monitoring the problem space and solution options; (f) implementing and monitoring the solution; and (g) adapting the solution" (Shin, et al., 2003, p. 8). These processes are essentially the same as those required in the PPDAC cycle of statistical investigation.

Authentic statistical inquiry is by nature messy (Moore, 1998) and concerned with ill-structured problems. Shin et al. (2003, p. 27) further indicate preference for the use of ill-structured problems in *unfamiliar* contexts for the purpose of learning, and emphasised the

---

point of view as valid, and transposes himself into the other to such an extent that he understands not the particular individual but what he says" (Gadamer, 1993, p. 385).

imperative to maintain the structural complexity of problems. It is the complexity that necessitates that students become aware of their thinking and reflect on their problem-solving processes. Based on these principles, I used ill-structured problems as the basis of discussions in order to stimulate the development of statistical reasoning. I chose contexts that might be unfamiliar to the students personally, such as the used car market and minimum wages, but that are likely to be discussed informally in social contexts. This decision was not made to advance the research component of the course, rather it was made in the spirit of statistics education reform, and reflects the way that I usually teach statistics.

In the South African context, where descriptive statistics makes up only 5% of the mathematics curriculum (effectively one 45 minute period every third week) and contributes merely 20 marks out of 300 (6,7%) to the final Mathematics mark at FET level (Department of Basic Education, 2010), my aim was to expose teachers as students to the nature and role of descriptive statistics in reasoning and problem solving, rather than to cover an extensive range of topics. My selection of content was driven by two key issues: the need to engage with the initial messiness of data contexts and the need for conceptual understanding of the statistical tools to be used. In the table below I broadly outline the course content, focused on descriptive analysis of single variable distributions and co-variation of two variables in a distribution.

Table 4: A summary of the content of the researched course.

Session	Focal situation/context/problem	Statistical focus
1	**Introduction to the course:** Information about the research and consent.	Reflecting on our views of the nature of statistics; difference between statistics and mathematics.
2	**Introducing an ill-structured problem:** What is a reasonable price for a used car? **Working with data about prices for used cars**: Categorisation of cases in order to compare prices meaningfully.	What are statistical questions? What is data? Operationalise vague terms "reasonable" and "used car"; identifying sources of variability; constructing reference classes. **Statistical vocabulary**: Cases and variables; shapes and summaries of distributions of a single variable. **Talking statistics**: How do we use statistical summaries as parameters for describing findings as typical, usual, unusual, very unusual. Comparison of statistical summaries with shape properties of distributions.
3	**Working with data about prices for used cars:** Introducing FATHOM. Finding data on the internet and importing data to FATHOM; graphs and statistical summaries on FATHOM. How to report statistical summaries in contexts: Writing up the solution to "What is a reasonable price for a used car?"	Comparing distributions represented in box and whisker plots, histograms; cumulative relative frequency plots Standard deviation; standardising data; using z-scores and p-values to compare cases relevant to different distributions.
4	**Conceptual understanding of the mean:** compared to procedural understanding. Everyday use of *average* versus statistical meaning of *mean.*	How does the mean come to represent a data set? We know the procedure to calculate the mean, but what is the object definition of the mean?
5	**Introducing an ill-structured problem.** Using data about load weight, distance travelled and fuel used, find a way to estimate transportation costs. Reflecting on mathematical reasoning about linear functions as pre-knowledge for statistical regression.	**Co-variation and scatter plots.** From a description of a context with numerical information to reasoning about co-variation. **Statistical vocabulary.** Bi-variate distribution, strength and correlation; least squares regression and trend; gradient and co-variation; predicted values and residuals. **Talking statistics.** Interpreting the gradient of the least squares regression line in context. How would typical learners reason from a mathematical point of view, and how can we bridge their reasoning to consideration of co-variation and regression; Average as a metaphor for the least squares regression line.

6	**Understanding the mathematical properties of the statistical objects:** Analysing the formulas of the correlation coefficient and the regression gradient; the logic behind r^2 and residual graphs.	**Correlation**; **residuals as unexplained variation.** The Anscombe data set: limitations of regression and the correlation coefficient. Residual graphs to verify if linear regression is an appropriate model; coefficient of prediction to verify if the chosen independent variable is a good predictor variable.
7	**Introducing an ill structured problem.** An argument in the media: South Africa's minimum wage is a scandal.	The difference between opinion and fact; Identifying the causal argument of the author; generating alternative causal lines: What would we take as evidence that the author is correct? Not correct? Finding data, building a statistical argument.
8	Hand in of 'analysis of wages' problem. A sampling experiment using three kinds of beans.	Sampling methodology.
	Study Break	
9	**Reading:** The mean as a signal in a noisy process (Konold & Pollatzek, 2004) Reflection on own understanding of mean, problem types to develop understanding of mean	Concept of error as deviation from the mean.
10	**Questionnaire design:** uses of cell phones among school children	Different questions yield different types of data; bias.
11	Discussion of final research projects Various exercises, revision.	
12	Presentation of reports for assessment.	

5.6 Sources of data

I used the video data that I collected during the course as my primary data. In addition, I collected copies of the written work of the students, including their final research projects, and I conducted a semi-structured interview with each student after the fourth session. The students' written work mostly comprised standard textbook exercises and did not yield additional data to support my research focus, namely their discourse. The video data was so extensive that the scope of my study did not even allow me to analyse half of the sessions. During the interviews, I asked the students questions to probe their experience of the course and asked how the course helps them to think about teaching statistics. Again, the scope of my study did not allow me to analyse their discourse in relation to my teaching and the interview data may yet be analysed in a different study. Based on Sfard's approach to the use of her own theory, I defined as the unit of my analysis the discourse during classroom discussions. Therefore I legitimately backgrounded other sources of data for this study.

I decided to videotape the course from beginning to end in order to have access to the complete flow of the course, including group discussions, talk between students as they worked individually and classroom discussions. Videotaping the sessions was also important in order to capture the interlocutors' gestures along with the inscriptions they may make, and refer to during discussions, as these would mediate the discourse (Sfard, 2008).

As researcher, I could not anticipate in advance when the richest discussions would occur that would enable me to analyse and describe the emergence of statistical discourse. The selection of episodes for analysis was therefore the first research task. Once the course was finished and I commenced with the analysis of the data, I watched all the video data in sequence (66 hours of data, excluding their assessment project presentations). As expected, discussions of ill-structured problems were rich and the students engaged with each other's contributions in addition to responding to my contributions as a lecturer. In contrast, in sessions where the students were engaged with textbook exercises or homework discussions, the interaction was more constrained, as the focus seemed to shift to what would be correct procedures or answers worthy of marks. During the initial viewing process I consulted the literature in my review (Chapters 2 and 3) and made notes and interpretative elaborations (Sfard, 2008), on which to base my selection of sessions to analyse. I decided on four sessions to transcribe for analysis. Two sessions involved students working on an ill-structured, context-based problem, and one session concerned an ill-structured problem about a statistical concept, hence it was a theoretical problem. The fourth session involved a discussion of statistically based claims in a newspaper. I analysed all four sessions, but decided to write up only three, namely the discussion of a reasonable price for a used car in sessions two and three; and the discussion of the meaning of the mean in session four. The discussion in session seven, which I did not include in my study, concerned the claim in a newspaper that South Africa's minimum wage is a scandal. I noticed early on in my analysis of the discussion in session seven that the discursive routines that I had identified in the other discussions were repeated. During the design phase of the course I included the newspaper claim as a problem about the analysis and interpretation of statistical claims. However, it transpired that the students proceeded once again with discussions to structure the data-context, using the same routines as those I identified in the discourse on a reasonable price for a used car. From the perspective of discourse development, and using the analytical tools I described earlier, the session yielded no new information[39] and confirmed my analysis of the two selected discussions. To summarise, the discussions I analysed and reported in this study are:

Sessions 2 and 3: What is a reasonable price for a used car?

This problem is grounded in the investigative cycle of statistical inquiry (Wild & Pfannkuch, 1999), and the statistical reasoning task was the structuring of a data-context in order to investigate the problem statistically. This discussion provides information about reasoning at the beginning of the PPDAC cycle of statistical inquiry as described in Chapter 1.

Session 4: Conceptual understanding of the mean compared to procedural understanding

This problem is concerned with the colloquial use of *average* versus the use of *mean* in statistical discourse. The statistical reasoning task was to construct the mean as a statistical object, as opposed to a statistical procedure.

Aiming for descriptive validity (Maxwell, 1992) I paid a postgraduate education student at the university to do the initial transcriptions of the videotapes of the selected sessions. Independently, I transcribed the same sessions with the intention of comparing our transcripts for accuracy. In comparison, it became clear that she was not suitably acquainted with terminology used in the course to do accurate transcripts, and I transcribed the videotapes of the selected sessions myself, and asked a video producer to compare the transcripts to the video. Although he is not acquainted with statistical terminology, he is experienced in comparing scripts to video footage for editing purposes.

Even with ‘accurate’ transcripts in hand, I kept going back to the video itself during the process of analysis. I could not sense the flow of the discussions from the transcripts. I concede that an experience of flow is likely to be subjective, but I nevertheless see such awareness as contributing to the descriptive validity of the analysis. Practical experience gained from years of teaching enabled me to become aware of the ebb and flow of energy in discussions, which cannot easily be captured by necessarily linear transcripts of the videotaped conversations. However, according to the principles of completeness, contextuality and directness, I provide extended transcripts as evidence for my interpretation

[39] I am not claiming that analysis of the session with different analytical tools and a different research focus would not yield new information. Indeed, the data hold promise for discursive analysis of argumentation skills.

of discussions. Any challenges to their interpretive validity should be addressed by reviewing the video data, rather than solely reading the transcripts.

5.7 Collaboration in my research

Confrey and Lachance (2000) as well as Ball (2000) insist that qualitative research benefits from collaboration. I regularly collaborated with a teacher educator colleague with an interest in statistics education during the course. Chris Human holds a masters degree in mechanical engineering, and designs course materials for teaching statistics to post-matric students who prepare for entrance to engineering courses at university. I was able to call upon Chris to soundboard my ideas during task design. As my analysis progressed, I also presented preliminary findings at conferences in order to engage with peers. Specifically, I presented a paper about the transition from everyday reasoning to statistical reasoning at the Southern African Association for Research in Mathematics, Science and Technology Education Conference in 2009, which culminated in a published paper (Lampen, 2011). I also presented a poster on my work in progress at the 33rd Conference of the International Group for the Psychology of Mathematics Education (Lampen, 2009) and presented an invited paper at the Eighth International Conference on Teaching Statistics (Lampen, 2010). On a local level, I presented my work to fellow students at PhD weekend conferences organised by the Wits School of Education. In addition to comments from a critical respondent who had access to the presentation and supporting data ahead of one of these presentations, lively discussion about my interpretation of data in relation to commognitive theory helped me to clarify my own thoughts.

5.8 Credibility of my research

Qualitative research is inductive. Hence the researcher has to establish confidence in the 'truth' of her findings. Since truth is a discursive object, the findings must be judged against the principles of her epistemology and theoretical context. As a commognitive researcher, I acknowledge my own complicity in the classroom discussions that became the data for my research, as well as the subjectivity of my interpretations. My research is an act of participation in a discursive process that produces "cogent narratives by which other human practices can be mediated, modified and gradually improved …" (Sfard, 2008, p. 35). The narratives I created are intended to mediate the practice of statistical education. How does a commognitive researcher establish confidence in the credibility of her narratives? Maxwell argues that validity in qualitative research refers primarily to the accounts, rather than data or methods (Maxwell, 1992). The practice of statistical education pre-existed my participation and I presented my engagement with existing narratives from statistics education discourse in my literature review. Hence the first source of confidence in my newly-constructed narratives resides in comparison with existing narratives.

If research is a process of participation in discourse, the commognitive researcher must take responsibility for enabling discussion of her research and narratives among members of the research community. Hence, my research is based on the following six principles (Sfard, 2012, p. 8):

a) The *principle of multivocality*: researchers' stories about the world cannot be equated with the world. We are creative storytellers rather than the world's 'ventriloquists'. Hence the researcher talks in the first person and is always in the quest after new, more convincing versions.
b) The *principle of operationality*: if research is the activity of sharing useful narratives about the world, the researcher's talk must be guarded, as much as possible, against misunderstandings. The first condition for communicational effectiveness is disambiguation and operationalization of the researcher's vocabulary.
c) The *principle of completeness*: the entire discourse related to a topic is the unit of analysis, and in data collection, the researcher keeps in mind that all kinds of interaction, even those describe as "non-interventional interviews", are events of learning.
d) *The principle of contextuality*: the researcher has to document human interactions as fully as possible, never considering participationists' utterances out of their conversational context (this practice stands in stark contrast to the tradition of ignoring an interviewer's part of an interview).
e) The *principle of alternating perspectives*: the analyst, considering the possibility of incommensurability between her mathematical discourse and the discourse of the participants in her study, makes a sustained effort to alternate between being an insider and an outsider to her own ways of using words.
f) The *principle of directness*: in their research reports, commognitive analysts begin with things said (and done) by the participant, not with their own renderings of these

data. They also make it clear that their stories are not directly about the world, but about the participants' narratives about the world.

These principles of commognitive research serve to establish trust in the interpretive validity of research narratives.

Since transcripts are also narratives about my research, the conventions of transcription must be clarified to facilitate communication between the researcher and her readers. I used the following conventions in the transcriptions of the video tapes:

a) Utterances in italics were pronounced with more emphasis than the rest,
b) Three dots between utterances indicate hesitation by the discursant,
c) Round brackets indicate gestures and other physical actions,
d) Square brackets contain my clarifying text as researcher. They contain my interpretations of possible missing text or implicit references to surrounding text, and
e) The numbering in the excerpts enables the reader to locate the excerpts in the full transcripts provided in the addenda.

5.9 Summary

In this chapter I described my research design as a case study of the development of statistical discourse with a convenient sample of Honours students who registered for a course that I taught. I presented my methodology as analysis of classroom discourse according to Sfard's notion of commognition, steeped in a participationist epistemology. I indicated that my primary source of data was video footage of the complete course, especially of the discussions of ill-structured problems in two key sessions. I chose the sessions for the richness of the discussions and for their relevance to statistics education literature: the discussion in sessions two and three is grounded in the investigation cycle of statistical inquiry, and the discussion of the meaning of the mean is related to previous research about the colloquial meaning of average and how average is related to the mean as statistical object. I stated upfront the principles of commognitive research as a pledge to enhance the credibility of my research. In the following two chapters, I will present the analysis of each of the two selected discussions. In Chapter 10 I will discuss the implications of my research narratives for the practice of statistical education and research.

Chapter 6: Data analysis: The purposes and properties of contextual questions

6.1 Introduction

In this chapter I present a commognitive analysis of the first task I analysed for this study, solving the ill-structured problem “What is a reasonable price for a used car?” I start by explaining the task demands and the commognitive properties of the learning support materials. Then I provide background about the properties of questions in statistical investigation and proceed to analyse the written questions that students raised in their endeavour to understand the data-context. Finally I provide my analysis of the classroom discussion of the questions. I will answer the following empirical questions, based on the research questions I formulated at the end of Chapter 4. Here I focus on the questions indicative of discourse, as reflected in the following formulation:

a) What are the objects and narratives that are realised when posing informal questions about the data-context?
b) What shifts are evident from the posed questions between colloquial discourse and literate statistical discourse?
c) What are constraining and productive narratives in the shift towards statistical discourse?

Everyday intuitive reasoning about the variation in the prices of used cars extended over two sessions in the course. The learning task I set for the classroom discussions in these two sessions was aimed at “the deconstruction, negotiation and refining of the problem in conjunction with context familiarisation” (Fielding-Wells, 2010, p. 2) at the start of the investigative cycle of empirical inquiry. As indicated in the PPDAC model, this phase requires grasping system dynamics and defining the problem (Wild & Pfannkuch, 1999). I gleaned from examples of idiosyncratic responses in the literature, that the context of used-car prices was not likely to be understood as a statistical context at the start of the investigation cycle. I wanted to evoke the context as fully as possible in the classroom situation, so that the structuring of the context would be a meaningful learning task. Reflecting on my own experience of buying used cars, I provided a reading pack with different kinds of information available to a person who would want to buy a car (see Appendix A). In the next section I will discuss the commognitive properties of the task and the support material presented to the students.

6.2 The commognitive properties of the task and support materials

6.2.1 The task

In acquisionist terms learning tasks and learning support materials serve to unpack the concept to be learned. In participationist terms, learning materials provide the focus of the discussion. The plethora of research on the design of meaningful learning tasks and support material suggests that 'a learning task' is a complex discursive object. As I indicated in Chapter 5 I chose ill-structured problems to introduce classroom discussions in order to elicit reflection and meta-level discussion of the data-contexts. I can best explain my selection of the task for this discussion and the support material with an adaptation of Vygotsky's (1986) geographical metaphor of concept development. According to this metaphor, the scientific concept to be learned is at the apex of a pyramid. At the base of the pyramid are spontaneous concepts. Among the spontaneous concepts are those with the potential to be developed to reach the scientific concept at the apex. Between the spontaneous concepts at the base and the scientific concept at the apex lies the formation of complexes and pseudo-concepts. I adapted Vygotsky's vertical metaphor of movement toward the apex of a pyramid, to a horizontal metaphor of shifting discourses as presented in Figure 7.

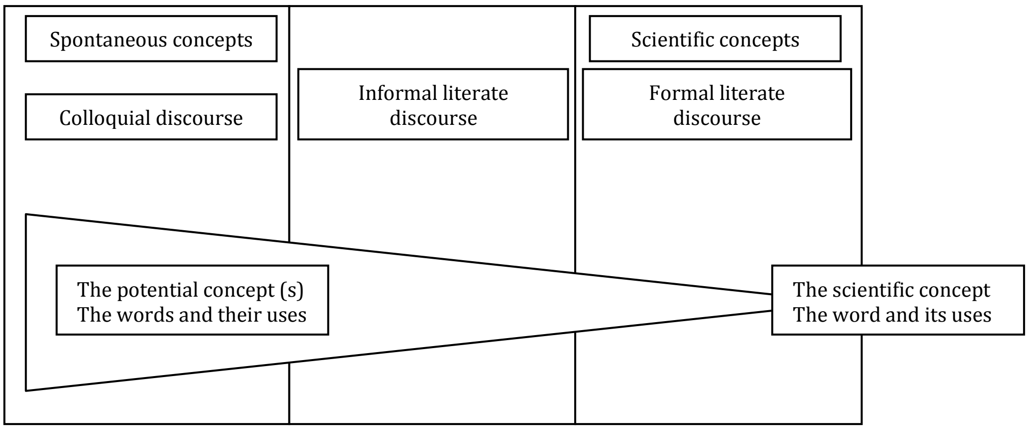


Figure 7: Concepts and discourses

Although Sfard (2008) does not explicitly map discourses to Vygotsky's stages of concept development, the mappings between spontaneous concepts and colloquial discourse, and scientific concepts and formal literate discourse is implied in her definitions of the two discourses. According to Vygotsky (1986, pp. xxxiii - xxxiv) "Scientific concepts originate in the highly structured and specialized activity of classroom instruction and impose on a child logically defined concepts; spontaneous concepts emerge from the child's own reflections on everyday experience". In comparison Sfard (2008, pp. 296 & 298) defines literate discourse as "discourse mediated mainly by symbolic artefacts created specifically for the sake of communication", and colloquial discourse as "discourse used in everyday life and developing spontaneously". Potential concepts (Vygotsky, 1986, p. 138) are "formed either in the sphere of perceptual thinking or in that of practical, action-bound thinking – on the basis of similar impressions in the first case, and on the basis of similar functional roles in the second". This notion is useful to identify words in colloquial discourse that are disobjectifications of the abstract discursive objects the students have to create. In my rendition of the pyramid of concept development, the ever-narrowing space created by the potential concepts and the scientific concept cuts through and connects spontaneous, informal literate, and formal literate discourses. In her analysis of informal discourse on number, Sfard (2008) acknowledges tasks framed in colloquial terms as appropriate, rather than in the scientific terminology that must be developed by a learner. For example, the question "how many are there?" frames a task close to colloquial discourse, compared to "what is the number?" which frames the task close to scientific discourse.[40] Similarly, the task I set, to find a reasonable price for a used car, employed colloquial terminology that has potential to be shifted to literate statistical discourse about typicality amongst variability.

I introduced the task with the verbal question: "What is a reasonable price for a used car?" The task required of the students to interpret the everyday concepts 'reasonable price' and 'a used car' statistically. I did not expand on the meaning of 'reasonable', and neither did the students ask questions about the meaning of the word at any time during the discussions. Dictionary definitions of 'reasonable' centre around the properties of being "in accordance with reason", "not extreme or excessive" and hence, "moderate, fair" (Merriam Webster

[40] Sfard (2008) argues that discourses have diffuse boundaries and that the discursants co-constitute the enacted discourse, hence I do not claim that my word use framed tasks uniquely within a specific discourse.

Online Dictionary). The online Oxford dictionary also gives "as appropriate" in as one of the synonyms of reasonable.[41] Both dictionaries indicate that a "reasonable person" would propose or accept such a price based on sound judgement. In their research about young children's understanding of average, Mokros and Russell (1995) identified 'reasonable' as one of five different mental constructs of average. The others are the mode, the mean algorithm, a midpoint and a point of balance. Of these constructs, only 'reasonable' is directly available in colloquial discourse, and the authors found that young students who interpreted average as reasonable referred to subjective contextual information in their elaborations. Hence, the colloquial term 'reasonable' was accessible as a spontaneous concept, and average as a measure of centre of a data set was the scientific concept in the literate discourse. Between 'reasonable' and its statistical abstraction are processes to establish reference classes, to compare many prices within a distribution, and decide on criteria for a statistical measurement that would be reasonable in the data-context.

The second aspect of the task, namely to construct the statistical use of the phrase "a used car", required students' awareness of a suitable reference class and the need for large numbers of observations. My reference to 'a' car, rather than 'the' car or 'this' car was intentional. In colloquial discourse, the indefinite article 'a' is used to indicate a single case, for example a single car, although not a particular car. The use of 'a' suggests a notion of generality and has the potential to evoke consideration of an aggregate of cars as a reference class. Alternatively I could have phrased my question as "What are reasonable prices for used cars?" using the plural to signify an aggregate of cars and an interval of prices. I could also have asked "What is the price of a typical used car?" which would emphasise the used cars as cases, rather than price as the variable. Both alternatives can be structured into statistical questions in more than one way, but common to the structuring process would be decisions about cases and variables, categories and reference classes. I chose the question format which I reasoned would be least directive, since I wanted to elicit reasoning about concomitant variables and constitution of suitable reference classes.

Based on the definition of informal statistical reasoning that I formulated in Chapter 3, narratives that show awareness of a suitable reference class, the need for many observations

---

[41] Retrieved from www.oxforddictionaries.com on 17 July 2012.

and variation belong to informal statistical discourse. Such narratives were described in statistics education literature as global rather than local views of data (Ben-Zvi & Arcavi, 2001). Extrapolating back to colloquial discourse, I hypothesised that practical, non-statistical discourse would be evident from local contextual narratives about one specific car that satisfies the personal criteria for decision making. Global colloquial narratives about an aggregate of cars in order to decide on a reasonable price for a general car would have the potential to develop to literate discourse. Narratives in informal literate statistical discourse, even without consideration of numerical data, would be about an aggregate of prices-of-cars to compare, with less emphasis on the cars as physical objects and aimed at description rather than decision making. Formal statistical discourse would be realised through narratives about prices abstracted as measurements, and the use of statistical objects as abstractions of 'reasonable' price.

6.2.2 The support materials

I designed support materials to "grow down" (Vygotsky, 1986, p. 194) from the scientific concept to its roots in colloquial discourse. Hence, the support materials reflect my decisions about realisations of key re-enactments of the scientific concept reasonable price as a measure of centre of a data-set of prices. I designed the support materials from the end to the beginning, re-enacting the desired discursive shifts: from the statistical summary back to statistical graphs of distributions, further back to bar graphs that use length as a metaphor for distance and time, further back to tables that represent simple discursive objects and measurements, and finally back to the actual process of finding information about concrete objects. I did not present the support materials in this strict linear order to the students, but rather allowed for them to move between the realisations in order to complete the task.

The reading pack included colloquial and informal statistical information about the used car market.[42] The boundaries between colloquial and informal statistical information are diffuse and the information can be placed on a continuum between discourses. For information to be colloquial, the chain of disobjectification to concrete objects and simple discursive objects

[42] My selection of contextual information was already a structuring of the data-context, but a reasonable one which constrained the learning task, but left it ill-structured. I controlled the general model of the cars and selected only Toyota RunX models. My choice was motivated by the serendipitous occasion that a colleague at Wits bought a used RunX at the time, and I noticed my undergraduate students' keen interest in the car.

must be short. I viewed information that is available to a person on the street who is interested in the context, and phrased in colloquial language, as colloquial information. Similarly, graphical representations that are available to a lay person interested in used cars, such as tables in a newspaper and on websites, are forms of colloquial information. Such tables are closer to informal statistical objects, since they signify aggregates, and sources of variation through the column headings. Informal statistical information uses literate language (terms that signify abstract discursive objects) and abstract graphical representations. Informal statistical information is likely to have been produced by a statistically literate agent for the purpose of communicating with non-statistical discursants.

Endorsed opinions

The first item in the reading pack is an electronic article which was published on the website of Fin24 on May 22, 2008 at 12:00 am.[43] On a continuum between colloquial information and informal statistical information, this article is slightly closer to colloquial information. The author addresses practical problems that potential buyers of used cars encounter. He starts out by evoking the practical problem of incomplete information about used cars offered for sale, and shares information about a new website that stores extensive information about all cars legally registered by their vehicle identification number (VIN). He then relates opinions of manufacturers about the resale value of their models, and opinions of officials in the used car market, such as that of a ‘solutions executive’. The article contains two statements that are colloquial but recognisably statistical. First, the author claims that there were at the time of his writing twelve million cars in South Africa, and that the average age of the cars was seven to eight years. Second, he claims that sales of used cars had not declined as much as the sales of new cars, which had declined by 15 percent compared to the previous year. The author does not provide the sources of his information. Since the article was published online, under the auspices of a listed company, Media 24, it represents an endorsed colloquial narrative about the used car market.

Personal opinions

The second item with colloquial information is a blog with entries related to car sales and the official trade values of used cars. Among other opinions and anecdotes, some blog entries

[43] I provide the hard copy of the article and of the blog entries in Appendix A rather than including a reference in my list of references, since I could not access them on the website again at the time of this writing.

question the accuracy of the depreciation tables compiled by TransUnion, a credit bureau proclaimed as South Africa's leading provider of automotive information. The depreciation tables are used by car sales people to determine the trade value of used cars. I quote some relevant entries:

> *LandRover figures wrong*
>
> …Your figures are wrong, when I tried to trade this in in 2005, the depreciation was R260,000…
>
> -- by LandRover on March 20 2007, 11:18
>
> *Depreciation overstated*
>
> Does the stated cost prices [sic] take discounts, including finance rates well below prime into account?
>
> -- by Anonymous on March 20, 2007, 10:34
>
> TransUnion should consider listing 2 columns: one for cars value [sic] if BOUGHT during the first half of a manufacturing year, and anopther [sic] column for the 2nd half of the year – hopefully this will go along [sic] way towards more accurate rating of…more
>
> -- by Anonymous on March 21, 2007, 21:54
>
> *flawed methodology* [sic]
>
> It ignores the fact that the new list price quoted excludes any optional extras that have been paid for when buying the car new, while the trade value includes these options, therefore the…more
>
> -- by me on March 22 2007, 07:51
>
> *What the article has to say on the matter:*
>
> Each figure is based on a car in standard condition with an average mileage. In addition, the figures do not include extra options or gadgets.
>
> -- by Jack on March 22, 2007, 12:53

The blog entries represent individual opinions and subjective, experiential narratives, rather than endorsed narratives. The pedagogical importance of the blog is threefold: many entries question a set of statistical summaries (the depreciation table) that is used as a norm in the practical context of buying and selling used cars; the entries raise awareness about sources of variability in prices of used cars; and they use informal statistical terminology such as "standard condition" and "average mileage". These narratives are about the system dynamics of the used car market and add relevant complexity to the data-context at the start of the investigative cycle of statistical investigation.

Endorsed statistical information

Where the blog entries provided informal opinions of a colloquial nature, the depreciation table compiled by TransUnion is a statistical tool that is officially endorsed in the used car market. The depreciation table provided official figures as facts, rather than as opinions, without any information about how they were compiled. Table 5 shows the depreciation table for Toyota RunX cars used in our task.

Table 5: The depreciation table for used Toyota RunX cars

New Car Prices compared to their Trade Values in 2007: Source Moneyweb.co.za									
Toyota	**New Car Price in 2004 (R)**	**Trade Value of 2004 model (R)**	**Depreciation of 2004 model % pa**	**New Car Price in 2005 (R)**	**Trade Value of 2005 model (R)**	**Depreciation of 2005 model % pa**	**New Car Price in 2006 (R)**	**Trade Value of 2006 model (R)**	**Depreciation of 2006 model % pa**
RunX 140i RT	133079	88200	-12.81	133079	100500	-13.1	139250	114000	-18.13
RunX 140i RS	143955	93600	-13.37	143955	106300	-14.07	145400	120700	-16.99
RunX 160i RS	159331	99800	-14.44	159331	113500	-15.6	161000	128800	-20
RunX 160i RX	177625	105100	-16.05	177625	121800	-17.19	179400	141000	-21.4
RunX 180i RX	188024	112000	-15.86	188024	128800	-17.23	189900	147900	-22.4
RunX 180i RSi	202475	119700	-16.07	202475	140100	-16.82	204500	163600	-20

From a commognitive perspective, the depreciation table in Table 5 is a visual realisation that belongs to literate statistical discourse. Yet, it is presented to people who do not necessarily participate in statistical discourse, but in the practical discourse of trading cars. In Chapter 4 I argued that the informal (popular) scientific narratives used by participants in literate discourse to communicate with lay people are qualitatively different from informal scientific discourse developed from colloquial discourse. The depreciation table is a popular science rendition of the formal concept of depreciation, and the statistical narrative of its construction is hidden from lay people.

Colloquial examples of used cars

The reading pack contains one additional set of colloquial data, namely a table of information copied from a website that advertised used cars for sale. Any person who wants to explore prices of used cars and has internet access can obtain such data.

The information I provided as examples of cars appeared on the relevant website already in tabulated form, which is an instance of statistics in the air (Nisbett, et al., 1993). Table 6 shows an excerpt from the data table copied in 2008 from the website www.cars4sale.co.za. In the first column of the table an image of each car was provided and in the second column a description of the car. The description gave the model of the car, for example TOYOTA RunX 180i RSi, or TOYOTA RunX 140i RT, as well as claims about the condition of the car (e.g. mint condition) and extras like tinted windows or a compact disc player. In five additional columns the year of manufacture, the mileage, the asking price, the location and the website reference of each car are provided.[44]

Table 6: Table with information about a selection of used cars

From your selection-10 vehicles have been found					Edit Search	New Search
Image	Description	Year	Mileage	Asking Price	Location	Parking lot
	**TOYOTA RunX 180i RSi** 141 kw. FSH, T/B, Smash and Grab Tinted Windows. Full House …more info AA Service Plan Available	2004	120,000 km	R120 000 Negotiable	Kwa-Zulu Natal, Durban	Ref #:569f3b
	**TOYOTA RunX 180i RSi** Mint condition …more info AA Service Plan Available	2006	35,757 km	R110 000	Gauteng, KEMPTON PARK	Ref #:798f2d

All the information in the table is data, but not all is statistical data. In fact, without a statistical question the information merely refers to properties of individual cars and none of the information is statistical data. The description column provides rich opportunities for discussions of suitable reference classes, drawing for example, on general knowledge that cars with bigger engines tend to be more expensive than smaller cars.

The content of the reading pack that I have described so far serves a role similar to that of concept cartoons (Keogh & Naylor, 1999), namely to provide a range of views, opinions and arguments about real-life situations that are related to the concept or learning task in focus. As teachers of statistics, the students will have to engage with written and spoken texts about data-based contexts, hence, in contrast to concept cartoons, I did not minimise the role of written language. The reading pack also contains informal statistical information that I

[44] I made a pedagogical decision to provide only data about Toyota RunX cars in order to manage the learning process. I judged that leaving the context any less structured would require huge investment in terms of time to obtain a shared dataset.

designed with the pedagogical goal of developing statistical reasoning, or shifting from colloquial to informal statistical discourse.

Informal statistical information about used cars

I used the colloquial information from the cars4sale website to construct an abstract statistical data table. The data set had information about 85 RunX cars, organised as cases, along with a selection of variables. Table 7 shows an excerpt from the statistical data table.

Table 7: An excerpt from the abstract data table

www.cars4sale.co.za 15/7/2008					
Model	**Year**	**Km**	**City**	**Colour**	**Price (R)**
Toyota RunX 140i RS	2007	3000	Pretoria	GOLD	129950
Toyota RunX 180i RX	2007	5000	Cape Town	SILVER BLUE	165000

To construct this table I removed the photographs of the cars and the detailed descriptions of condition and extras. During the design stage of the course, I deliberated that the detailed information may be relevant for a person in the act of buying, but they are not salient concomitant variables in relation to price. However, I retained the information about colour and the cities in which the cars were located, in order to generate opportunities to discuss different kinds of variables, such as categorical or measurement variables.

In anticipation that the students might simply look for the price in the middle of an ordered list, I decided not to order the data by price. Similarly, anticipating that the students might simply look for a middle car (not too old and not too new) if the year of manufacture is ordered, I refrained from ordering by year of manufacture. Such quick decisions would be counterproductive for developing discourse about price as a variable. The kilometre readings of the cars were sufficiently variable that finding the middle value, even in an ordered list, would not be too obvious a task, so I ordered the data by the kilometre readings. I posited the acceptance of the abstract table as a tool for the investigation of prices of used cars, as the first outcome of structuring the data-context.

Lastly, in preparation for the classroom discussion about a reasonable price, I constructed a selection of graphical realisations of the data in the abstract table. Two graphs provide horizontal case value bars which facilitate focus on price and mileage as measurements of individual cars. I decided on this format based on Gravemeijer's (2002) reasoning that value bar graphs retain the relationship between the case and its measurement, and that the horizontal bars provide a metaphorical link with the property measured. Specifically, the length of a horizontal bar representing a kilometre reading is spatially related to length as distance travelled. Vertical bars are more readily associated with frequency of occurrence. In order to conceptualise a set of measurements as a statistical distribution, the discourse must shift from narratives about comparing cars to narratives about comparing prices as measurements. Gravemeijer (2002) argues that dot plots serve as a transitional symbolisation en route to density distributions of measurements. For this reason, I included in the reading pack dot plots of the prices and the kilometre readings of the data in the abstract table, to facilitate focus on the measurements. The dot plots are informal statistical representations, but not colloquial at all, since they are not used in colloquial sources of information. Figure 8 shows a horizontal value bar graph and Figure 9 shows a dot plot of the same data.

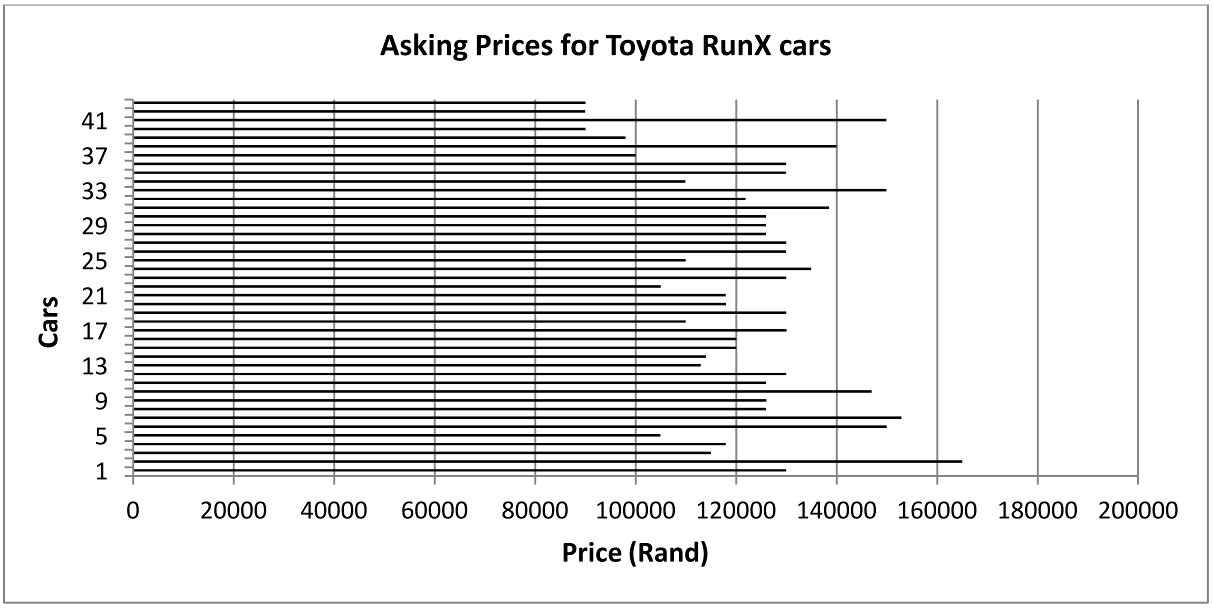


Figure 8: A value bar graph of asking prices of a sample of used RunX cars

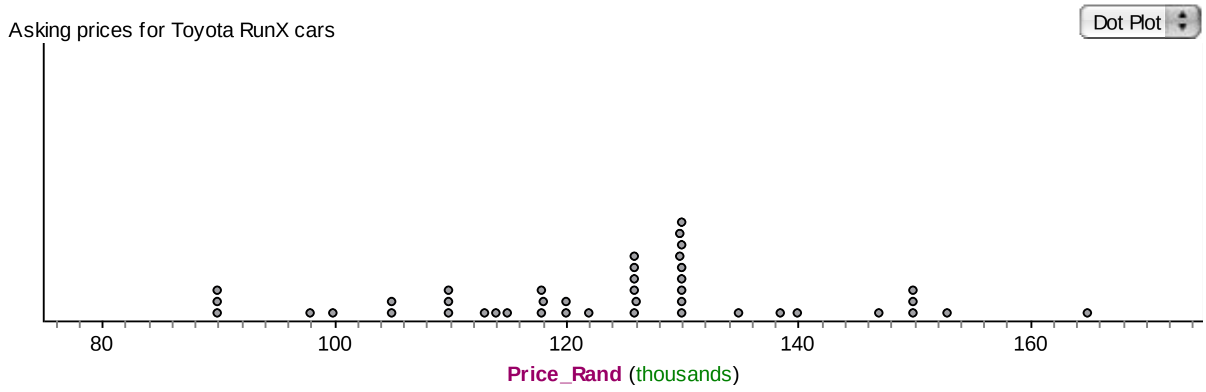


Figure 9: A dot plot of the asking prices of a sample of used RunX cars

The dots represent the end points of the value bars and the dot plot is an informal visual representation of the density of the data at different positions along the measurement line.

From a commognitive perspective, the reading pack provided narratives about used cars in different discourses. The electronic article and the blog entries belonged to colloquial discourse, where the mediators are concrete objects or simple discursive objects. Similarly, the website information about used cars for sale were colloquial narratives. On the other hand, the depreciation table belonged to literate statistical discourse, but was presented as informal statistical information. Lastly, the discursive objects designed for learning statistics, namely the abstract data table, the case value bar graphs and dot plots, belonged to informal literate discourse.

6.3 The role of questions in statistical inquiry

Question-posing plays a key role in relating context and data throughout the investigative and interrogative cycles (see Figure 1, Chapter 1) and various researchers support a call for focused development of problem-posing skills through teaching (Arnold, 2008; Pfannkuch & Horring, 2005; Sanchez & Blancarte, 2008). Arnold (2009) distinguish between question-posing and question-asking. Both processes are intended to directly facilitate the use of statistical methods. While question-posing involves the formal structuring of a context necessary to proceed to a statistical investigation, question-asking weaves through the investigative process spontaneously and continuously. The purposes of question-asking are to *interrogate* and check procedures and findings during investigation and to *analyse* data in order to make decisions. Similarly, Arnold distinguishes two types of posed questions according to the purpose of the poser: *survey questions* are aimed at getting data and *investigative questions* can be answered with the data at hand. There are three types of investigative questions: Summary questions (ask for a description of a data set); comparison questions (about comparison of data sets across a common variable) and relationship questions (about the interrelation between two paired variables) (Graham, 2006; Pfannkuch & Horring, 2005). A slightly different use of the phrase ‘investigative questions’ is suggested by Allmond and Makar (2010) since in their study there were no data at hand, but objects about which statistical questions had to be formulated.

Allmond and Makar (2010) assessed the investigative questions formulated by nine year old children, drawing on the work of Chin and Kayalvizhi (2002) and Arnold (2009) to categorise the questions at seven different levels of a hierarchy. From the lowest level, the hierarchy is described as follows:

a) *Irrelevant/Off topic questions:* Questions with no clear connection to the stimulus
b) *Non-mathematical questions:* Questions that do not require mathematics to reach a solution (e.g. How do peaches grow?)
c) *Non-investigative questions:* Questions that are impractical to investigate in terms of classroom constraints; can be answered simply by checking a reference; or have no real purpose or interest for others (e.g. How many peaches does an average peach tree grow in a year without any growth food?)
d) *Closed questions:* Questions that have a definite answer that can be worked out using information directly available (e.g. How much fat, sugar and salt are in a can of peaches?)
e) *Questions with potential for investigation:* Questions that can be refined into investigative questions (e.g. What size are the peaches inside [a can]?)
f) *Investigative questions:* Questions that can be answered by using statistical methods (e.g. Is there the same amount of peaches in every tin?)
g) *Inquiry questions:* Investigative questions containing ambiguities that need to be negotiated (e.g. What peach manufacturer sells the healthiest tinned peaches?) (Allmond & Makar, 2010, p. 3).

The analytical focus of the researchers was on the appropriateness of questions for learning statistics in the confinement of a school classroom. Depending on the data/information at hand, the example of a non-investigative question could have been an inquiry question, and the example of an inquiry question could have been a non-investigative question. Both questions require colloquial concepts to be operationalised: What is an average tree? What is meant by healthy? Yet, judged against my definition of informal statistical discourse, both questions show awareness of variation and the need for a large number of observations, and both questions require specification of a relevant reference class.

For clarity, Figure 10 graphically summarises the categories of statistical questions derived from statistics education research.

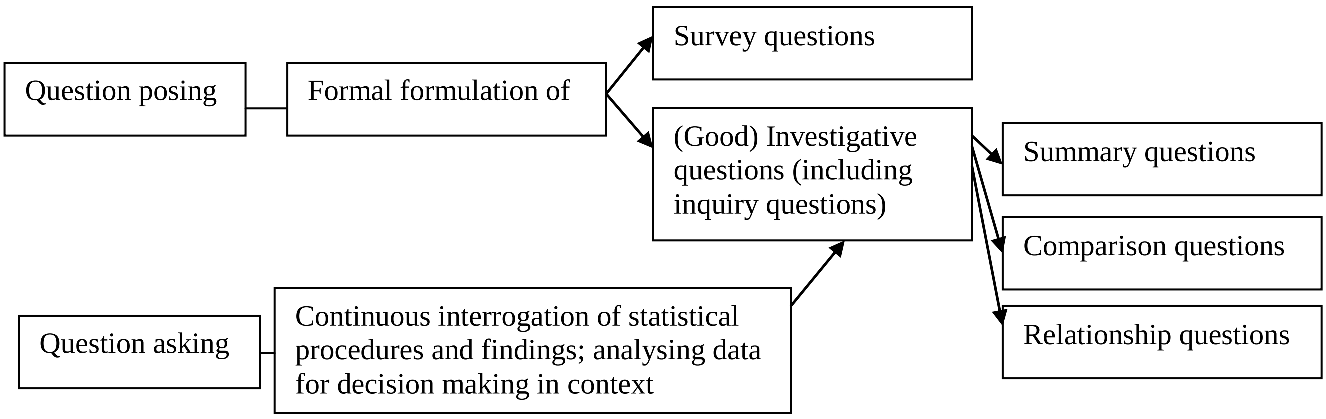


Figure 10: Graphical summary of types of statistical questions

Although I was aware of the need for posing good investigative questions, my interest was in what my students' questions reveal about their coming to understand the data-context in relation to statistical exploration. As part of a discourse, even irrelevant, off topic or non-investigative questions may reveal contextual concerns and information about the attitudes and dispositions of the discursants. The existing categories for statistical questions therefore did not fall within my research interest. I will discuss my codes for analysing my students' contextual questions in the next section.

6.4 Coding contextual questions

I referred the students in my study to the reading pack as a source of contextual information and asked them to formulate questions that can be answered with the data at hand, hence from my perspective as lecturer I requested investigative questions as described by Arnold (2009). The students wrote down their questions and spontaneously engaged in small group discussion during the process.

With my interest on the students' informal statistical discourse about the data-context I analysed the written questions according to their purpose, as is evident from my interpretative elaborations. Interpretation of the purposes yielded two clearly distinct groups of questions. I named the first group *evaluative questions*. I define evaluative questions as contextual questions that are concerned with evaluation of the trustworthiness of agents in the data-context, and evaluation of the consequences of aspects of the context for someone who wants to act in the data-context.

In order to code questions as evaluative, I looked for references to agents in the data-context. Specifically, the use of the term "who" indicated reference to a person or a group of people as agents. Questions that referred to agents or the active role of specific agents, such as "the car dealer" were also coded as evaluative. In addition to references to agents, questions with implicit or explicit evaluations in terms of good, bad, fair, and credible or biased were coded as evaluative questions. Such evaluations are subjective opinions rather than facts.
In Table 8 I present the evaluative questions with my interpretative elaborations.

Table 8: Evaluative questions about the data-context

Evaluative Questions (Questions related to social interaction and practical action)	**Interpretative elaborations.** What are the questions about?
E1: Should I have the car checked myself or believe it has been AA[45] approved? (GG)	*Trustworthiness of agents*: Who takes responsibility? *Consequences for practical action*: One has to safeguard oneself in social dealings. An authority is needed.
E2: Who collects statistics? Are they biased toward new car dealers? Second hand car dealers? (DH)	*Trustworthiness of agents*: Who provides the information? *Consequences for practical action*: Is the information objective or biased toward the seller? One has to safeguard oneself against use of biased statistics.
E3: Where does data come from – who is the custodian of the database? Any industry regulations and checks in place? (DH)	*Trustworthiness of agents*: Who takes responsibility for the data? *Consequences for practical action*: One has to safeguard oneself against data unregulated and unchecked by an authority.
E4: Credibility of second hand car dealers? (DH, RC)	*Trustworthiness of agents*.
E5: Why do motor dealers not give you the full history of the car? How credible is information about pre-owned cars: accident history, mileage, owner, where driven? (S, SM)	*Trustworthiness of agents*: Credibility, motive behind selective information. *Consequences for practical action*: One has to safeguard oneself against exploitation by agents.
E6: How does the car dealer take the seller for a ride when he uses the price guide? (GK)	*Trustworthiness of agents*: Procedure behind adverse use of a market standardisation tool. *Consequences for practical action*: One has to safeguard oneself against use of a market standard.
E7: Depreciation: Is this good news for the buyer or the seller of a second hand car? (NM)	*Consequences for practical action*: implications of a market standard for an individual. Good news for the buyer may be bad news for the seller. Someone is at a disadvantage.
E8: What exactly is the potential pitfall of the guide (mileage)?[46] (GK)	*Consequences for practical action*: Dangerous/adverse consequences due to the use of the market standard. One has to safeguard oneself against use of a market standard. How does the system work?
E9: How reliable are the depreciation figures? (RK)	*Consequences for practical action*: Consequence/implications of a market standard
E10: How do they check out the book value of the car? (SM)	*Practical procedure* used by an agent to apply a market standard in practice.

The second group of questions were also based on anecdotal opinions taken from the blog, yet moved beyond the need for social trust and imagined action, towards exploring the variable under consideration, namely price. I named the questions in this group *exploratory questions*, which I define as contextual questions aimed at investigating the structure of the data-context and possible relationships between concomitant factors. Exploratory questions

[45] Automobile Association
[46] After a blog entry that argued that the official book price does not take into account the difference in kilometre reading of a car offered for sale in January and in December of the same year. Seemingly unaware that the book price takes into account 'average mileage' the blogger felt this is a pitfall of the book price.

are alienated and abstracted. They do not refer to human agents, but use nouns to refer to variables that influence each other. Table 9 presents the students' exploratory questions with my interpretative elaborations.

Table 9: Exploratory questions about the data-context

Exploratory Questions (Questions related to the world out there)	**Interpretative elaborations.** What are the questions about?
X1: Reputation of the make of car – is this the determinator (sic) of percentage depreciation, or purely economic factors, or track record/good service/high standard? (DH)	*Exploring structure*: How does the system work? Separating "purely economic factors" from social factors such as "reputation", suggesting concurrence between depreciation and causal factors.
X2: Why is depreciation on pre-owned cars lower than on new cars? (GK)	*Exploring structure*: How does the system work? Awareness of differences in variation between subsets of cars.
X3: What is the cost of maintenance? (SM)	*Summative conclusion*: No indication of variation.
X4: What is the most reliable (resale) brand of car to buy second hand? (GG)	*Summative conclusion*: Awareness of variability among "brands" but wants an extreme value (most).
X5: Which car manufacturer depreciates the least? (KH)	*Summative conclusion*: Awareness of variability among "manufacturers" but wants an extreme value (least).
X6: New versus previously owned – which is the better deal? Which is the better percentage depreciation? Reliability? (DH)	*Summative conclusion*: Awareness of variability between categories "new" and "previously owned".
X7: Are private deals cheaper – bypass second hand car dealers? (DH)	*Summative conclusion*: Awareness of variability between subsets "private deals" and "second hand car dealers". Investigate the assumption that private sales are cheaper than dealer sales.
X8: When car prices go up, is depreciation changed by same percentage across all models? (DH)	*Exploring a relationship* between change in antecedent and consequent. Awareness of co-variation, looking for trend or pattern. How does the system work?
X9: "Green cars" – influence old petrol car sales? (DH)	*Exploring a relationship:* correspondence between price and a change in the system. Sales may refer to frequency or supply-demand effect on price.
X10: Influence on price if petrol price goes up? (DH)	*Exploring a relationship*: Awareness of co-variation between price and a change in the system.
X11: How does mileage influence the price of the vehicle? (GG, SDS)	*Exploring a relationship*: Awareness of co-variation between mileage and price.
X12: Repossessions up with lending rate up, means more pre-owned cars to sell, does price go down? (DH)	*Exploring a relationship*: Co-variation between supply and price. How does the system work?
X13: Is there a correlation between the asking price of a car and its (a) mileage (b) age (c) its extras? (KH)	*Exploring a relationship*: Correlated change between price and concomitant variables. How does the system work?
X14: What price is appropriate for each year model for each area? Or even colour? What criteria is your focus? (SDS)	*Exploring variation*: Awareness of variation in price in different reference classes.
X15: What price is appropriate (not too cheap not too expensive) for a 2nd hand RunX? (SDS)	*Exploring variation*: Judgement based on comparison of aggregate of prices.

I also analysed the students' exploratory questions in terms of the kind of answer anticipated (Franklin & Garfield, 2006). This perspective relates question and possible answer in an argument structure. The question provides a claim, for example "Why do motor dealers not give you the full history of the car?" (Question E5 in Table 8), claims that motor dealers withhold information from a prospective buyer and that they have a reason for doing so. An answer to this question has to be supported by some kind of evidence. Kuhn (1991) classifies evidence in arguments as non-evidence, pseudo-evidence or genuine evidence. Since I discussed Kuhn's threefold classification of evidence in Chapter 3, I provide only a brief summary here:

a) *Non-evidence* denotes not only the absence of a need for evidence but also evidence that merges with the claim so that effect is cited as evidence of the cause.
b) *Pseudo-evidence* are scripts of a personal and experiential nature, consisting of examples and illustrations that are unfalsifiable to the reasoners immersed in the context. Pseudo-evidence include unquestioned, personal acceptance of statements by people in positions of authority.
c) *Genuine evidence* distinguishes clearly between claim and support. Indirect genuine evidence provides plausible analogy in assumption of the proposed relation, and direct genuine evidence relates claim and evidence by referring to correspondence, explicit co-variation or correlated change between instances of the claim and instances of the evidence.

Answers to the evaluative questions in Table 8 are bound to be supported by at most pseudo-evidence, since judgement of people's motives is never an objective endeavour. In the evaluative questions, explicit references are made to sources of authority like the AA (Automobile Association) (Question E1) and the "custodian of the data base" (Question E3). Implicit questions about credibility can best be answered by experiential evidence from other agents.

Continuing the analogy between questions and evidence, genuine questions are questions for which the answer is based on genuine evidence that refer to facts about objective properties of objects. In Kuhn's classification, genuine evidence is external to a claim. For example, answering the question "If there are more pre-owned cars on the market, do the prices come down?” (after Question X12) requires evidence of co-variation between the number of cars offered for sale and the price of used cars.

Using the type of evidence needed to answer the question as a coding principle, I derived the following three types of questions:

a) *Non-questions:* Questions that make assumptions about motives of agents cannot be answered by providing evidence, only by experiential narratives in support or refutation of the assumptions. An example of a non-question is E5: “Why do motor dealers not give you the full history of the car?”
b) *Pseudo-questions:* Questions that interrogate motives can be answered by citing plausible opinions of some authority. An example of a pseudo-question is E9: “How reliable are the depreciation figures?” and X2: “Why is depreciation on pre-owned cars lower than on new cars?”
c) *Genuine questions:* Questions that can be answered by investigating the objective context, for example question E7: “Depreciation: is this good news for the buyer or the seller of a second hand car?” and X7: “Are private deals cheaper–bypass second hand car dealers? Genuine questions refer to correspondence, explicit co-variation or correlated change between concomitant variables.”

There were no non-questions among the exploratory questions. I further analysed the exploratory questions that were coded as genuine questions in terms of their potential for informal statistical discourse. To be coded as informal statistical questions, the questions had to indicate awareness of variability, suitable reference classes and the need for multiple observations. All the exploratory questions are informal statistical questions,[47] although they do not necessarily refer to measurements as data.

6.5 Analysis of the questions

6.5.1 Colloquial questions and statistical questions

Both exploratory and evaluative questions are framed in colloquial discourse. They make no use of the formal discourse of statistics. The exception is the reference to correlation in X13

[47] Question X3 is included as a question that can be answered statistically, but only if the assumption of variation across a sample of cars is made. I cannot claim that participant SM assumed variation based on this question alone.

that might have been taken from statistical discourse.[48] Table 10 shows that sixty percent of the questions were exploratory (15) compared to forty percent evaluative (10). Evaluative questions were asked exclusively by only two out of the twelve students (RC and NM), both with previous experience of introductory statistics. KH and SDS, also students with previous knowledge of statistics as a subject, asked only exploratory questions. The rest of the students asked both evaluative and exploratory questions.

Table 10: Frequency of evaluative and exploratory questions

Type of question	Total asked	Number of students who asked
Evaluative questions	10	2
Exploratory questions	15	2
Both		8

Evaluative questions in my analysis are likely to be categorised as irrelevant, non-mathematical, non-investigative or closed, when using Allmond and Makar's (2010) categories, or even as non-questions using Arnold's (2009) categories. Nevertheless, I argue that although these questions do not refer explicitly to the statistical data at hand and were not clean questions for the purpose of immediate exploratory data analysis, they reveal actual contextual concerns and assumptions at the start of the investigative cycle. The order in which I listed the evaluative questions in Table 8 shows a subtle shift from questions that are about social trust (Questions E1 to E6), to questions where the notion of a procedure that guides action is evident (Questions E7 to E10). The variety of questions hint at how difficult it is for teachers to achieve an appropriate exploratory disposition when the discourse is not explicitly statistical.

6.5.2 Properties of the questions

Spontaneous positioning as evident from the discourse of the questions

I view evaluative questions and exploratory questions as related on a continuum of spontaneous discursive positioning in the data-context, rather than as two distinct categories. At the extremes of the continuum are complete immersion in the data-context and complete ignorance of the implication of answers in the data-context. This proposed continuum is an

[48] Participant KH who wrote down the question previously did a statistics course and taught statistics at high school level.

interpretative act on my behalf, which allows me to relate to my students' apparent discursive positioning.

The evaluative questions are indicative of a desire "not to be taken for a ride" in the event of getting practically involved in the used car market. As I discussed in Chapter 3, Kuhn found that the majority of her participants provided pseudo-evidence (Kuhn, 1991) for their own causal claims. Applying Kuhn's definition of types of evidence to the kinds of answers that would apply to my students' questions, I argued that evaluative questions can be answered by pseudo-evidence, hence they are pseudo-questions when it comes to the norms of good everyday reasoning. I further argued in Chapter 3 that Kuhn's examples of pseudo-evidence are indicative of the participants' immersion in the context, and here the evaluative questions suggest similar immersion in the data-context.

The exploratory questions seem to suggest greater freedom from the constraints of imagined action and situational decision-making. The askers may still position themselves as agents in the context, but their action is investigation rather than decision-making. A related finding was reported by Bakker, who had to instruct children in his study to position themselves as data analysts rather than actors in context in order to elicit more profound answers and "objective and precise arguments" (2004b, p. 115). From a commognitive perspective, this difference in positioning is related to a shift in discourse through the process of alienation, which has been defined by Sfard (2008, p.295) as "using discursive forms that present phenomena in an impersonal way, as if they were occurring of themselves, without participation of human beings."

None of the exploratory questions referred to the role of agents in the data-context. Comparison of two questions from either category, where they are close to each other on a continuum of positioning, clearly illustrates the difference in alienation between the evaluative and exploratory discourses:

> *Evaluation question* E10: "How do they check out the book value of the car?" (SM)
>
> *Exploratory question* X1: "Reputation of the make of car – is this the determinator [sic] of percentage depreciation, or purely economic factors, or track record/good service/high standard?" (DH)

The person asking the evaluative question was highly cognisant of the role of agents in determining the book value of the car as presented in the depreciation table. In contrast, the person asking the exploratory question referred to an alienated and objectified "reputation" or other "determinators of percentage depreciation" without reference to human actors.

Dispositions and ways of reasoning

I cannot claim that the students were exclusively considering the kind of questions that statistics can answer, as I instructed in posing the task. However, the mere fact that they were attending a statistics course is likely to have raised such considerations. More importantly, their questions revealed beliefs and hypotheses about the practical context and the need to understand the system dynamics of the used car market. The assumptions and tacit judgements in context that are revealed in the evaluative questions, suggests deterministic reasoning that is likely to influence subsequent data analysis and conclusions. As much as a good investigative question sustains and focuses reasoning throughout the statistical investigative process, unresolved and unstructured situation models have the potential to prevent students from making sense of statistical processes and conclusions.

A clear assumption that emerged from the evaluative questions is that buying a used car is a risky endeavour and requires a watchful eye, both on being done in by the system and by individual agents. Hence a disposition of scepticism is evident which is indicated as a desirable disposition in the fourth dimension of the model of statistical inquiry (Wild & Pfannkuch, 1999).[49] In this model, the scepticism is directed toward the meaning of statistical data and findings in relation to the context. Here, the scepticism is about the role of agents in imagined transactions. Attitudes of scepticism foreshadow possible investigations. For example, a student who suspects that car manufacturers somehow control the prices of used cars will be sensitive to differences in price between a private seller and a dealership. Strategic thinking and seeking of explanations are general types of thinking that are evident from the evaluative questions, despite the questions being aimed at evaluating social relationships in the event of practical action. These attitudes are valued by Wild and Pfannkuch (1999) and are included in the second dimension of the model of statistical inquiry.

---

[49] See Figure 1 in Chapter 1.

The absence of the sense of risk emanating from the exploratory questions is striking in comparison to evaluation questions. The range of exploratory questions is indicative of dispositions of imagination, curiosity and awareness (Wild & Pfannkuch, 1999) of various concomitant factors that influence prices of used cars. These dispositions manifested in the absence of statistical data when an ill-structured problem was posed about a variable quantity, namely that of price. The goal of exploratory questions was to explore possible relations between the prices of cars and factors such as mileage, age, buying from a private person or a car dealer, and different types of cars.

6.5.3 Informal statistical discourse

Statistics requires variation and uncertainty for its existence (Watson, Kelly, Callingham, & Shaughnessy, 2003). Similarly Nisbett and his colleagues (1993) propose that awareness of variation and suitable reference classes and the need to consider a large number of observations are necessary, although not sufficient, for informal statistical reasoning in everyday contexts. While variability is omnipresent in everyday contexts and people are intuitively aware of variability, it does not follow that people generally use this awareness in reasoning. Kuhn (1991) shows that the propensity to provide genuine evidence related to co-variation is hardly the pattern in adult reasoning. Awareness of variability in the context of prices of used cars is clearly evident in the exploratory questions. Yet, despite the evident awareness of variability, some exploratory questions require local answers (E.g. X5: Which car manufacturer depreciates least?) rather than global answers. According to Ben-Zvi and Arcavi (2001) local views of data do not support understanding of data as distributions. In addition, such local, summative questions hint at a concern with specific and extreme data-values, a phenomenon that is widely reported in statistics education literature.

Local and global views of the data-context

I remind the reader that the students' questions were not based on a statistical data set, by their own choice, but on contextual information. Yet, evaluative questions and exploratory questions differ also in their reference to specific objects, or aggregated objects. Evaluative questions, focused on imagined personal action, not surprisingly referred mostly to specific cars or prices. This is evident from the use of the pronoun *the* in combination with imagined personal action:

E1: Should I have the car checked myself or believe that it has been AA approved? (GG)

Personal action is not explicit in the following example, but "check out", a colloquial term that means *to verify as correct*, evokes the image of a specific car with a specific book value. When the narrative concerned other agents, the students used the plural form to aggregate the agents as if there is no variation among them. Compare the reference to "they" in question E10: How do they check out the book value of the car? (SM)

In contrast, the exploratory questions showed consideration of a contextual aggregate of used cars. The students' alienated discursive reference to cars, prices or dealers in plural, or *a car* in general, is indicative of the constitution of a global view of the context, rather than a local consideration of one specific car. Even reference to *the car* in questions X11 and X13 in relation to co-variation cannot be interpreted as referring to any specific car as a case.

Awareness of reference classes

The notion of a contextual aggregate as a reference class is admittedly vague at this stage, and not adequately delineated for statistical treatment. The huge conglomerate of used cars had not been organised for meaningful comparison of price yet. The students had not used differences between for example Mercedes and Toyota as cars to suggest comparison of prices within categories. Nevertheless, categories for comparison are hinted at in the majority of exploratory questions. For example, intrinsic properties like 'model', 'brand' or 'make of car' (Questions X1, X4 and X5), and "green cars" (Question X9) became available as categories alongside which to structure reference classes, while systemic properties like 'private sale' or 'dealer sale' (Question X7) and the 'lending rate' (Question X12), became available as sub-categories for the comparison of the variable 'price'.

I indicated in Chapter 3 that role-governed categories are harder to form than categories based on observable features of members. In order to form role-governed categories, the relationships must first become clear. Relationships between variables in a data-context are based on proposals of causes related to observations. I will illustrate with an example:

Question E5 asked: Why do motor dealers not give you the full history of the car? How credible is information about pre-owned cars: accident history, mileage, owner, where driven?

Question X13 asked: Is there a correlation between the asking price of a car and its (a) mileage (b) age (c) its extras?

Both questions list factors that are somehow related to the price of a used car. The implicit causal relationships are plausible: if a car had been involved in an accident it should cost less than a car that had not been in an accident; a car with higher mileage should cost less than a car with less mileage; a car that belonged to the proverbial ‘old lady’ should cost more than a car that belonged to the proverbial 20 year old male; a car with valued extras should cost more than its basic counterpart. Yet, there was no evidence in E5 that these relationships are understood in the practical context. The asker of E5 did not relate the factors to variable prices in a way that is useful for comparison. Which cars will be included in a reference class based on accident history? Will these cars be similar enough to compare in terms of price? Nor is it plausible to use ‘owner’ or ‘where driven’ as criteria to create a reference class in which to compare prices. Similarly, it is not plausible to structure a reference class based on extras. On the other hand, the relationships between mileage and price, and age and price in question X13 are useful to create reference classes. One can plausibly investigate the prices of used cars in a reference class of cars with less than 100 000km on the odometer, or in a reference class of cars less than ten years old.

Contextual questions as good investigative questions

As a group, exploratory questions are indicative of summative, comparison and relationship questions, which other researchers classify as statistical questions in tasks based on statistical data sets (Graham, 2006; Pfannkuch & Horring, 2005). Summary questions can be answered by a statistical summary value (X3 to X7, X14 and X15), comparison questions aim to compare two data sets across the same variable (X8), and relationship questions can be answered by statistical investigation of correspondence or co-variation between variables (X10 to X13).

Question X15: “What price is appropriate for a 2nd hand RunX?” is closest to a *posed* question, ready for statistical investigation, since the reference class is clearly stated, where second hand RunX cars are the cases and the statistical task is clear, to determine a price that is not too expensive or too cheap, hence an *average* price. Answering question X15 statistically requires a representative (random) sample of prices of second hand RunX cars and determining a central value for the variable ‘price’. The students had such a dataset available, yet only one student consulted the data and asked the question. Neither was question X15 taken up as a possible route for statistical treatment during the collective discussion later in the session. Consideration of the other exploratory questions reveals the oversimplification of the context that would have resulted from a narrow statistical answer to X15.

6.6 Summary

I will summarise my findings in relation to the questions posed at the start of this chapter.

a) What are the objects and the narratives that are realised when the students pose and discuss questions about the data-context?

The objects of evaluative questions tend to be agents in the imagined context. The used car dealer, the seller of a used car and even the car that is considered are simple discursive objects, since they can be literally pointed out in the actual situation. The narratives of evaluative questions are about the *effect on me* or *consequences for me* or at best *the effect on or consequences for the person on the street* (as a generalised me), when getting actively involved in the context. Exploratory questions tend to refer to abstract discursive objects like depreciation, cost of maintenance and references to relationships between properties. Hence, the narratives of exploratory questions are less overtly concerned with practical consequences and are narratives concerned with how the system works that encompasses used cars.

b) What shifts are evident from the posed questions between colloquial discourse and literate statistical discourse?

A continuum is suggested by the differences in focus among the questions. The focus ranges from trusting agents in the event of personal action in the context, through awareness of general and formal procedures in the context toward increasing focus on objective properties of the context. The notion of a continuum suggests possibilities for shifting from enacted colloquial discourse toward an increasingly alienated discourse on the relationships between the variable properties of objects. I refrained from creating a hierarchy of questions, since my purpose was to describe the questions in relation to students' structuring of the context and not to evaluate the quality of the questions.

c) What are constraining and productive questions in the shift towards statistical discourse?

At this stage, I hypothesise that questions that positions the reasoner as an agent in the imagined context may constrained shifts toward statistical discourse. I expected students with previous experience of statistics to interpret a question about *reasonable price* to refer to data about prices and variables that influence prices. An appropriate initial response to the task would have been: We have to find a reasonable price, so let us first look at price data of a sample of used cars. Yet, issues of social trust were of pressing concern for all students, including those who had previous experience with statistics. On the other hand, those students who did not have previous exposure to statistics were not unable to explore factors that influence price. Neither group considered the statistical data about prices as adequate information to pose or answer questions about *reasonable price* thus far. Rather, the anecdotal blog entries were appropriated overwhelmingly in the formulation of questions.

Exploratory questions have generative potential for informal statistical discourse, since they suggest narratives about the structure of the context, rather than imagined action in the context. These questions inter alia show awareness of a global view of the context, co-variation between variable contextual properties, the emergence of hypothetical causal lines and the need for typical values.

Figure 11 below graphically summarises the key findings from my analysis of the written questions. The questions are placed on a proposed continuum constituted by evaluative questions and exploratory questions.

At this stage, I do not have evidence for possible causal lines between the types of questions posed and the properties I obtained from the analysis. For example, I do not claim that a shift in spontaneous positioning in relation to the context (from actor to observer) is responsible for a shift in abstraction of the discourse. I do conclude that the questions provided by the students are an early indication of the importance of the authentic statistical reasoning task of "grasping system dynamics" as part of the investigative cycle of statistical inquiry.

In the next chapter, I will analyse the classroom discussion of the students' questions. I will show that the narratives about the data-context is characterised by routines that enable and constrain the development of statistical discourse.

Figure 11: A continuum of contextual questions

E1, E2, E3, E4, E5; E6

Focus: trust, agents

E7, E8, E9, E10

Focus: trust, procedures

X1, X2

Focus: Structure of the System

X3 – X6

Focus: Summative, Local view of properties of the context

X7 – X13

Focus: Co-variation, global view of context, variable properties related by causal hypotheses

X14 – X15

Focus: Co-variation, global view of context, variable properties related by causal hypotheses; typical values required

Evaluative questions ⟷ **Exploratory questions**

Evaluative questions	Exploratory questions
<u>Positioning of the asker</u> Imagine self as an agent, e.g. potential buyer	<u>Positioning of the asker</u> May imagine self as actor, but increasingly less focus on other agents as social actors and more focus on observed objective properties of the context
<u>Disposition</u> Sceptical of the role of other agents; risk involved in decision making	<u>Disposition</u> Diminished sense of risk; curiosity about relationships between concomitant factors
Immersion in the data-context allows only limited awareness of price as a variable	Alienation and abstraction of the concept price and related concepts
<u>Pseudo-questions</u> Can be answered by citing experience or opinions of trusted agents	<u>Genuine, potentially statistical questions</u> Can be answered by measuring and comparing
Case based questions (local view of data-context)	Aggregate based questions (global view of data-context)

Chapter 7: Data analysis: The colloquial narratives behind contextual questions

7.1 Introduction

Wild and Pfannkuch (1999) emphasised the role of context knowledge at the inception of a question or idea that can be investigated statistically. They said "The earliest stages are driven almost entirely by context knowledge. Statistical knowledge contributes more as the thinking crystallises" (Wild & Pfannkuch, 1999, p. 228). This applies even in the case of professional statisticians, who take up "grasping the dynamics of a system, problem formulation, and planning and measurement issues" as a statistical task (Wild & Pfannkuch, 1999, p. 225). This prompts the question: How does thinking crystallise?

7.2 Analysis of the group discussions of the written questions

My analysis of the written questions formulated by my students showed that previous knowledge of statistics was not sufficient to formulate good investigative questions. For some of the students, imagined action in the data-context and social relations were so prominent that their questions were not related to the variable *price* in any direct way. Although the written questions were the result of crystallising aspects of the context, the questions were embedded in spontaneous small group discussions, and elaborated on in a 'whole class discussion'. Analysis of these discussions showed that the students were even more disproportionately concerned with subjective practical and social quandaries, than they were with objective aspects that would facilitate statistical reasoning. In this section I analyse narratives that emerged in the discussion of the written questions. I will show that despite awareness of variability and sources of variability, and despite appropriate references to cars and variable prices in the plural, instances of informal statistical discourse were limited.[50] Further, despite noun use that suggests objectification of the desired variable *price,* even the exploratory questions remained embedded in subjective, enacted stories about the data context. In terms of teaching and learning, the whole class discussion served as opportunity to endorse or reject individual narratives.

Sfard (2008, p. 300) defines a narrative as **"**a series of utterances, spoken or written, that is framed as a description of objects, or processes with or by objects, and is subject to endorsement or rejection, that is, to be labelled as "true" or "false.""

From the class discussions I identified two distinct types of narratives, namely narratives aimed at judging social trust and validity and narratives about the implications of contextual information for hypothetical action. I called such narratives evaluation narratives, since they map onto the evaluative questions in Table 8 (Chapter 6). I searched for investigation narratives aimed at constructing a base for understanding the context in terms of objective properties, as well as in terms of the comparison of different variables in relation to prices of used cars; but such references were hardly ever more than individual utterances to share an exploratory question with the class.

I will exemplify each of the two kinds of narratives with excerpts from the discussions. In selecting the excerpts I took care to draw together narratives by excluding utterances that wove through but belonged to other narratives. The complete transcript of the discussions in Session 2 is available as Appendix B. The excerpts are identified in terms of the small group discursants, or as whole class discussion and the numbering of the turns reflect the temporal order of the contributions during the discussion.[51]

---

[50] I defined informal statistical discourse as discourse about variability, sources of variability, large numbers of observations and appropriate reference classes.
[51] The complete transcript in Appendix B reports the small group discussions one after another, although they occurred in parallel, filmed by two video cameras.

7.2.1 Social evaluation narratives

Social evaluation narratives are characterised by questioning the motives of social agents in the context. The narratives describe agents in the data-context as having malicious intent.

Table 11: Excerpt 1. First example of a social evaluation narrative

Excerpt 1. Session 2: Group B: Discursants KH, SM, RK, GK		
Turn	**Discursant**	**Utterance**
16	Lecturer	So what have you learned about the second hand car market?
18	KH	Some models depreciate more than others.
19	GK	The second hand car depreciate less than the new car.
20	Lecturer	You mean the older a car is the slower the value depreciates?
21	GK	And then the other thing is what they they they advertise is not how it is (inaudible).
22	Lecturer	In what sense?
23	GK	In the sense …that…ah…they give you the things you want to hear…they'll tell you …ah…about the kilos, about who was the previous owner, and then they don't tell you about…what's that now…
24	RK	If it's been his for a long time.
25	Lecturer	Mm, they spin you yarns?
26	GK, RK	Yes.
27	Lecturer	Yes, yes I think you are right. Look at the information, be careful, there is an opinion here that wow the car is great value. I think carry on with what you are suggesting, but ask questions now…Maybe on your own, get a range of questions…
28	KH	We must get proper questions, now. Which can be answered with this data.
…	…	…
37	GK	Mmm. But is it true the older cars they depreciate slower than the new cars?
38	KH	(Nods affirmatively). They say if you drive a new car out of the showroom, as you drive it out it depreciates.
39	SM	MmMm.
40	GK	So the older car, isn't that the same thing?
41	KH	No, because it already has depreciated.
42	GK	(Frowns, turns her head away)
43	KH	(Inaudible) You can buy an old car, an old skedonk[52] for a thousand rand, it cannot depreciate much more.
44	GK	Maybe if we analyse this last…we can get a good question…
45	GK	Formula for the trade price…
46	KH	There isn't a standard depreciation. Different cars depreciate

[52] A colloquial Afrikaans term for a dilapidated car.

		different amounts.
47	GK	...about when in year a car is bought compared to the trade price
48	KH, GK	Who is disadvantaged by this guide?

From Excerpt 1 (Table 11), it is clear that issues related to trust were important to the participants. In Turns 21 to 27, GK and RK and I engaged in an evaluation narrative which signalled mistrust in the motives of used car dealers. What "they" advertise is "not how it is" (Turn 21) and they seem to withhold information on purpose (Turn 23). I endorsed the narrative, based on the evidence in the reading pack. In the case of being actively involved in buying a used car, one should not accept information uncritically. That is practical wisdom. KH seemed to have noticed that practical wisdom is not at stake in this discussion and suggested they should find "proper questions" (Turn 28). While GK had earlier noticed the difference in depreciation between new and used cars as an aspect of the system (Turn 19), the group continued to create a narrative in which depreciation became a tool to be used to the disadvantage of buyers or sellers. In Turns 37 to 48, GK and KH discussed the practical implications of depreciation. KH accepted the pattern that older cars depreciated less than new cars. However, this proposed pattern seemed to pose a problem for GK, who frowned and intentionally turned her head away (Turn 42). If GK understood the relationship between age and value of a car as linear and negative, changing rates of depreciation would create cognitive conflict. My hypothesis is supported by GK's proposal to find a "formula for the trade price", hence a fixed rule by which to calculate the trade price of a car (Turn 45). KH (Turn 46) countered with an explanation of variation and "non-standard" depreciation, but GK (Turn 47) followed up by observing an apparent unfairness in the variation, namely that the official trade price does not take into account when in a year a car is bought. The narrative ended with a question about a tool (the guide with official trade values, or the depreciation table) that is used to disadvantage people.

In the whole class discussion, the issue of the intended use of the depreciation table to determine trade values for specific cars was realised again in a social evaluation narrative.

Table 12: Excerpt 2. Second example of a social evaluation narrative

Excerpt 2. Session 2: Whole class discussion		
Turn	**Discursant**	**Utterance**
70	GG	Well, if…the one in January…doesn't do much driving…and the one in December has decided to do a road trip for example,…so it's mileage, ja…and I suppose it's also the condition of the car as well.
71	KH	They don't look at that. They take a look at the car, take out the table and say this is the trade value of the car.
72	RK	What if the cars are used on different terrains, you know like…the person uses the car on a good surface, and then a person uses the car for a short time…driving you know, not on the road…
73	RK	I am saying if they are using mileage to to to assess the value of…then they might be misleading this customer.

GG (Turn 70, Excerpt 2, Table 12) argues that if a buyer had a choice between two cars that were made in the same year, she should choose the used car with the lower mileage. KH retorts that mileage would not be taken into account by dealers using the official trade value guide. RK (Turns 72 and 73) finds even more reason to be suspicious of agents. Even if they take mileage into account, it would be misleading. To him, a car with low mileage, but driven on bad roads, has less value than a car with higher mileage driven on good roads. His conclusion is that used car dealers are misleading buyers if they take into account mileage and not terrain. He therefore judges the motive of such dealers, and introduces a source of variability in the 'condition' of cars, if not in their value.

The next social evaluation narrative is also about the role of other actors in the context, in this case, those who produce the data.

Table 13: Excerpt 3. Third example of a social evaluation narrative

Excerpt 3. Session 2: Whole class discussion		
Turn	**Discursant**	**Utterance**
94	DH	Uhm, one other thing that relates to where the data is from. Actually, are the people who produce these tables real statisticians, do they work for the car dealers themselves or do they work for the second hand car dealers? Last week we had one set of data and we argued both sides from one set of data. So….who's actually gathering the data and analysing this, are they statisticians, or are they just smiling…because of (inaudible).

95	Lecturer	Who gathers the data and who produces the statistics? And you are saying they do it for their own reasons…
96	DH	Do they get paid by the new car dealers, lots of money to make it look better, or are they paid by (inaudible)?
97	Lecturer	Are you interested in that?
98	All	Agreement and interest.
99	KH	It comes back to the question how they are calculating the depreciation.
100	RK	Ja, cos I think that obviously they want …(inaudible).
101	KH	If you can't find the validator in the house, and you rely on this one set (of data), how do you know where it comes from, how did they get it…
102	SDS	Is there a board, Mrs Lampen?
103	Lecturer	There is an association of second hand car dealers, yes, and uhm…the second hand car prices are, uhm, fixed within a band… While you are interested, there was a fight in the newspapers about a year ago, exactly this. Second hand cars, new cars were very cheap about a year or two ago, you could get cars with everything that opened and closed, cd-players and so on, for the price of what you would expect for a car with nothing extra. OK? So they tried to sell the new cars, and if they, and then they manipulated the price of second hand cars…said no you can't give more discount than this or this, otherwise people won't buy new cars. Right? So yes, there, there are there is politics, and there are power games for sure. That is why we must be informed.

In Turn 94, DH questions the motives of those who gather data. A narrative ensues about the problem of data obtained from so-called real statisticians and those in service of other social agents. In Turn 96, KH suggests that those who gather data may benefit financially and in Turn 100, RK seems to agree. The whole class endorses the problem statement by indicating their interest in the issue (Turn 98). So strong is the suspicion that social agents are using data to their own benefit that the only safeguard seems to be a "validator" or a board as watchdog (Turns 101, 102). My contribution in Turn 103 serves to support the notion that data must not be accepted as value free. On their own, the narratives may seem to be of passing interest only, but issues about credibility resurfaced until the end of the discussion, as the next short excerpt shows:

Table 14: Excerpt 4. Fourth example of a social evaluation narrative

Excerpt 4. Session 2: Whole class discussion		
Turn	**Discursant**	**Utterance**
184	Lecturer	All right. So let's remember this: if the answer is certain, or the question is what is the most or the least, what is the most expensive, the least expensive, then we don't need stats...(unclear) but when the answer is uncertain, one math answer, or ten maths answers can't give us the picture. Then we need stats.
185	Lecturer	I want to push on. The credibility of data. Is that something we can answer with stats?
186	Class	No.
187	Lecturer	Credibility here is a judgement that you make on trust...That's not something you can measure, therefore stats can't give you an answer. Stats can only deal with something you can measure – measurement also mean count.
188	RK	Why don't you clarify the credibility by doing a little study into what you're saying?

7.2.2 Action evaluation narratives

Action evaluation narratives are characterised by evaluating variable properties of objects (cars) in terms of their implication for personal action in the data-context. They tell stories of subjective judgements and opinions of the prudence of buying cars.

I present two excerpts to illustrate narratives about the implication of the contextual information for prudent action in the context. The first narrative is extreme in its subjectivity, and followed directly after KH's call (Turn 28, Excerpt 1, Table 11) to get proper questions.

Table 15: Excerpt 5. First example of an action evaluation narrative

Excerpt 5. Session 2: Group B: Discursants KH, SM, RK, GK		
Turn	**Discursant**	**Utterance**
29	SM	It is not worth to buy a car.
30	KH	Mmm [tentative].
31	GK	What do you mean it's not worth to buy a car. You mean it's not worth to buy a new car or an old car.
32	SM	It does not matter, if it is a new car or an old car. As long as it is a car.
33	GK	Mm?

34	SM	Ja.
35	GK	The reason it is (inaudible) the petrol price?
36	SM	The interest rate, service…cars are very expensive, OK? And then, when you buy it it depreciates.

SM's declaration (Turn 29, Excerpt 5, Table 15) that it is not worth buying a car, introduces an action evaluation narrative. He mentions variables that might influence prices of used cars that are tightly bound to owning a car: service cost, interest rate and depreciation... SM's lament (Turn 36) that "cars" are expensive and then "…when you buy it it depreciates" illustrates how subjective judgement and objective properties of cars were heaped together in a syncretised whole. His judgement of "expensive" – without any reference to actual prices – suggests that his personal financial constraints provide a norm. The suggestion is confirmed when I compare his written questions with his contribution to the discussion. SM contributed three written questions: two were evaluative questions (E5 and E10, Table 8, Chapter 6), which question the motives of used car dealers. Question E5 questions the credibility of the information dealers provide about specific cars, and question E10 questions dealers' use of the official book value to determine the price of a car. His third question (X3, Table 9, Chapter 6) was: "What is the cost of maintenance?" Hence, the only question he formulated in relation to money was not related to the price of cars, but to cost. The implication of his judgement is that whatever the book value or the relative value of a typical used car may be, it has too little or no value for him, and buying a car would not be prudent. This is an example of extreme evaluation discourse that holds little potential for appropriating an abstract discourse about the concept of a *reasonable price for a used car*. SM opts out of further exploration at this stage in the group discussion, and later in the whole class discussion he raises the question about cost again: Turn 53: "Cost of maintenance of new car and old car." (See the complete transcript of Session 2, Appendix B). He does not engage in discussion about his understanding of relationships between maintaining a new car and an old car.

A second example, this time of a more sophisticated action evaluation narrative, is presented below. Although this narrative is also characterised by evaluating variable properties of objects in terms of their implication for personal action, the action is explicitly hypothetical and purposely imagined in order to develop a reasoned judgement of price.

Table 16: Excerpt 6. Second example of an action evaluation narrative

Excerpt 6. Session 2: Whole class discussion		
Turn	**Discursant**	**Utterance**
81	SDS	OK, if I was trying to buy a car, and I was getting this data, I would look at specific criteria. Like if I was looking for a specific colour, I would choose all the ones with the colour I want and look at what is an appropriate price. Not too cheap so that there is probably something wrong with it, also not something too expensive, because then I can't afford it. So I'm looking for, I wanna say an appropriate price for each criteria – colour, or model, or area…
82	Lecturer	(Writing) What is an appropriate price for different criteria. Um, what shall we call those? Variables?...you mentioned you don't want to pay too cheap, cos then you would suspect…
83	SDS	I'd be suspicious, ja.
84	Lecturer	…and not pay too much, then you'll think you'd be done in.
85	SDS	Mm!

SDS' narrative in Excerpt 6 (Table 16) is one of personal agency rather than helplessness in a complex context where motives of other agents are questioned and relationships between concomitant variables are still opaque. In Turn 81, she poses personal preference and affordance as anchors for decision making. In addition, she suggests avoiding extreme prices, with the implication that some price between "too cheap" and "too expensive" would be a safeguard. As shown in Chapter 6 (see Table 9) SDS contributed the following written questions: X14: "What price is appropriate for each model for each area? Or even colour? What criteria is your focus?" and X15: "What price is appropriate (not too cheap and not too expensive) for a secondhand RunX?" Comparing her narrative in the group discussion with her written questions suggests that despite the apparent alienation of the written questions, they are embedded in an evaluation narrative about prudent decisionmaking in the data-context.

Since the discussion was based on the written questions that individual students produced, it seemed logical that the properties of the written questions would be transferred to the discussion. Hence, I expected investigation narratives (based on the written exploratory questions) in addition to the evaluation narratives. When exploratory utterances were realised, they remained embedded in evaluation narratives as in Excerpt 6. The discussion provided pockets of information about the students' contextual discourse that were not evident from their written questions. The major contribution of the analysis of the discussion was the identification of words and their uses that prevented the discourse from shifting to informal statistical narratives on variability in prices of used cars.

7.3 Words and their uses that prevented the emergence of informal statistical discourse

The objects of the exploratory questions (Table 9, Chapter 6) are transaction procedures, and properties of cars in the used car context, seemingly "as if they occur without human participation" (Sfard, 2008, p. 295). Exploratory questions are therefore the result of alienation. Yet these questions did not evolve into discussions of relationships between variables. My interest was drawn to the possibility of different discourses in the classroom discussion, by SDS's narrative in Excerpt 6 (Table 16). SDS uses the term "appropriate price" in relation to cheap and expensive. But she does not judge cheap and expensive relative to the same measure. 'Too cheap' indicates "something wrong with the car" and 'too expensive' indicates "…I can't afford it" (Turn 81, Excerpt 6, Table 16) 'Too cheap' is referenced to a judgement of the condition and value of a car, and 'too expensive' is related to a judgement of its cost to her as an individual. Similarly, GG, KH and RK relate a judgement, "condition", to "trade value" and "the value of [the car]" in their narrative (Turns 70 to 73, Excerpt 2, Table 12). While trade value is represented in monetary terms,[53] and the value of a car can also be represented by an amount of money, condition is a judgement for which the discursants' criteria were implicit and subjective. A closer look at their word use is needed to describe the students' discourse.

[53] If not explicitly in their narrative, then in the reading materials.

7.3.1 The uses of price, value and cost in the discussion

The terminology used in the discussion was appropriated largely from the reading pack. In these examples of colloquial discourse, the complex relation between price and value is glossed over. In Table 17 I present some examples from the articles and readers' blogged comments (See Appendix A).

Table 17: Words and their uses in colloquial discourse on price and value of cars

Example		Interpretive elaboration
	From "Investment Insights: Car values continue to crash"	
1	"We compared the prices of vehicles when they were new to their present trade value."	New vehicles have prices (pl.); used vehicles have a trade value (sing.)
2	"…the Crossfire, plummeted in value by more than 40% pa."	The type of car lost monetary value
3	"…the real price was always below the stated retail price…"	A car has a real price (sing.) that differs from a stated retail price (sing.)
4	"…an increase in new prices is always good for residual values and that is why the three German manufactured cars held their value."	If new car prices (pl.) increase, some types of cars hold their value (sing.) [in comparison to their price when new]
	From Fin24: "The truth about used cars"	
5	"The value is given as a range, since many factors influence the price"	The trade value (sing.) is a range for comparison to the price (sing.) of a specific car
	From the blogged comments	
6	"Why is the price of second hand vehicles not declining. It remains very hard to find good value for money?"	The [stated retail] price (sing.) of all second hand vehicles is too high in comparison to [prices of new cars] to be good value for [a buyer's] money

In these examples, price and value are sometimes used as synonyms, e.g. the stated retail price (Example 3) is a synonym for the trade value. The plural phrase "residual values" refers to the difference between the trade value and the total amount used to finance the sale of a specific car at the end of the finance period. Hence, residual values vary among cars. Apart from this instance, the term 'value' is used in the singular or as a mass noun, while 'prices' are used as often as 'price'. The message from the car trade is that the trade value is a stable amount in comparison to car prices that vary.

Example 6 (Table 16) in particular presents a challenge to a discursant who does not have technical context knowledge of the car market: 'value for money' seems to be judged by this discursant on two scales. One scale is the additive difference between new car prices and the trade value of used cars. On this scale, the smaller the difference, the less incentive there is to buy a used car. The other scale is a subjective scale of affordability. What must the monetary difference between new price and used value be in order for the lower price to be value for money for an individual buyer?

I proceeded to compare the uses of the words price, value, and cost uttered during the discussions in Session 2. The results are presented in Table 18.

Table 18: Comparison of word phrases including price, value or cost.

Word phrase including 'price'	N=39	Word phrase including 'value'	N=15	Word phrase including 'cost'	N=21
(Second hand) car prices	9	Car values	1	The cost of maintenance, service	10
Prices of cars	4	Value of cars	1	Cost of a car	3
A price for it	1	Prices de-value	1	Maintenance cost	3
Its price	1	Resale value	2	Calculating the cost	1
The price of the car	7	Car with this value	1	Service will cost; it will cost me; car will cost more	4
The asking price	1	Car is great value	1		
The selling price	1	The value of [the car]	1		
(The) petrol price	5	Lost their value	1		
The trade price	2	The value (depreciates)	1		
(Worth) the same price	2	Trade value	1		
An appropriate price	3	(Are the) value of the cars (the same)	4		

In relation to cars as objects, the plural *prices* was realised thirteen out of 39 times during the discussion, and the singular nine times. This suggests that the students were appropriately aware of variation between prices of cars. On the other hand, the singular *price* was used nine times altogether to refer to the discursive objects *asking price*, *selling price*, *petrol price*, and *trade price*. This suggests that the students did not talk about these as variable entities in the same way as variable car prices. It is evident from their contributions that the students were aware of variation. Compare for example DH's utterance in Turn 52 of Session 2[54]: "The influence that petrol price has on car prices", and her proposal to investigate the relationship: "You take the petrol price over as long a time as possible; we need new car prices, obviously by some kind of average, initially, you might break it down…" (Turn 107, Session 2).

Value was realised less than half as often as *price* during the discussion and only once in plural form. In Turn 2 of Session 2, I said: "So, I found on the internet, after a news item this week, information about car values." I referred to the trade value table (See Appendix B) with values for different models and makes of cars, and not to values of individual cars. I used the term to imitate literate use, without awareness of subtle differences between price and value, and oblivious to possible confusion that may arise due to this usage.

Nineteen out of 21 times 'cost' was used in relation to service or maintenance of a car, hence something that has to be done. 'The cost of a car' was used only three times out of 21. All three times 'the cost of a car' was invoked as a comparison for cost of maintenance in an argument that older cars will have higher maintenance cost than newer cars. In Turn 129 (See the full transcript of Session 2 in Appendix B), KH suggests a way to compare maintenance costs of cars in relation to age: "Well I think, if you compare them, stats is a good way to compare. So I mean, it can be something as trivial as a pie chart, you know, saying this is the cost of the car and this percentage is maintenance…"

I have shown that during the classroom discussion 'price' was realised as monetary amounts more than any of the other nouns. Scanning the word phrases that contain 'price', it seems that price was reified and used appropriately as a noun. That means 'price' was an object that could be acted on discursively, yet the students did not proceed to compare prices with the aim of realising *reasonable price*. The term

---

[54] The transcripts of the utterances that I refer to without referencing an excerpt are available in Appendix B, the full transcript of the discussion in Session 2.

“appropriate price” was realised by SDS as “not too expensive and not too cheap” in one written question (X15) and again by SDS once only in the discussion (Turn 81, Excerpt 6, Table 16). I will argue in the next section, that price was not alienated as yet, as its uses remained tied up with uses of value, and hence the students could not act on the object ‘price’.

7.3.2 Price and value in ontological collapse

Ontological collapse is evident when our word use flattens discursive hierarchies so that discursive objects and primary objects are given the same status as “things in the world” (Sfard, 2008, p.57). Our use of the words ‘price’ and ‘value’ collapsed into intrinsic properties of cars and supported evaluation narratives. In my introduction to Session 2 as lecturer, I used price and value indiscriminately to refer to prices of individual cars and official trade prices of aggregates of cars. In Excerpt 7 (Table 19) I present my whole narrative at the start of the session.

Table 19: Excerpt 7. Session 2: Introduction of the data-context

Excerpt 7. Session 2: Introduction of the data-context		
Turn	**Discursant**	**Utterance**
2	Lecturer	So, the main thing that I think we have to start finding out is what can statistics do and what kind of questions can we answer with statistics…So, I found on the internet, after a a news uhm item this week, information about car values. And I know the men here like cars, and the women here like cars. And we like to have smart cars, don’t we, and new cars. But the information here says we have to consider: Is a car an investment? So I am giving you the information I found, from a website, car prices…and I want you to spend the time to look it through, scan quickly to see what people on the internet said about prices of cars. After that you will see there is a data set that I got for you, of second hand prices of one type of car, a Toyota RunX. After that, page 11, are official data that says how car prices devalue. So, your first task now… is: get to know the context, what happens to the value of second hand cars? Is it worth buying a new car? I am giving you about 20 minutes to look at the data and you come up with questions that you can ask that we can answer with the data. So, I want a list of questions that is answerable with the data that you have and that is about the situation with second hand cars.

I used the term ‘value’ in various forms, intended to signify objective comparison between monetary amounts, yet the reference classes for comparison were unclear and my uses did not operationalise ‘value’ for measurement: In Excerpt 7 (Table 19)

> “…information about car values” may have referred to values of individual cars, or to types of cars as in the official guide or depreciation table.
>
> “…is a car an investment?” required a judgement about the change in the value of a car in general. An investment increases in monetary value, and some cars are investments.
>
> “…car prices devalue” may have referred to change in individual car prices, or to change in prices of cars in general.
>
> “Is it worth buying a new car?” required a judgement about the change in value of a new car in general, compared to change in value of a used car.

Similarly my uses of ‘price’ were open to interpretation in terms of reference classes, although price is measured in currency:

> “…information from the website, car prices…” may have referred to the new car prices in the depreciation table, or to the list of prices in the abstract data table (See Tables 5 and 7, Chapter 6).
>
> “…what people on the internet said about prices of cars” may have referred to statements about prices of cars in general, or prices of individual cars, or types of cars.
>
> “second hand prices of one type of car” referred to the prices of the cars in the abstract data table.
>
> “official data that says how prices devalue” may have referred to new car prices of aggregates of cars as indicated in the depreciation table (see Table 5, Chapter 6), or to both new car prices and trade value in consecutive years as in the depreciation table.

Yet, despite my ambiguous word usage, the students did not seem to be confused about reference classes, but rather about the relationship between price and value against their narratives of mistrust and prudent action in the data-context. GK summarised the key concern in Turn 66 in Excerpt 8 (Table 20).

Table 20: Excerpt 8. Value and price in an evaluation narrative

Excerpt 8. Session 2: Whole class discussion		
Turn	**Discursant**	**Utterance**
62	GK	About the apparent pitfall…how is the…pitfall….disadvantaging the buyers? You know, that they say the price, uh, they compare the prices of the new car and…uh…the old car within a period of time. Maybe they work out the interest, I mean the depreciation. They say the depreciation ranges from January to January the following year. If you buy the [new] car in between the year, and then you want to sell it for the next year, where are you going to fall there? Are you going to be disadvantaged?
63	Lecturer	Are you saying, if you buy a car in 2005, and you buy it in December 2005…
64	KH, GK	Will we sell it at the same price?
65	Lecturer	When bought new – Jan v Dec 2005, what will be better buy in 2006.
66	GK	Maybe, can we say are the cars worth the same price? Are the value of the cars the same?
67	Lecturer	(Writing): Are those cars still of same value? OK, and I think you have identified here (referring to questions written down in turn 54), that one thing that plays a role there is the mileage.
68	GG	But, if a person bought it in January, and didn't do much, say, ja, there is no mileage on the car, then it is probably better to take the one from January.
69	Lecturer	Why?
70	GG	Well, if…the one in January…doesn't do much driving…and the one in December has decided to do a road trip for example, so it's mileage, ja…and I suppose it's also the condition of the car as well.
71	KH	They don't look at that. They take a look at the car, take out the table and say this is the trade value of the car.
72	RK	What if the cars are used on different terrains, you know like…the person uses the car on a good surface, and then a person uses the car for a short time…driving you know, not on the road…
73	RK	I am saying if they are using mileage to to to assess the value of…then they might be misleading this customer.

GK's question in Turn 66 (Excerpt 8, Table 20) realises *worth*, *value* and *price* in relation to each other, rather than as synonyms. I rephrased her question as I wrote on the white board. In my rephrasing (Turn 67), I omitted the term *price*, and took up *value* instead. In addition, I related increasing mileage to change in value. My motivation in the moment was based on common sense reasoning, that older cars should be cheaper than newer cars. If cars are equally old in years, older can be measured by mileage. Indeed I aimed for the students to structure mileage as a concomitant variable related to price. In Turns 68 and 70, GG confirms the proposed relationship between *value* and *mileage*. When KH (Turn 71) discounts the use of mileage to determine the trade value, rather than the value of the hypothetical car, the subjective use of *value* emerges. RK's (Turn 72) inclusion of different terrains as a variable that influences the value of a car, suggests more than a direct relationship between price and value, and casts my incorporation of mileage also in a subjective light. It is intuitively valid to assume that use on bad roads would have consequences for the condition (Turn 70) of a car. Indeed, an endorsed statement from the reading pack states: "Each figure [in the official trade value guide] is based on a car in standard condition with average mileage."[55] How to judge standard condition and average mileage is left to intuition. GG (Turns 68 and 70) seems to have an intuitive sense of over-use and under-use in terms of mileage. A "road trip" implies a sharp increase in mileage over a short period of time. A buyer will not be able to judge the probability of such a road trip if she does not have an intuitive awareness of average mileage.

At this stage, I want to draw together three excerpts to conclude my argument that the ontological collapse evident from our use of price and value as intrinsic properties of used cars support evaluation narratives. They are Excerpt 5 (Table 15), SM's opinion that it is not worth buying a car; Excerpt 6 (Table 16), SDS's opinion about cars with extreme prices, and Excerpt 8 (Table 20) above, which relate condition explicitly as an aspect of a car's value. These excerpts tell a story of colloquial discourses that are incommensurate with literate discourse about the data-context. Although all the narratives in these excerpts are evaluation narratives, they differ in degrees of abstraction. I looked for evidence of alienation and indentified shifts between value-for-me discourse, value-for-money-discourse, value-of-a-car discourse and relative value-of-cars discourse.

7.3.3 Evaluation narratives in a value-for-me discourse

Contrary to the use of value in the official trade value guide, the concept of value was appropriated on a personal scale by the students during the classroom discussion. It had

---

[55] See "Investment Insights: Car values continue to crash" in the reading pack in Appendix A.

the meaning of “relative worth, a good value at the price.”[56] Objective judgement of relative worth of a car presupposes a reference class of other cars, similar enough in other properties than price. Yet, I have shown that an extreme evaluation narrative prevents comparison of prices of used cars to judge relative worth. Instead, the worth of a used car is judged relative to personal, subjective factors. SM’s narrative in Excerpt 5 (Table 15) suggests a discourse on ‘value-for-me’.

A less extreme discourse on value-for-me is realised by SDS in Excerpt 6 (Table 16). SDS judges “appropriate price” on two different scales. At least one of her judgements belongs in a value-for-me discourse: “Not too expensive” is judged on the basis of her own financial position, “because then I can’t afford it.” SDS posed question X15 (Table 9, Chapter 6), which hints at the statistical average through the realisation “not too cheap not too expensive”. Yet in her narrative, it is clear that not too cheap and not too expensive lie on different scales: the condition of the car scale (‘something wrong’ to ‘excellent’) and the affordability scale (‘cannot afford’ to ‘cheap in relation to my pocket’). A mean price for a used car cannot be placed on these scales.

Narratives in a value-for-me discourse cannot be endorsed or rejected by anyone else, and hence prevents shift toward literate discourse. Values of the variable price in literate statistical discourse have to be alienated from personal judgements in terms of expensive or cheap. Without the discursive alienation, the meta-discourse about statistical summaries may be bypassed so that the mean price remains a contextual judgement on a personal scale of value-for-me, rather than an objective description of a data-set.

7.3.4 Evaluation narratives in a value-for-money discourse

GG’s judgement in Turns 68 and 70 (Excerpt 8, Table 20), is an example of placing cars on a relative scale of value-for-money. This evaluation scale is tied up with intuitions and experience about what one can get for a given price. Value-for-money discourse is

[56] Online Merriam Webster dictionary, accessed on: 15 October 2012

more alienated than value-for-me discourse, although its narratives remain evaluative. The evaluation is not completely subjective anymore (there is no indication of “too expensive for me”), but related to measurable properties of cars, in this case *mileage*. Her evaluation is that it is “probably better” (Turn 68 in relation to Turn 66, Excerpt 8, Table 20) to take the car with less mileage than a car of the same price with higher mileage.

7.3.5 Evaluation narratives in a value-of-a-car discourse

SDS (Excerpt 6, Table 16) as well as GG and RK (Excerpt 8, Table 20) relate the value of a car to *condition.* SDS (Turn 81, Excerpt 6) says if a car is too cheap “there is probably something wrong with it”. I interpret RK’s reference to “different terrains” (Turn 72, Excerpt 8), on which a car was used as a proposal of a causal relationship between condition of the road, condition of the car driven on the road, and the value of the car. Less explicitly, the value of a car is also related to overuse and underuse through measurement of mileage (Turn 68, Excerpt 8). Similarly, GK (Turn 62, Excerpt 8) voices her unease with the practice of assigning the same book value to cars that were *new* in January and December of the same year, respectively. For GK, the book value as a measurement is not yet an object parsed from value as a property of a specific car. In fact, GK explicitly questions the conflation of trade value and value of a car (Turn 66, Excerpt 8). Although a judgement of *condition* is likely to be on a subjective, evaluative scale between bad or something wrong and excellent, here judgement is based on properties of cars that are perceived to influence its value.

7.3.6 Evaluation narratives in a relative value-of-cars discourse

In Turn 71 (Excerpt 8, Table 20), KH endeavours to shift the discussion towards accepting the system and compare values of aggregates of cars, relative to each other, regardless of specific differences. But KH’s narrative is rejected. In Turns 72 to 75 (Excerpt 8) RK realises his doubt that the official book price is a reliable indication of the value of a car. He seems to suggest that the book value is used in practice to negate differences in value of individual cars. Evaluation narratives are incommensurate with narratives in a relative-value discourse.

7.3.7 Ontological collapse: Fusion of evaluative and descriptive scales

Sfard defines ontological collapse as follows: "[The] phenomenon of taking all the objects that are being talked about – discursive objects [like price, value, and condition] and extra-discursive objects [such as used cars and the money in my pocket] – as belonging to the same category of "things in the world" that pre-exist discourse, with their mutual relations similarly "objective" and mind-independent" (Sfard, 2008, p. 300). The notion of ontological collapse is not foreign in statistics education research. Bakker (2004b, p. 101) observes that "An underlying problem is that middle grade students generally do not see 'five feet' as the value of the variable 'height', but as a personal characteristic of, say, Kate. In addition to this case-oriented view, students should learn to disconnect the measurement value from the object or person measured, and to consider data against a background of possible measurement values." Bakker's observation is based on interviews with seventh grade students about the meaning of average in context, and related to calculation of the mean. I have shown that the teachers in my study showed similar reasoning when they try to understand the data-context.

In Figure 12 I present the students' narratives on value and price based on my interpretation of their questions and their discussions. The composition of the figure reflects a pedagogically desired trajectory towards investigation of the variable price.

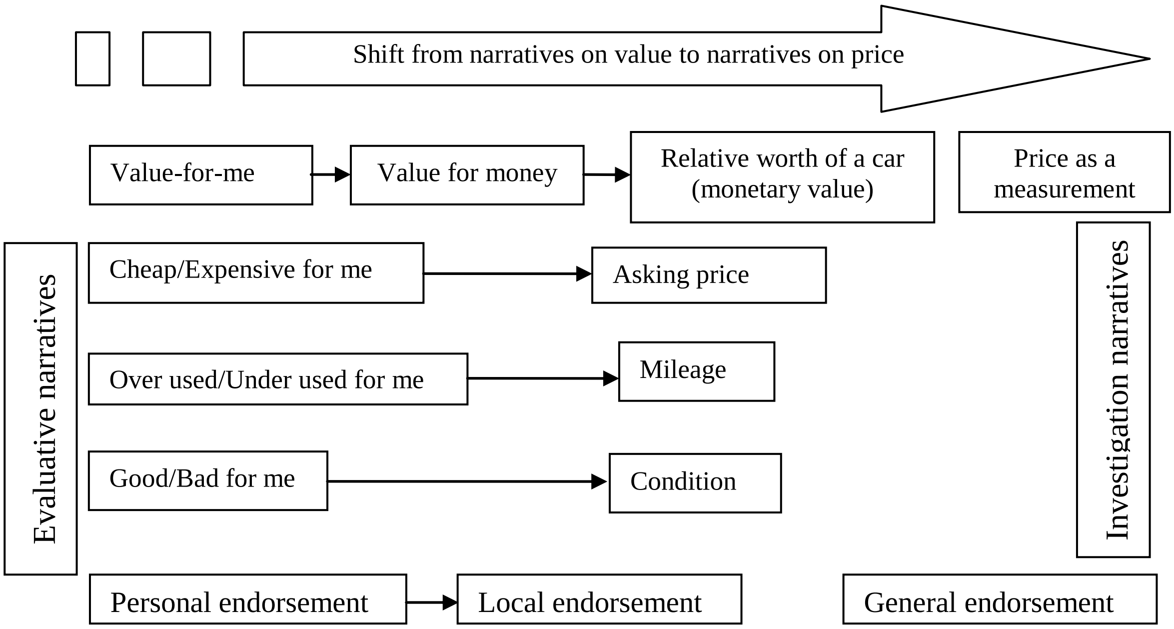


Figure 12: Shifts of narratives on value and price.

With descriptive statistics in mind as the target discourse, investigation narratives must replace evaluation narratives. Investigation narratives should be aimed at proposing relationships between *value* measured on an objective monetary scale, and other measurable properties of used cars across an aggregate. Such investigation narratives can be endorsed if *relative monetary value of cars* is operationalised in terms of mileage, year of make and model, or other measurable concomitant factors. Price must be alienated as the monetary amount that describes, rather than evaluates, the value of cars in relation to the measurable indicators of relative value. In Figure 12 I use arrows to show shifts in narratives that were evident from the discussions. By omitting arrows I signal shifts that did not occur, but that should occur in order to develop informal statistical discourse.

7.4 Summary of findings

In this chapter I decided to take one step back from the formulation of questions at the start of the statistical investigative cycle, in order to learn more about colloquial discourse in a data rich context and in relation to an implicitly data-based question. A statistician would take the question “What is a reasonable price for a used car?” as a signal to organise used cars into categories in order to compare like with like according to price. A statistician would insist on defining and operationalising *reasonable* and *price* and *used car* before gathering data. A statistician would then summarise the data and compare price data to data of other influencing variables. What they would judge as reasonable would depend on the purpose of the original question, but judgements will be based on the statistical answers. Teachers as students in an introductory statistics class are not statisticians, but they are reasonable adults who are exposed to uses of statistics in the run of life. The students in my study are mathematics teachers, familiar with working with numbers and calculations in an abstract discipline. How did these teachers go about grasping the system dynamics in order to proceed to posing investigative questions? I will summarise my conclusions with reference to my definition of informal statistical reasoning, as presented in Chapter 3.

7.4.1 Awareness of variation, sources of variation and the need for large numbers of observations

The students used price (singular) when they talked about aggregates of cars, and did not endorse the use of a “trade price” or “trade value” to smooth out differences in price between cars. Hence, I claim they were appropriately aware of variation in prices of cars in the data-context. The students realised various sources of variability in prices of used cars. These include the mileage and the condition of the car as the most salient sources. They did not proceed to compare prices objectively, and hence did not explicitly show the need for large numbers of observations.

7.4.2 Awareness of appropriate reference classes

Although I have no reason to doubt that the students would logically compare prices of cars of the same kind, I claim that subjective and largely intuitive comparisons interfered with objective comparison of prices of cars, despite awareness of appropriate reference classes. I have shown that colloquial uses of *price* and *prices*, and *value* and *values* obscured the reference classes, even in narratives that are endorsed by agents in the used car market.

7.4.3 Narratives that prevented development of informal statistical discourse

I identified a major obstacle in shifting the discourse toward informal statistical discourse. I took for granted that the students would use prices as measurements. I was wrong. The prices remained properties of particular cars. The literature prepared me for this phenomenon, but there were no explanations in the literature of why people fail to abstract measured properties. My analysis shows that the concept of *price* remained embedded in evaluative narratives that merged price with value. Value, in its turn, was related to condition as an observable property of cars, for which no objective measurement is available. I have shown that the discourse on value ranged from extreme value-for-me narratives through value-for-money narratives and eventually to relative value-of-cars narratives. Evaluative narratives conflated subjective and intuitive measurement scales (such as expensive to cheap), and objective quantitative measurement scales, so that price did not emerge as an alienated descriptive measurement. I showed that this conflation led to ontological collapse, where price gets the same status as the concrete object about which it tells a story: price seems to exist outside our discourse on it. Without price emerging as a measurable variable alienated from subjective evaluative connotations, there is no need to distribute prices on a measurement scale in order to obtain a statistical distribution.

A last note of caution: Shifting from colloquial to statistical discourse should not imply that evaluation routines should be severed from the discourse. Once statistical questions have been answered and we need to act on the basis of such answers, the evaluative questions need answering. Thus, both evaluative and exploratory discourses are valid in the statistical process. Evaluative questions have potential to provide checks and balances to compensate for the lack of certainty that statistical treatment inevitably yields in terms of practical action.

In Chapter 8, I present my analysis of the concluding discussion, aimed at grasping the system dynamics of the used car context. I will show how the students finally break free from their experiential constraints and relate price objectively to other concomitant variables.

Chapter 8: From colloquial reasoning to informal statistical reasoning: Grasping the system dynamics of the data-context

8.1 Introduction

In Chapters 6 and 7 I analysed students' contextual questions and their discussion of the questions. I argued that the questions revealed their dispositions and reasoning in the context evoked by the task question: "What is a reasonable price for a used car?" The question was ill-structured and framed in everyday discourse. The students had access to anecdotal contextual information downloaded from a blog about various aspects of the used car market as well as a dataset of a sample of used RunX cars (see Appendix A). The dataset provided numerical information of the year of make (model), the kilometre reading (mileage) and the geographical region where a car was located. Through commognitive analysis, I described the students' written questions as evaluative or exploratory questions. I found support for my classification in their narratives during discussions of the questions. I reported that evaluative narratives persisted during the discussions, to the exclusion of exploratory narratives. The evaluative narratives told stories of suspicion about the motives of agents in the used car context and imagined personal involvement as a buyer of a used car. I argue that the persistence to tell evaluative stories had the effect that the students made judgements on personal and qualitative scales (e.g. expensive versus cheap) instead of describing variable measurements on an alienated quantitative scale. I identified this lack of alienation of variable concepts as a major inhibiting aspect in terms of shifting colloquial discourse to informal statistical discourse.

As I embarked on the analysis of the discussion in this chapter, I was acutely aware of the students' persistent immersion in the data-context and the implication for development of statistical discourse. During the teaching, I was intuitively aware of this situation as well, but I did not have a theoretical explanation or an objective way to address the issue. I was as immersed in the teaching and learning context as my students were in the data-context. As I began the analysis I found theoretical support for my intuitive observations in cognitive research about text comprehension. Zwaan's (2004) construct of an immersed experiencer[57] explained the students' enacted approach to the ill-structured problem they had to solve. Zwaan and Kaschak (2009) argue that such immersion is a necessary means of re-situating oneself in order to comprehend text. My students' non-statistical reasoning was therefore an indication that they were intent on making sense of the data-context (and of what was expected of them in the course). The evaluative questions and narratives that I identified and described in Chapters 6 and 7 are therefore descriptions of the contextual immersion of my students. However, in order to develop statistical reasoning, the students must gradually emerge from their immersions and structure their comprehension of the data-context objectively and in terms of statistics. Based on research on informal statistical reasoning, I defined informal statistical reasoning as informal reasoning about everyday contexts that takes into consideration the need to base inferences on a large numbers of observations; the causal as well as random components of contexts; and the constitution of a suitable contextual aggregate (reference class) for comparison and induction. I showed in Chapters 6 and 7 that despite all these properties being realised in the students' narratives, their discourse remained non-statistical.

My task in this chapter is to map my students structuring of the data-context in terms of translation of discourses. Translation from informal discourse to informal statistical discourse will be evident when the narratives take into account the need for large numbers of observations, suitable reference classes, and sources of variation; even without explicit data at hand. A further translation toward formal statistical reasoning would be evident when the narratives are about new abstract discursive objects, such as measurements of attributes of the concrete discursive objects. Formal statistical discourse will be evident from narratives about, yet again, new abstract discursive objects like distributions, measures of centre, and measures of dispersion.

---

[57] I discussed the construct of 'The Immersed Experiencer' in Chapter 3.

I will show that shifts from immersed colloquial discourse to informal statistical discourse involve and may be dependent on argumentation and opposing views among the discursants. The discussions I analysed in this chapter occurred in the third session of the course. The full transcript is available in Appendix B.

8.2 Introduction and motivation of the task

At the end of the second session[58] in the course, I gave the class the following homework, to:

> access the website ‘www.cars4sale.co.za‘ and use the search options on the website to get a sample of 30 used Toyota RunX cars offered for sale. We want to use the sample to investigate prices of used RunX’s in order to decide on a reasonable price. Download the data onto a spreadsheet to bring to class for discussion in Session 3.

As lecturer I was aware at the end of the second session that the students’ discourse was mostly action-based and that properties of used cars were not alienated as variables that could be described quantitatively. I judged in the moment that they needed scaffolding to become explicitly aware of measurements of concomitant variables that influenced price, and of ways to structure relationships between variables. One source of such structure was available in the data-context itself, namely the interactive search options on websites that advertise used cars. Since I had sourced the data about used cars in the reading pack they received from the same website as to which I had sent them, I was aware of the structuring decisions I had made in order to obtain an appropriate dataset of RunX cars. The website has a basic search option and a detailed search option (Figure 13 below). Both search options allows a potential buyer to select the geographical region (e.g. Gauteng or Western Cape) where the car is located, the manufacturer (make) of the car and a suitable price range. The detailed search option offers more filters, such as: the engine model, the year model, and the kilometre range. The following sorting options were available: price, with photo, city, manufacturer,

[58] I analysed the questions students formulated in Session 2 in the previous chapter.

model, kilometre reading, year, model, and colour. I had two goals with using the website as a scaffold. Firstly, the students would be confronted with measurements of the properties of used cars; hence they would have numbers to compare and second, and have to make filtering and structuring decisions in order to obtain the data.

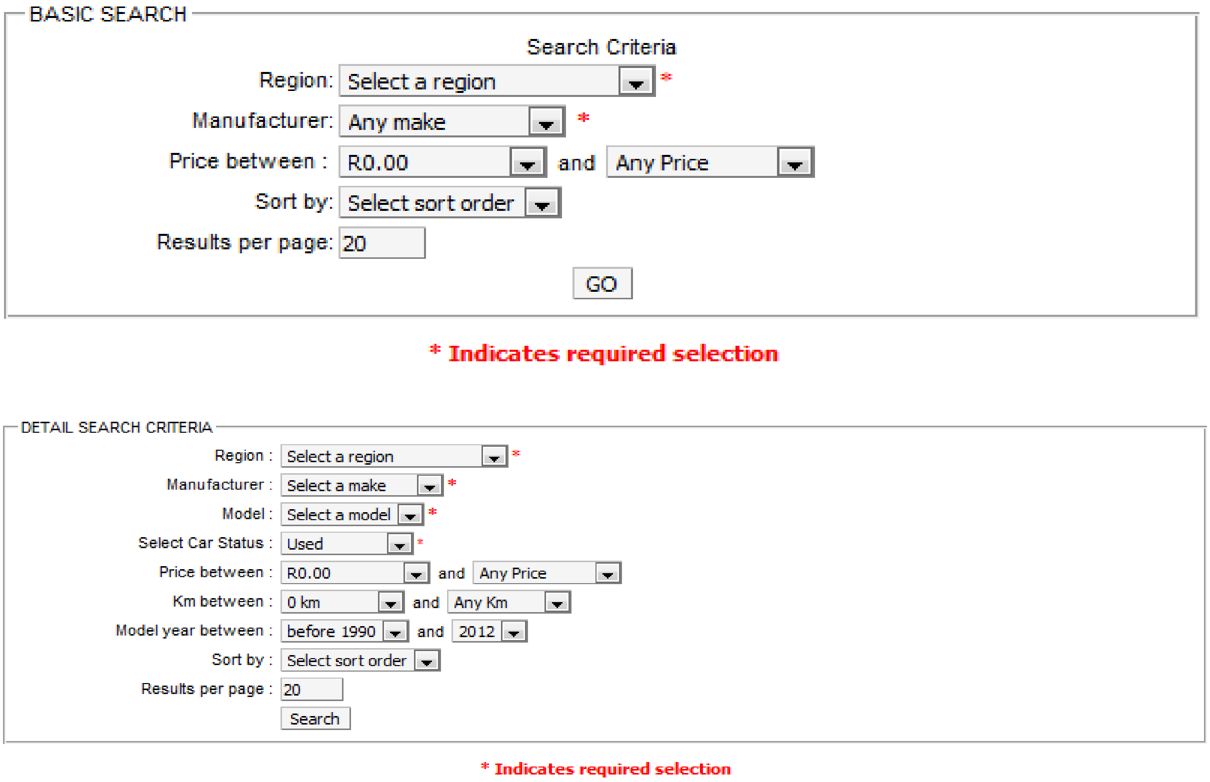


Figure 13: The search options on the website www.cars4sale.co.za.

Opening the discussion in Session 3, I referred to the epistemic properties of questions in the context of used cars, in order to confirm our epistemic goal in the forthcoming discussions. Excerpt 9 (Table 21) provides the narrative of our short interaction:

Table 21: Excerpt 9. Session 3: Whole class discussion

Excerpt 9. Session 3: Whole class discussion		
Turn	**Discursant**	**Utterance**
1	Lecturer	Thinking about second hand cars. What kinds of questions can be answered by statistics and what not? Can you give me some feedback?
2	DH	Where there are measurements involved, measures or counts.
3	KH	Trends.

4	Lecturer	We can take the used car business as example, but I'm trying for us to get to know in general the kind of questions statistics can answer.
5	KH	Specific kinds of questions, like, what is the average price for an Opel Corsa, that kind of question.
6	Lecturer	That it can answer, but can stats answer what car do *you* like?
7	GK	Uh, uh. (Disagreement).
8	Class	No.
9	Lecturer	It can't, right? Can statistics answer a question like: What is a reasonable price for a car to pay?
10	Class	(Agreement)
11	GK	Yes.

GK's confirmation (Turn 11) that the question "What is a reasonable price for a car?" could be answered by the use of statistics introduced a discussion about the students' search processes in order to obtain data from the web. The ensuing discussion formed the basis of my research task, to analyse the discursive shift from the evaluative narratives realised in Session 2, to informal descriptive narratives.

In the following section I present excerpts of four episodes from the classroom discussion in the order in which they emerged to give the reader a sense of the flow and development of the discussion. These excerpts tell the chronological story of differently structured searches for data on the website, and how the search processes prioritised contextual variables. They also allow me as a researcher to tell a story of shifting discourses, by means of the construction of realisation maps.

8.3 Mapping the discourse: realisation maps

Sfard describes the construction of realisation trees as a commognitive research task. A realisation tree is a "hierarchically organized set of all the realizations of the given signifier, together with the realizations of these realizations, as well as the realizations of these latter realizations, and so forth" (Sfard, 2008, p. 301). Realisations are produced by discursants through "a procedure that pairs a signifier with another primary object of the product of such a procedure"; while in turn, a primary object is "any perceptually accessible entity existing independently of human discourses, including things we can see, touch (material objects, pictures) as well as those that can only be heard" (Sfard, 2008, p. 301).

I focused my analysis on the students' realisations of the concept phrase *reasonable price for a used car*. Contrary to Sfard's use of realisation trees, I could not identify stable discursive objects by which students increasingly disobjectified the concept phrase. Their realisations were imagined actions, which did not lend to the formation of stable or generally endorsed hierarchies, as in the case of mathematical objects.[59] Trying to structure realisation trees of a colloquial concept like *reasonable* led me to understand how far removed colloquial discourse and literate statistics discourse are from one another. Once one has access to scientific statistics discourse, the phrase *a reasonable price for a used car* signifies a completely different object compared to everyday discourse. I understood that the statistical concept *a reasonable price for a used car* had to emerge from my students' discourse on cars in a way similar to the scientific concepts in Vygotsky's (1986) seminal research. With the scientific concept as target, framed within the obligation to answer the question by means of statistics, the students would have to go through processes described by Vygotsky as heaping, reasoning with complexes and pseudo-conceptual reasoning on the way to appropriating the statistical concept. Indeed, as the discussion developed, the students' realisations reflected increasing stages of objectification and alienation of their imagined actions. I used such stages of objectification and alienation, which I observed as organising principles for mapping the students' realisations 'from the bottom up'.

8.3.1 Coding of realisation maps

In Figure 7 (Chapter 6) I presented Vygotsky's geographical metaphor of concept development, to which I related colloquial discourse, and informal and formal literate discourses. I will use the diagram as the basis for my construction of discourse maps. My literature reviews and analysis of my students' contextual questions[60] provided evidence of colloquial discourse in a data rich context. The existence of formal statistical discourse is undisputed, and commognition theory predicts the existence of a

[59] In relation to the concept of function in mathematics, widely endorsed realisations are graphs, formulae and tables of paired numbers.
[60] See Chapters 6 and 7.

connecting discourse. The scientific concept at the apex of the concept pyramid is *reasonable* as a statistical *measure of centre*. In the discourse community of my statistics course, I was the experienced interlocutor, who made decisions about the direction of movement through the discourses. As I explained before, *reasonable* is a dictionary-endorsed colloquial realisation of *average* as a measure of centre, and I made the pedagogical decision to focus on upward movement by means of signification, from reasonable to measure of centre. It was an unexpected finding that 'reasonable' could be further realised toward the bottom of the pyramid. Vygotsky argues that not all spontaneous concepts can be developed toward the target scientific concept, therefore shifting towards informal statistical discourse inevitably requires a backgrounding of contextually useful, but statistically unproductive, narratives, such as the evaluative narratives that I described in Chapter 7.

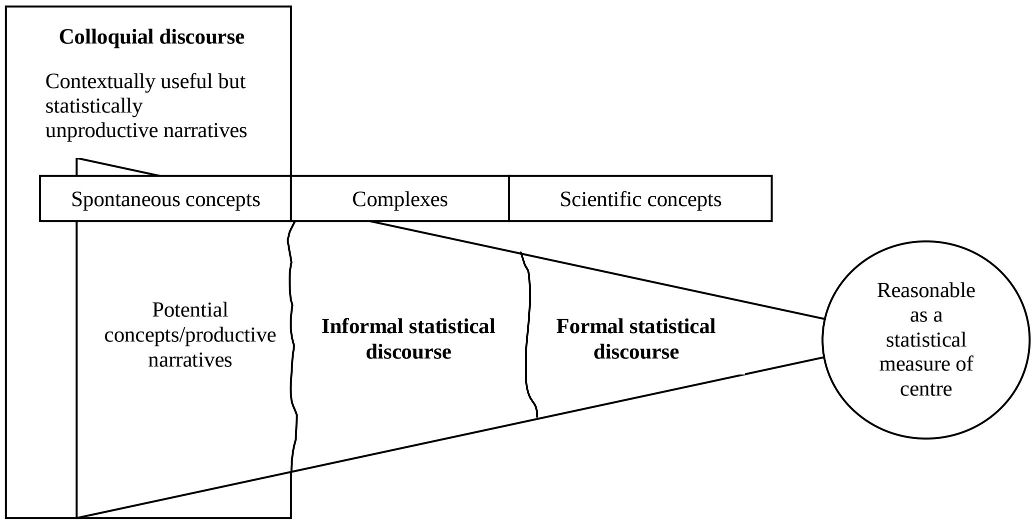


Figure 14: A spatial metaphor for shifting from colloquial discourse to formal statistical discourse

In Figure 14, I suggest permeable boundaries between the different discourses to emphasise that "discourses are in constant flux and infiltrate each other, and do not have well defined borders" (Sfard, 2008, p. 131). Often a single utterance involves more than one layer of the discourses as I defined them, almost like an eddy in the main current of a river. For example, within informal statistical discourse there might be utterances with properties of the adjacent colloquial discourse. In such cases, I recorded parts of the utterance in different discourses.

Sfard describes the analysis of discourses as a search for family resemblances, rather than detailed and over-specified properties or “universal commonalities” (2008, p. 133). I defined *colloquial discourse* as uninformed by formal statistics discourse and aimed at personal decision making or action. Realisations of imagined social relations in the data-context, such as opinions or active roles of imagined co-actors in context also received a colloquial discourse code. Lastly, consideration of single cases rather than aggregates of cases were assigned to colloquial discourse. These are instances of case-based reasoning and a local views of data (Ben-Zvi & Arcavi, 2001), which are meaningful in everyday, practical situations, but unproductive for the development of statistical reasoning.

Informal statistical discourse consists of utterances that were exemplary of colloquial reasoning that considers objective properties of concrete objects with an awareness of variability, reference classes and the need for many observations. Consequently, I coded utterances that indicated comparison between cases (such as ranking) and procedures to enable comparison (such as sequenced actions which imply ranking) as realisations of informal statistical discourse. In Figure 15, I give an example of coding an utterance between colloquial and informal statistical discourse.

Utterance: GK: "I will compare the model…of the RunX that I am looking for…thereafter I'll compare the year…that it is made. Thereafter go check on the price I want, the price that is affordable to me." (Turn 12, Excerpt 9 below)

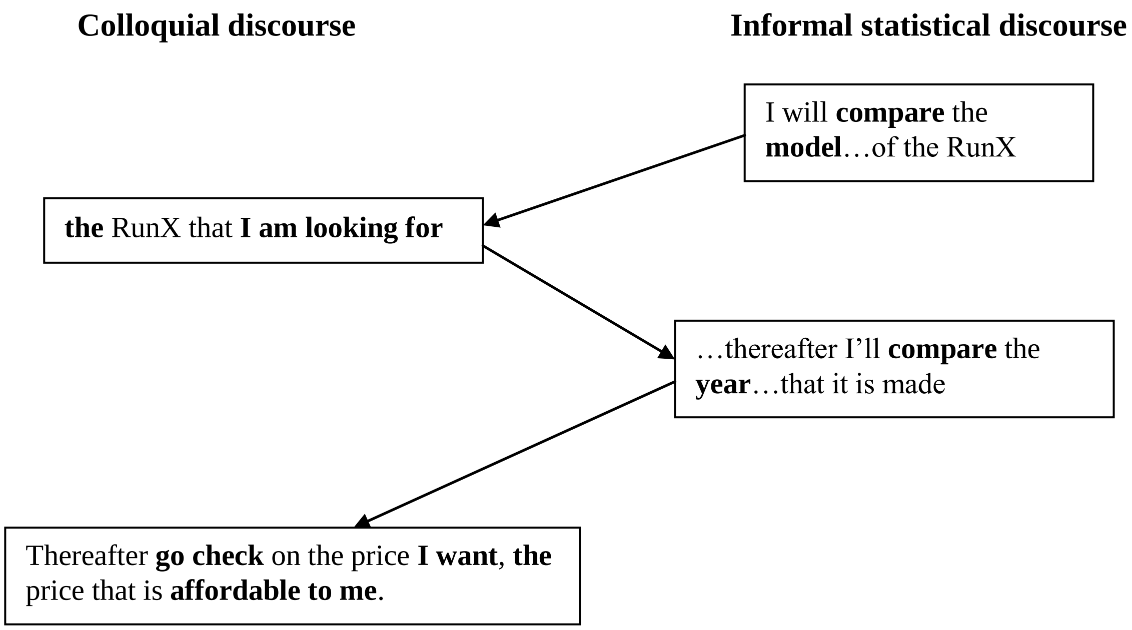


Figure 15: An example of coding an utterance in relation to colloquial and informal statistical discourse.

GK's realisations of sequential comparison of various different properties of cars were coded as informal statistical discourse, although they were realised within a larger colloquial narrative of personal decision making and case-based reasoning (she will "check out" as opposed to "compare" the price).

Literate statistical discourse as a layer of discourse mapping is theoretically viable, but I did not expect many realisations to be coded as such in the discussions at the beginning of an introductory course. Utterances that indicated exploration of alienated variables; posited relationships of variation between variables, and procedures to enable such explorations were coded as indicative of the emergence of *literate statistics discourse*. The use of standard statistical terminology was coded as literate statistical discourse only if their realisations were at least in informal statistical discourse. For example, utterances that realised a *reasonable price of a used car* as the mean or median of a sample of prices were coded as scientific statistics discourse, if they were realised by a comparison of variables.

8.3.2 Commognitive analysis: GK's tentative shift from colloquial to informal statistical discourse

In Excerpt 10, GK explains her search on the website for data to use to determine a reasonable price for a used car. It is clear that her search was aimed at making a practical decision:

Table 22: Excerpt 10. Session 3: Whole class discussion

Excerpt 10. Session 3: Whole class discussion		
Turn	**Discursant**	**Utterance**
11	Lecturer	How? What will you take as…What will you look at, GK if you want to determine what is a reasonable price for a RunX?
12	GK	I will compare the model…of the RunX that I am looking for…thereafter I'll compare the year… that it is made. Thereafter go check on the price I want, the price that is affordable for me.
13	Lecturer	If I think in terms of statistics then, I see…I imagine GK taking the RunX model she wants, a 140i…but just one. Are you going to take a look at one RunX 140i?
14	GK	Noo, I don't think it's a 140i, I take, no I will compare them, you know, I'll take… their year of make.
15	Lecturer	OK, so she's not going to take RunX all of them, you'll compare them in years?.
16	GK	Mm (agrees with me).
17	GK	And thereafter I'll check out their kilometres as we said, how far they've travelled so far…thereafter check the price of each…and maybe the colour.
18	GK	Mm (agreement).
19	Lecturer	Right?
20	GK	Then I take the one that suit[s] me best.
21	Lecturer	But the question was: 'what is a reasonable *price* to pay?'
22	GK	Mm!
23	Lecturer	So do you say your price will depend on the kilometres? What is reasonable will depend on the kilometres?
24	GK	(Looks around for support) Not just that.
25	KH	Yes it does.
26	Lecturer	You are not willing to pay the same price for a car with 100 000 on the clock and one that has 50 000 on the clock…
27	GK	Mm.

GK realises eddied narratives. The strongest current is a narrative of imagined action and evaluation, but within this narrative she realised comparisons of objective properties of cars. The mere availability on the website of different filters which yielded different aggregates of used cars, allowed her to shift her narrative from focusing on a single case to comparing groups of cars. GK emphatically denies the narrow case-based search, which I offer as substantiation of her narrative (Turns 13 and 14, Excerpt 10, Table 22); hence, comparison is a stable realisation. However, despite the deliberate and sequentially organised process of comparison, GK does not realise the properties of cars that she deems important, as alienated variables. GK uses a *specific value* of the main variable implied in the question, namely *price,* as a criterion for selection rather than as a variable. This is indicated by her use of the phrase "thereafter check the price of each" (Turn 17). GK's attending procedure is a potentially repetitive cycle of searching for the most ideal car for her price.

In GK's narrative, the closing condition was a match between the price of a car on the web and the price she can afford. That implies that her subjectively-fixed price provides a pivot for her search and for comparison of other variable properties. Since GK did not mention any numerical values in her narrative, presumably because what she could afford was not of concern for anyone else, this pivotal value did not emerge from prices of used cars as data, as should a statistical anchor such as a mean or median price.

Lastly, in GK's narrative, variable properties are not ranked objectively according to importance in relation to price, as GK is not willing to assign more importance to kilometre reading (Turns 23 and 24, Excerpt 10, Table 22) as a variable that influences the prices of used cars. Price for her depends on all the factors she mentioned, and she does not filter or structure any further. In analogy to Sfard,[61] GK is caught in her action field, which allows only narrow comparison. GK's realisation procedures (her use of the filtering and sorting options on the website) are direct in their reference to concrete objects (the car she wants at the price that suits her) and her actions as an imagined

---

[61] Sfard said we would be "captives of our visual field" if we could only talk about the things we could point to (2008, p. 111).

buyer. They are not aimed at explicitly logical relationships between variables. GK's discourse map is presented in Figure 16.

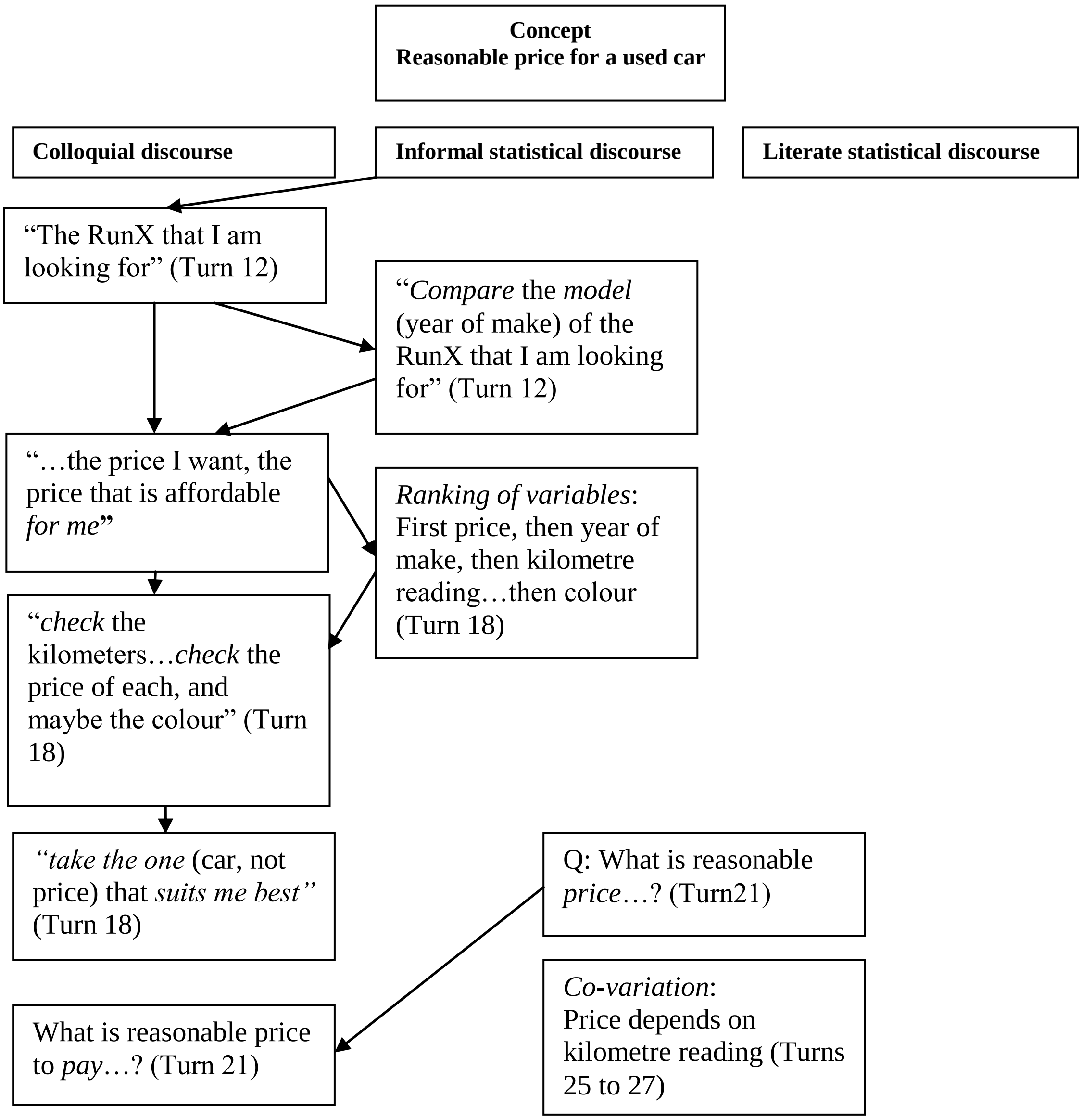


Figure 16: A discourse map for GK's narrative on 'reasonable price for a used car'[62]

I challenged the appropriateness of GK's narrative in Turn 21: "But the question was what is a reasonable *price* to pay." My challenge is also eddied: I emphasised price without reference to a specific imagined car, but follow up with an action on price. My

[62] Turns 11 to 27 in this discourse map belong to Excerpt 10, Table 22.

utterances are not taken up by GK. The discourse map in Figure 16 shows that the closest to literate statistical realisation in this discussion is my suggestion of co-variation in between kilometre reading and price.

8.3.3 Commognitive analysis: GG's informal statistical discourse on 'reasonable price of a used car'

GG proposes a different search procedure. She takes up my utterance and KH's acceptance of the causal relationship "price depends on kilometre reading" (Turns 23 to 26, Excerpt 10, Table 22). GG's discourse is also characterised by a narrative of imagined action, but case-based reasoning yielded to a more stable comparison of aggregate measurements of objective properties of cars. GG is further along the process of alienation and abstraction than GK. Excerpt 11 below shows the difference.

Table 23: Excerpt 11. Session 3: Whole class discussion

Excerpt 11. Session 3: Whole class discussion		
Turn	**Discursant**	**Utterance**
30	GG	Oh no, I was just going to say that…I will first look at the price, then decide on the kilometres. So you look at…what…at the price range you are willing to pay and then you look at the kilometres and all the other variables…(inaudible) what GK says…
31	Lecturer	OK, so, think carefully here. There are two ways that GK and GG propose to go about it. GK says I know which model I want and then I will go and see what can I get for that and what can I expect to pay. GG says I know which model I want and I have a certain budget, I can afford a certain car, a price, so I will fix the price and see what I can get for that price.
32	GG	Mm.
33	Lecturer	So these are different ways to go about it right?...Which means the relationships that you set up between your variables are a little bit different. Your classification. So, if we just think of what kind of data, how you must deal with your data to work in GG's way. See if this makes sense. GG says, I have a budget,…so say your budget is what?
34	GG	A hundred and twenty… five thousand?
35	Lecturer	(At white board) [A] maximum [of a] hundred and twenty five thousand rands. And she wants to know what she can get for that price. And now she goes to the website. What should she choose first [in setting up the search]. Do you only want a RunX?
36	GG	Yes.
37	Lecturer	OK, so we'll stick to a RunX. We can leave it…the models, still open…you might be able to get a bigger RunX for this price that's still good.

38	GG	Yes.
39	Lecturer	Now let's imagine GG goes to the website. You are going to select price range first, right?
40	GG	Mmm (Agreement).
41	Lecturer	And you're going to get [a window which asks] prices between [which range] you are willing to look at?
42	GG	Probably between hundred and thirty and hundred and twenty (gestures a range while she talks). OK so my highest is hundred and twenty five thousand, so hundred and twenty five thousand is my max then my minimum would be …. Probably a hundred thousand. I wouldn't really go for anything less than that, mm…in case, yeah, that's what I'm willing to pay, so I'll pay up to that amount, probably not less than hundred thousand.
43	Lecturer	OK. So you look for cars in the hundred thousand to, you said hundred and thirty thousand bracket.
44	GG	OK.
45	Lecturer	…which is interesting for me. You want to pay a maximum of hundred and twenty five, but you look a little higher.
46	GG	Well it all depends on the other variables that fall under it.
47	Lecturer	Uh huh?
48	GG	So that's why, if I find a car, a RunX that is hundred and twenty five thousand, and see that there is a newer model, with less mileage on it, that's a little bit more expensive, I might consider buying that one.
49	Lecturer	OK, save a bit and put it with that price 'cause it is now really a good buy for that little bit extra.
50	GG	Ja.
51	Lecturer	OK, good.
52	Lecturer	You have decided you are not going to look only for cars that cost exactly hundred and twenty five thousand rands, which makes sense. She is willing to go a bit lower to see maybe there is a bargain on this side, and a bit higher, maybe there is good value on that side. OK, can statistics tell her to do that?
53	Class	No.
54	DH	No, but you need data to do that for yourself.
55	Lecturer	Ja, but that is a good reasoning decision that comes from common sense, from knowing the world a bit. Do you agree? Statistics can't deal with it [such decision] can't help us to do that.
56	Lecturer	So before we deal with statistics, we must make good decisions about what we questions we want to ask, what ranges we want to put… to the data we are going to look at. That doesn't just jump out of the pages. So a good deal of thinking has to go into that and justify that, before you go looking at the data itself.
57	Lecturer	Right, so GG is now looking for cars. Can you imagine she's on the website, she has set the range for price at hundred to hundred and thirty thousand, and that is her first variable. Now she is going to get a *lot* of cars for that price range. Different models, different kilometre readings, different prices, different colours, sold in different places… What's the next important thing for you, GG?
58	MM	Extras.
59	Lecturer	Lots of different extras…

60	GG	Probably the kilometres, like how much mileage it has done. After that…
61	Lecturer	Why? What is your reason for now choosing kilometres, um…
62	SDS	…above colour?
63	GG	Well, colour is actually my last one. Like my order would be…first the price, then the kilometres, then the model, … and then the colour. So, I think kilometres is to see how much mileage the car has done. To see if it's…If it has been driven too much then…maybe I don't want a car that already has thirty thousand kilometres on it, maybe I want a car with only 16 000 kilometres on it.
64	Lecturer	OK, let's stop there. She said kilometres is important. Do you all agree? Why? Why do you say kilometres is important?

GK's realisation procedure suggests a different ranking of the variables than GK. GG also controls the variable *price*, but in contrast to GK, she considers a range of prices and singles out kilometre reading from "all the other variables" (Turn 30). I endorse her narrative on a range for the variable price (Turn 31) in which to compare kilometre reading and other variables, and evoke the structure of the search options on the website.

GG's explicit goal is also evaluation of price for the purpose of imagined action. Yet, her narrative has strong undercurrents of exploration of logical relationships between variables. In effect, GG is exploring what one could get for R125 000 in terms of other variables associated with the price interval R120 000 to R130 000 (Turn 42). She uses her personal maximum amount as a measure of centre for her search, if not for the data, and entertains a downward spread to R100 000. This is centre and spread in everyday reasoning, and is based on context knowledge of the used car market.[63] However, GG does not alienate price as an objective measurement, the property *price* allows her to explore value for money: she wants a *reasonable car* for her money. As I have indicated in Figure 12 in Chapter 7, this is a positive shift compared to GK's (best) value-for-me narrative. My interpretation of GG's narrative as value-for-money is supported by her assumption that she will not get her reasonable car at less than R100 000 (Turn 42). In particular, GK endows "not too expensive and not too cheap" with numerical limits. Similarly, GK's use of kilometres as an indication of relative condition supports a

[63] GG said in an interview that she had a boyfriend that was interested in cars and she enjoyed talking to him about the task.

value-for-money narrative: “kilometres is to see how much mileage the car has done. To see if it’s…if it has been driven too much” (Turn 63). Her use of kilometres in relation to the condition of a used car implies that there is an implicit contextual norm against which to judge whether a car had been driven too much. GK also endows “driven too much” with a numerical value, albeit without objective substantiation (Turn 63). I endorsed GG’s narrative regarding value-for-money as informed contextual reasoning, and thus a precursor of formal statistical reasoning (Turns 46 to 56).

The discourse map in Figure 17 shows how GG’s narrative was realised mostly in informal statistical discourse, as compared to GK’s narrative.

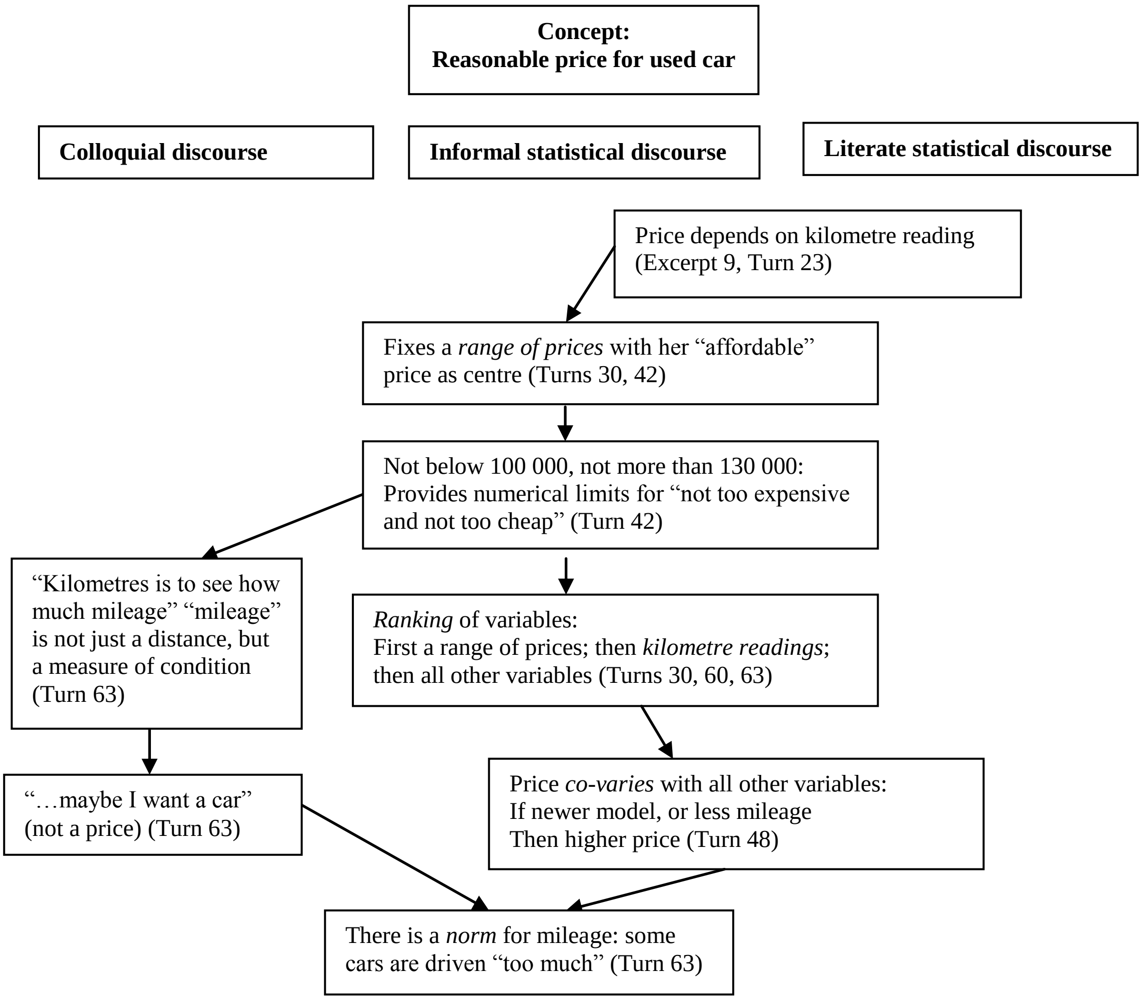


Figure 17: Discourse map for GG’s realisations about a ‘reasonable price for a used car’

GG's realisations of subjective properties like mileage as a proxy for condition, and her tentative declaration that a car with less mileage might be what she would want (Turn 63, Excerpt 11, Table 23), deserves another explanation. It is plausible to infer that her colloquial realisations were not aimed directly at action in the data-context, but served as a meta-discourse on her comparison of variable properties of used cars. She might have realised the colloquial utterances in order to substantiate the logic of her own narrative.

GG's main contribution to the shift in discourse was the introduction of a sense of co-variation of variables and accompanying tentative language: "So that's why, if I find a car, a RunX that is hundred and twenty five thousand, and see that there is a newer model, with less mileage on it, that's a little bit more expensive, I might consider buying that one" (Turn 48, Excerpt 11, Table 23). In addition, she followed up with numerical comparison of mileages, suggesting that 16 000 km might be value for money, while 30 000 km might not (Turn 63, Excerpt 11, Table 23).

At the end of Turn 64 (Excerpt 11, Table 23) two subtly different search processes were available to the class for consideration. Both search methods assigned importance to properties that influenced *price*. If the two interlocutors GK and GG were reasoning about the concept *reasonable price for a used car*, then *reasonable* was at that stage related to value-for-me and value-for-money. A further shift was needed towards relative value of cars and ways to measure value in order to structure price as an abstract numerical variable.

8.3.4 The role of dissent and evidence to enable discursive shifts

In my role as lecturer, I seized on GG's realisation that kilometre reading was an important variable to consider in relation to price (Turns 48 and 63, Excerpt 11, Table 23) in order to facilitate discussion about co-variation and dependent and independent variables. I asked the class to substantiate the relationship between price and kilometre reading: “OK let’s stop there. She [GG] said kilometres is important. Do you all agree? Why? Why do you say kilometres is important?” (Turn 64, Excerpt 11, Table 23).

Contrary to my expectation the discussion returned to practical, evaluative, colloquial discourse, as SM completely discounted the importance of kilometre reading in relation to price. However, SM’s strong voice of dissent provided a crucial change in the pace and complexity of the discussion. The norms for agreement on what constituted important factors in relation to price, required objective, genuine evidence.[64] As the interlocutors disagreed and tried to convince each other, alienation and meaningful, stable relationships between variable properties of cars emerged.

Table 24: Excerpt 12. Session 3: Whole class discussion

Excerpt 12. Session 3: Whole class discussion		
Turn	**Discursant**	**Utterance**
64	Lecturer	OK let’s stop there. She said kilometres is important. Do you all agree? Why? Why do you say kilometres important?
65	Class	(Scattered affirmative responses).
66	SM	I think it is not important.
67	Lecturer	Not Important? Why?
68	SM	No, it’s not.
69	Lecturer	Not for you? OK, let’s get the “important” argument, no let’s get the “not-important” argument. Why do you say it is not important?
70	SM	It depends on how often the car has been taken to service and then, after…how many kilometres. So kilometres, for me, is not important .
71	Class	(Silence).
72	NM	For me, it’s just a model. I feel the model is more important than…
73	Lecturer	Do you mean by ‘model’, the year of make?
74	NM	Ja.
75	GG	Oh, I was taking it as the model of the car. Ja.
76	Lecturer	The engine capacity? Let’s call that now the engine capacity.
77	Lecturer	Why, NM? Why do you take the year of make as more important than the number of kilometres on the clock?
78	NM	Eish, just a hunch.
79	GG	Maybe … it seems nicer to have a 2007 model than a 2004 model.
80	NM	Just that.
81	Lecturer	So your criteria for a car, I am ‘putting words in your mouth’ - you actually want a brand new one?
82	NM	(Nods).

[64] Kuhn (1991) defines genuine evidence by the following criteria: (a) be distinguishable from description of the causal sequence itself and (b) bear on its correctness. See Chapter 3 for a full discussion of types of evidence.

83	Lecturer	So the newer you can get it, the better?
84	GG	But you can get a nice car for hundred and thirty thousand rand if you're willing to pay for it. I am just saying. So you shouldn't actually look for second hand if you're willing to pay hundred and thirty thousand rand for a new car.
85	Lecturer	That much…
86	GG	Ja.
87	Lecturer	Do you hear what she is saying? But let's get to why you say kilometres are not important. And then again it is about what do you assume happens…with cars as they drive, and what do you assume happens when it goes for its service, 'cause you said, kilometres is not important if it goes regularly for a service, you're happy.
88	SM	Yes.
89	Lecturer	Why?
90	Class	(Silence).
91	Lecturer	Mmm?
92	MM	Do you want to say something?
93	Class	(Laughter).
94	RC	No, well I think, ah, even if you put a car in for a service, they're not gonna put in a brand new engine, so if you have done 100 000 kilometres on this engine, they're not going to replace it, it's still going to have the same wear and tear than a car that has done 100 000 kilometres. Um, I mean, they will replace the brake pads in a service, but they are not going to replace important bits that you get wear and tear…(inaudible).
95	SM	Um, that depends on the lifespan of the engine. Most of the cars, maybe, you can drive them for more than 500 [thousand] kilometres, only to find that the engine is still running very strongly.
96	RC	Yes, but…if…I mean…
97	SM	So we should not judge…
98	RC	…if most, if your car has a lifespan of 100 000 kilometres you'd rather buy one of 10 000 kilometres than 90 000 kilometres (inaudible) More life…for you.
99	RK	Just a comment. I, I think uh, from experience, I have seen some cars even that aren't strong, they don't even last as long as some second-hand cars, you know…it depends on the model, the company, I don't know…So, going for a new car doesn't necessarily mean you're doing good buying a good car. It depends on the model of the car.
100	Lecturer	I want you to listen to something: how, as we are trying to understand this (the situation), our argument shifts. Right? It's true what you are saying, in your view, but what is a good car then for that price, if we look at different models? So there is so much information we can work with, our brains can't deal with it all at once. So we have to categorize, we have to make groups. We have to say first we look at RunX, and then we are going to look at another car. But we have to do it in pieces, because our brains can't do it. Good.
101	Lecturer	So, let's say…OK, I think, I think you had a good argument there, that if you think the maximum life is measured in kilometres, then rather buy a 'baby' in terms of kilometres.

102	GG	…so that it can 'grow up' with you…
103	Lecturer	So *kilometres* is now the next one. If you look at value for money, then we must also think, what does kilometres have to do with value for money? It all has to do with our understanding of how a car gets older and what it does when a car gets older. In the end it is the moving parts that get 'wear and tear'.
104	Lecturer	Um, you have data about the depreciation of cars over years, is kilometres a factor in there?
105	GK	No.
106	Lecturer	Look at that data.
107	SM	It's not a factor.
108	Lecturer	It gives you prices: the new price, and the price a year later and a year later. Do they mention kilometres?
109	Lecturer	So if kilometres is an important one, what have they done with kilometres to compile that table, what do you think?
110	NM	I just want to say… the way I think about the car…it is as if I'm looking at a washing machine, somehow in my head (roll hands like wheels rolling)… so I'm thinking, if the washing machine can still do the job (roll hands energetically) still wash the clothes, even if it's old, but it can still do the job… I am somehow struggling to understand why… it doesn't matter how many washes the washing machine has done, but in future it can still do the job, washing that washing. So I'm also looking at the car in that way. So what matters is it can still move, it can still take me from one place to the next place, as much as the washing machine can still wash those clothes.
111	SDS	I think the important thing is for how long? If you get a 'baby car', it will last you much longer than time than an old car, where its gonna reach that end at some stage.
112	NM	Mm.
113	DH	Example of someone else's washing machine 16 years old that broke with the washing in it.
114	Lecturer	We can accept [the] machinery needs service, and it's parts does get wear and tear. There comes a point where it is just too expensive to still keep this thing on the road.

SM's realisations in Excerpt 12 (especially Turn 70) signal a return to the complexity of the data-context. His response to my request for justification of his claim that kilometre reading is not important, is a causal claim of correspondence with an external factor. According to Kuhn (1991) SM provided genuine, but weak evidence in the form of a counter-factual argument. He proposes that in reality, the demise of increasing kilometre readings is countered by good care. His position suggests that he holds an extreme position on the new variable (which I will call "condition") that he introduced: service makes a car as good as new and completely erases the effect of distance done. SM's evidence is discounted by RC (Turn 94). In response SM provides additional counter-factual evidence: he introduces the concept "life span" (Turn 95) which implies a fixed time limit presumably 'built into' the car.

SM is not willing to judge the condition of a used car by its kilometre reading. RC argues that "life span" can be measured in kilometres and creates a numerical scale for the imagined comparison of kilometre reading (Turn 98). Her position is met with RK's counter-factual evidence of quality differences between cars, which introduces yet another contextual factor, namely the *model*[65] of the car (Turn 99).

According to Kuhn (1991), the argument between SM, RC and RK is sophisticated in the way they use counter-arguments, and indeed the raised complexity was not accessible for all the students. In an interlude of non-evidence, NM fell back on her "feeling" and "hunch" that "it's just a model" that is important in relation to price (Turns 72 and 78).

The discussion in Excerpt 12 (Table 24) is a meta-level narrative of substantiation. The students are vying for endorsement of their arguments by the whole class, but the evidence offered was weak and complicated, rather than a well delineated argument. Nobody wants to make a decisive statement. Yet, the counter-factual arguments are utilised to argue about correspondence between properties of cars in general. Slowly a global view of the data-context emerges through the interlocutors' references to "most of the cars" (Turn 95), "some cars" (Turn99), or unspecified subgroups of cars, through reference to their "model". When RC refers to "your car" (Turn 95) she doesn't signify a single car, but rather a hypothetical car with that typical property. The narratives are shifting to informal statistical discourse. In this argument kilometre reading emerged as an objective measureable proxy of wear-and-tear (Turn 94) and age (Turn 102), and numerical values were used to exemplify intuitive extreme values, rather than central values.

[65] The word model was used inconsistently in the discussions to refer to year of make, manufacturer, or engine capacity. I assume model as referring to the manufacturer here, since it makes more sense to compare a RunX of a given engine capacity to a Volkswagen or BMW of the same capacity, than to compare it to another RunX with a different engine capacity or year of make.

Causal analysis of argumentation to establish priority of variables in context

In Chapter 3 I described Cummin's (1995) theory that informal causal deduction is influenced by the identification of alternative causes, that question the necessity of a cause to produce a given effect; and disabling conditions that question the sufficiency of the given cause. Structuring the discussion in Excerpt 12 in terms of alternative causes and disabling conditions revealed how individual narratives about simple linear causal relationships between variables, gave way to co-structuring of the system relationships of the data-context in a productive way. The argument started with the contested relationship between kilometre reading and the price of used cars established by GG in Excerpt 11. The relationship implied the following causal statement:

Introductory causal line:

If a car has a high kilometre reading it will have a lower price (than a car with a low kilometre reading)

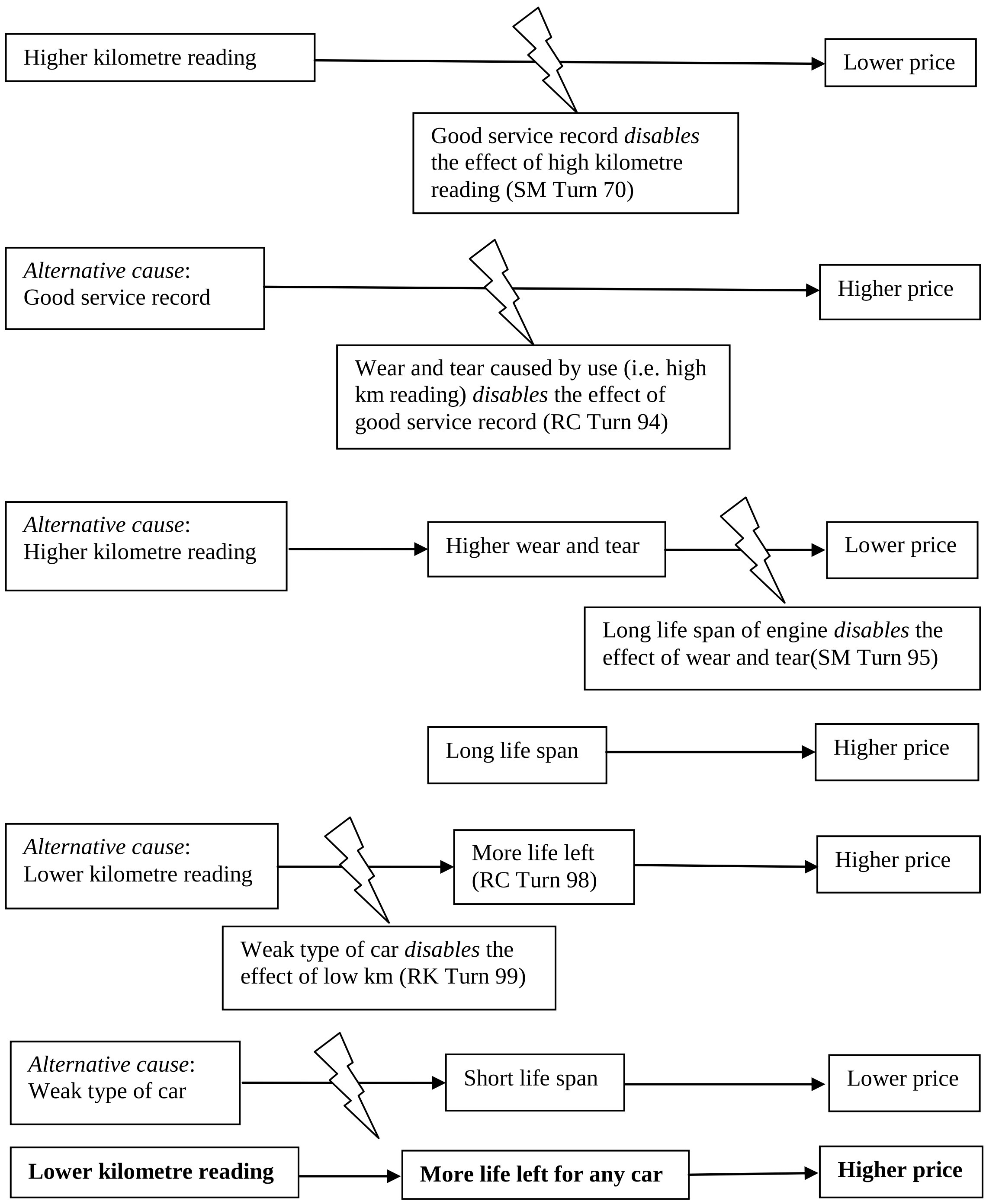


Figure 18: Causal analysis of the argument about the importance of kilometre reading in relation to the price of a used car

Concluding causal line:

> If a car has a lower kilometre reading it will have a higher price (than a car with a higher kilometre reading).

It is not important that the concluding causal statement was in the negative, because it is essentially the same as the introductory statement. Without analysing the discourse that came between, a researcher or teacher would have no way of deducing the commognitive shift that supported the concluding statement. At this stage in the endeavour, to grasp the system dynamics (and bring the statistical object *reasonable price* into being), the discourse had shifted from any one student's personal, contextual decision making to co-structuring of the factors that influence the value of used cars. The argument illustrates the struggle to relate variables unambiguously as the contextual understanding of what happens to cars in their lifetime developed. Where the *age* and *condition* of a used car were important factors related to *reasonable price* from the onset, at the introduction of the argument in Excerpt 11 it was not clear how age or condition were defined and logically related. At the end of the argument, kilometre reading emerged as an endorsed proxy for age and condition.

Ontological collapse

Not all students considered the discussion about sources of variability of prices of a used car as useful. For at least two students, discursive shifts were in the opposite direction that that of the other discursants. I will explain their plight as ontological collapse leading to (presently) unresolved commognitive conflict. Sfard defined ontological collapse as [the] phenomenon of taking all the objects that are being talked about–discursive (d-objects) as well as extradiscursive (p-objects)[66]–as belonging to the same ontological category of "things in the world" that pre-exist discourse, with their mutual relations similarly "objective" and "mind independent" (Sfard, 2008, p. 300).

---

[66] See Chapter 4 for a discussion of ontological collapse and primary and discursive objects.

SM's ontological collapse

During the discussion in Session 2, SM declared that buying a car made no sense, considering various costs and depreciation (see Excerpt 5, Session 2 in Chapter 7). SM's dissent in Session 3 was again based on an aspect of cost he had mentioned in Session 2, namely service. SM's contributions to the discussion in Session 3 were solely to strengthen his original opinion, that buying a car made no sense: he argued that kilometre reading is not important, since regular service restores its condition (Turn 79,Excerpt 11), and since services are expensive; a good car will last longer (Turn 95, Excerpt 11), but good cars are expensive and they depreciate in value; kilometre reading is not explicit in the depreciation table, and is therefore not a factor (Turn 106, 107, Excerpt 11). The flow of SM's discourse was against that of the other discursants and must be explained. I propose that SM's extreme evaluation discourse was incommensurate with the discourse of exploration of the relationships between price and other concomitant factors. In his discourse, properties of cars were never abstracted and remained aspects about which he had to make practical decisions. It seemed that for SM, the condition of a car was "mind independent" (Sfard, 2008, p. 300). For him, it was possible to see condition and fix condition with a service, rather than define condition in terms of kilometre reading or another measurable property.

NM's ontological collapse

In order to tell the story of NM's ontological collapse, I present two excerpts. Excerpt 12 is taken from the end of the whole class discussion in Session 2.[67] Excerpt 13 is taken from the end of the whole class discussion in Session 3.

At the end of Session 2, NM interrupted the discussion of the questions with a challenge to the rest of the discursants, to justify why we need statistics at all to answer our questions.

[67] The full transcript of Session 2 is available in Appendix B.

Table 25: Excerpt 13. Session 2: Whole class discussion

Excerpt 13. Session 2: Whole class discussion		
Turn	**Discursant**	**Utterance**
130	NM	It now sounds…it now sounds, those questions, we are forcing the stats.
…	…	…
147	DH	It's asking you to make a comparison. It doesn't ask what the cost of maintenance was, but, how do you make a comparison for it, unless you compare it to something else.
148	NM	'Cause all I see. I am seeing numbers here. I am seeing numbers from one garage or wherever, and I'm also seeing numbers, yes, it's clear, I can see it with just financial maths. Now that part about stats, I need an example a specific example, it's just...
…	…	…
152	GK	Can I just say what I am thinking. I am thinking that…no…what we're comparing, even if we use financial math to get the figures and so on. In the end you want to see what is this figures tell us, how can we interpret them. And then presently there we are talking about the maintenance of this make this make of car, let's say, what ah, Renault, ah, at the maybe at the age…after five years…ah…if I took it for maintenance it will cost me this much, after five years, after two years, you get what I'm saying, after each year, or maybe two years you service it. And from there you compare…you compare the prices of the … (lots of weighing gestures).
153	NM	I'm looking at the numbers.
154	GK	And you see what they tell you. Do they tell you that this is the best car to buy, because maintenance wise it's cheaper? Or is it more expensive?
155	NM	All I am saying is I can get the answer form just looking at those numbers. Done. What else can I do?
156	GK	You want to buy a better car? One that won't cost you more?
157	NM	Yes. The answers are just there. After doing financial maths, I just get the numbers there. This is more expensive, this is less.

For NM, the numbers in the data table hold the answers to her questions. If any calculations had to be done, and the numbers represented money, she would merely be doing financial mathematics. She is aware of variation: "this is more expensive, this is less" (Turn 157, Excerpt 13), but as evident from her earlier utterances, variation does not matter to her.

NM rebuffed the argument about the importance of kilometre reading, saying "For me, it's just a model" (Turn 72, Excerpt 12, Table 24). GK and NM earlier indicated that they would look for the single "best" car in terms of year of make since they felt it was good to have a new car (Turns 72 and 79, Excerpt 12) or even a better make of car (in case that was the meaning of model as used by NM). In Excerpt 14 (Table 26 below), NM does not even hold on to the desire for a new car. She falls back on an extreme binary view of the data-context and declares that a car is like a washing machine (or for both a car and a washing machine), in the sense that all that matters is if it works or not (Turn 110, Excerpt 14).

Table 26: Excerpt 14. Session 3: Whole class discussion

Excerpt 14. Session 3: Whole class discussion		
Turn	**Discursant**	**Utterance**
104	Lecturer	You have data about the depreciation of cars over years, is kilometres a factor in there?
105	GK	No.
106	Lecturer	Look at that data.
107	SM	It's not a factor.
108	Lecturer	It gives you prices. The new price, and the price a year later, and a year later. Do they mention kilometres?
109	Lecturer	So if kilometres is an important one, what have they done with kilometres to compile that table, what do you think?
110	NM	I just want to say… the way I think about the car…it is as if I'm looking at a washing machine, somehow in my head (roll hands like wheels rolling)… so I'm thinking, if the washing machine can still do the job (roll hands energetically) still wash the clothes, even if it's old, but it can still do the job…I am somehow struggling to understand why…it doesn't matter how many washes the washing machine has done, but in future it can still do the job, washing that washing. So I'm also looking at the car in that way. So what matters is it can still move, it can still take me from one place to the next place, as much as the washing machine can still wash those clothes.
111	SDS	I think the important thing is for how long? If you get a baby car, it will last you much longer than time than an old car, where its gonna reach that end at some stage.
112	NM	Mm.
113	DH	Example of someone else's washing machine 16 years old that broke with the washing in it.
114	Lecturer	We can accept machinery needs service, and its parts does get wear and tear, do get wear and tear. There comes a point where it is just too expensive to still keep this thing on the road.

NM seems to give up in the face of the emerging complexity of the data-context. She seems not able to analyse the situation in order to make reasoned decisions about the prices of cars on any other grounds than personal preference. In a discourse on price that was constituted by evaluative and action-based narratives, price-for-me seemed to be a pivot around which to distribute the price of a hypothetical reasonable car. In such a discourse, the numbers tell one at a glance what one wants to know: whether a price is cheap or expensive, distributed on a subjective evaluation scale. NM's position also signals ontological collapse.

8.4 Discussion: Repeating discursive patterns

The discussions I analysed in this chapter conclude the students' efforts to grasp the system dynamics of the used car context.[68] The discussions to structure the data-context started with their formulation of questions and our discussion of those questions in Session 2. With the added information from my analysis of the discussions in Session 3 I identified discursive patterns that were repeated across the discussions.

According to Sfard (2008, p. 301) a commognitive routine is a set of meta-rules defining a discursive pattern that repeats itself in certain types of situations; this set can be divided into three subsets: applicability conditions, routine course of action (or routine procedure), and closing conditions (or closure).

The applicability conditions of routines are subsets of meta-rules that delineate *when* a routine is appropriate; while the routine procedures indicate *how* the routine is to be performed. Identification of routines is a research task, based on inferences about the discursants' purposes in discussions. The same procedure or utterance may have different applicability conditions and different purposes for discursants and hence they will be performing different routines.

I propose that the narratives to structure the data-context revealed a social-evaluation routine; an action evaluation routine; and an investigative routine. Commognition theory classifies routines, according to their meta-rules, into deeds, explorations and rituals.[69] The closing condition, or the goal, of a deed routine is a change in objects. In

[68] In the course the teaching and learning tasks proceeded to statistical descriptions and comparisons of data sets by means of measures of centre and measures of spread.
[69] I discussed the different kinds of commognitive routines in Chapter 4.

the case of a ritual routine, the closing condition is alignment with others. Exploration routines aim to produce endorsed narratives about a context (Sfard, 2008). Sfard exemplified deed, ritual and exploration routines from a mathematical discourse perspective, and her definitions must be interpreted from a perspective of statistical reasoning. I will attempt an interpretation and base this on insights from my literature review in order to avoid circular reasoning.

Statistical reasoning is an invitation to explore variability in contexts, rather than perform a practical action. A statistical exploration routine, with the purpose of producing endorsed narratives about a data-context, would be based on a routine course of action that identifies contextual variables and relationships between them, defines measurement units and reference classes, proceeds to gather and investigate data, and finally, considers the relevance of statistical findings in the data-context. This is the routine described by the model of the Dimensions of Statistical Inquiry (Wild & Pfannkuch, 1999, p. 226), which I presented in Figure 1 in Chapter 1. Within this general interpretation of statistical investigation routines, the invitation to consider what would be a reasonable price for a used car implied an investigation of varying prices for used cars. Grasping system dynamics implies asking both 'why' and 'how' questions: why do prices vary and how do they vary? Statistically, the question *how* would suggest establishing an appropriate reference class, gathering a representative sample of prices and distributing them on a measurement line in order to describe the distribution by shape, centre and spread. The *why* question would suggest hypothesising relationships between contextual variables and embarking on the statistical investigation of co-variation of two variables, one of which is postulated as the cause and the other as the effect. Indeed, as a lecturer I expected the students to propose an investigative procedure like "Let's look at prices of RunX's on the web and see what the average price is" or "the price depends on […] so we will have to find an "average" for a relationship between variables." Yet, such a routine did not emerge from the discussions I analysed.

With statistics in mind, a deed routine, with the purpose of changing some extra-discursive object or situation, may plausibly be seen as making decisions that can be

described as statistical misconceptions. Such deeds will be evident from persistence to reason deterministically in the face of variation, making decisions based on unclear reference classes and subjective, personal experience as a sample of one observation.
Statistical rituals, or repeated discursive narratives to attain social approval or alignment, were described by Gigerenzer (2004, p. 587) as “mindless” statistics and exemplified with evidence of habitual, uncritical use of the null hypothesis with a 5% significance level. In the statistics classroom, rituals may be evident from narrow, rule-based use of statistical measures, without having recourse to their contextual implications.

Based on my attempt at interpreting routines from a statistical discourse perspective, I will now describe and classify the routines I identified through my analysis of the classroom discussions in sessions two and three.

8.4.1 Social evaluation as a deed routine

I have shown in Chapter 6 that forty percent of the questions contributed by the students told stories about efforts to evaluate the motivation of agents in the used-car context, such as used car dealers, statisticians and officials who construct and use market standards. The prevailing assumption was that such agents cannot be trusted, and that market tools cannot be trusted, in the event of buying a used car. Issues of trust continued to be of importance in the discussion of the questions, as reported in Chapter 7. Hence I propose these narratives indicated a deed routine[70]:

> *Social evaluation routine*: repeated narratives about imagined motives of social agents in a data-context, aimed at safeguarding oneself in the event of becoming an actor in the context.
>
> a) *Applicability condition:* The data-context signals interaction with social agents.
>
> b) *Routine course of action:* Assume social agents have questionable and selfish motives and devise ways to safeguard one against such agents.

---

[70] A deed routine is a repeated discursive pattern for which the closing condition is a change in objects (Sfard, 2008, p.297).

c) *Closing condition*: When the system relations of the context are understood sufficiently to allow the immersed experiencers choices and options to safeguard themselves, closure is reached.

8.4.2 Action evaluation as a deed routine

Repetitive action evaluation narratives were described in Chapter 7 and Chapter 8. In such narratives, the students told stories about the prudence of buying a used car, or stories that indicate processes to ensure prudent action in the data-context. I propose that evaluative narratives about imagined prudential action in the data-context are deed routines:

Action evaluation routine: repeated narratives about imagined prudent action in the data-context.

a) *Applicability condition:* Immersion in the data-context signals practical decision-making and action.
b) *Routine course of action:* Discursively enact the required actions and procedures.
c) *Closing condition*. When the enacted procedures have resolved imagined practical problems or have lead to prudent action, closure is reached.

Social evaluation routines and action evaluation routines are deeds, since their purpose is to achieve an imagined practical goal of safeguarding the discursant against infamous agents, or to ensure prudent actions.

8.4.3 Contextual investigation as an exploration routine

An exploration routine is a repeated discursive pattern with the goal to produce an endorsed narrative about the context (Sfard, 2008). Sixty percent of the contextual questions formulated by the students in Session 2 were exploratory questions. These questions were aimed at understanding the context in terms of real properties of cars and relationships between those properties. Narratives based on exploratory questions were limited in the ensuing discussion (see Chapter 7), but in this chapter, I showed how the narratives based on GK and GG's internet searches for data increasingly turned to exploring relationships between price and other concomitant contextual variables.

Hence, I propose that the repeated narratives about relationships between properties of cars constitute an exploration routine:

Context investigation routine. Repeated narratives about proposed relationships between objective properties of cars.

a) *Applicability condition:* Discursants' individual structuring of relationships between contextual variables produces competing narratives in need of substantiation.
b) *Routine course of action:* Restructure the competing narratives logically through argumentation.
c) *Closing condition*: When the restructured narrative is endorsed by the discursants, closure is reached.

The routines I have identified were at times produced by individual discursants and at other times emerged from the collective discussion, such as the co-structuring of the relationship between price and kilometre reading in Excerpt 11 (Table 23), and represented in Figure 18 in this chapter.

8.4.4 Implications of deed and exploration routines for the development of statistical discourse

The deed routines that I described as social evaluation routines and action evaluation routines were unproductive for shifting colloquial discourse to informal statistical discourse. Social evaluation routines are not indicative of any of the properties of informal statistical reasoning, and therefore fall outside the concept triangle I used as the metaphor for shifting discourses (See Figure 14 in this chapter). The value of action evaluation routines was also limited in terms of shifting the discourse. In Chapter 7 I argued that word use in action evaluation routines remained subjective. For example, price of cars and value of cars were evaluated on a personal scale of value-for-me. The value-for-me scale ranged from cheap to expensive, with a reasonable price somewhere in between as determined by personal affordability. Similarly, *mileage* signified the condition of a car, rather than an objective measurement of the distance a car had been driven. This routine prevented alienation of the contextual variables; they remained subjectively defined by human judgement.

The students' decisions about filtering of contextual variables on the website in order to obtain data were the focus of the discussion in Session 3. The narratives about the two search procedures (of GK and GG) were emerging personal investigation routines. Although initially aimed at action (looking for a car to buy), they told competing stories about the relationship between price and other variables. Through the argumentation about these stories, *mileage* as *kilometre reading* had been elevated to the most important variable in relation to prices of used cars, and numerical values were realised as markers against which to compare cars (See Excerpts 10 and 11 in Tables 22 and 23), while other important variables like the year model and the manufacturer model could now be used as reference classes in which to aggregate data.

The need for structuring the context would not have been evident from the data sets that the students downloaded. Every student had a spreadsheet with 30 cars and information about their geographic location, mileage, model, year of make, colour and even the "extras" they possessed. Their data sets were structurally identical to the dataset I provided in the reading pack in Session 2. Only on closer inspection and cued by the classroom discussion did I notice the consequential restrictions in some of the data sets. For example, that GK's data set consisted of cars made between 2000 and 2007, and GG's dataset consisted of cars priced between R100 000 and R130 000.

8.5 Summary

At the end of this chapter, I reflected on the discussions about the data-context that I analysed in this chapter, as well as in Chapters 6 and 7. With the benefit of a meta-view of these analyses, I identified repeating patterns in their narratives, which and I then defined as routines. Telling my story about the classroom discussion in the language of these routines enables me to propose relationships between the routines and the shift in discourse.

In this chapter I mapped the discourse in an extended classroom discussion that was aimed at grasping the system relationships in the context of the prices of used cars. The discussion ensued from the students' search processes on a website to obtain data about used cars. I compared two search procedures, both structured according to personal action goals, but different in their explication of relationships between price and other variables. I mapped the narratives of their search processes mainly onto two layers of discourse, namely colloquial discourse and informal statistical discourse. The discourse maps shows that GK's deed routine, which I called an *action evaluation routine*, constrained her narrative on the relationship between price and other variables to colloquial discourse. The closing condition for her routine was to find a car that she liked and that she could afford. She structured the relationship between price and other variables subjectively. In comparison, I showed that GG's emergent exploration routine, aimed at ranking price related variables logically, shifted her discourse toward informal statistical reasoning. She considered a range of prices, which indicated shift away from a local view of the context. Yet, her proposed range of prices was still structured subjectively around an amount she could afford. Contrary to GK's deed routine, GK's exploration routine allowed her to realise numerical ranges against which to evaluate cars. In particular, "not too expensive and not too cheap" received numerical limits in GK's narrative.

I have also shown how conflicting structures of the relationships between price-related variables evoked argumentation that allowed the discursants to logically restructure such relationships. Causal analysis of the argument in Excerpt 11 suggested that the intuitive everyday reasoning tendency of looking for disabling conditions or alternative causes (Cummins, 1995) supported the co-structuring of the context.

Amidst the complexity of the discussion, extreme evaluation routines led to ontological collapse, as in the cases of SM (Excerpt 12, Table 24) and NM (Excerpts 13 and 14, Tables 25 and 26). In their discourses, abstract consideration of relationships between variables made no sense, since that is not what they would do in practice. The extremely contextual purposes of SM and NM's routines eventually constrained their agency in the classroom discussion that shifted to objective relationships between contextual properties of cars.

Without the benefit of the deep discussion of the context of used cars, I would have been no wiser about the students' evaluation routines in their colloquial discourse and constrained shift toward statistical discourse. I would have been unaware that measurements might hold for them a meaning on a personal scale rather than on a numerical scale.

The question may well be asked whether the purposefully introduced complexity of the data-context and the task formulation did not lead to these specific routines. Would these routines have emerged if the focus was on statistical measures, rather than the structure of the data-context? My narrative in the next chapter is aimed at answering this question as I analyse the students' discourse on the statistical concept *mean*.

Chapter 9: Data analysis: The mean-ing of the mean

9.1 Introduction

In this chapter I report on my students' discourse on a decidedly statistical object, the arithmetic mean. Groth and Bergner (2006, p. 60) call for research to help university instructors better understand how to overcome "interference with conceptual learning that may be caused by previously built procedural knowledge" of the mean. Yet, not only limited procedural knowledge interferes with conceptual understanding of the mean as a statistical object, everyday conceptions of *average* also influence students' understanding of the statistical mean. In literate statistical discourse the statistical average is a balance point and a sensitive measure of central tendency of a set of data. In colloquial discourse the statistical mean is equated with a notion of *average* which is usually explained as "somewhere in the middle", "not too much and not too little". I suspected that my students' understanding of mean as average would likely rest on a subjective qualitative evaluation scale, similar to their reasoning about the concept *value*[71] of a car. Hence, their colloquial judgements of average may be vague qualitative anchors based on personal experience and constraints, rather than objective relationships between measurements as numbers. Such a qualitative judgement scale may be in use even where numerical data is available to execute the calculation that produces the statistical mean. Similar to numerous other studies, the students in my course knew how to calculate the mean. Yet, in Sessions 2 and 3 only one student (SDS) suggested finding the average, or the mean, as a strategy to determine a reasonable price for a used car, but her suggestion was not adopted by the rest of the class. The students therefore did not associate the statistical mean of a set of prices with the colloquial term *reasonable* price. After the discussion in Session 3, on which I

[71] See Chapter 7.

reported in Chapter 8, I introduced the mean and median as representative measurements of price through their positions on a graph of prices of used cars. Reflecting on the introduction of the mean, I wrote in my field notes:

> [I] Suppose I should have expected that they will not use the mean as a measure of centre without being told to calculate the mean. Enough literature suggests it. I introduced the mean and median with the graph. [I] named and samed median with a point they suggested they would quote if asked to use one number to describe the price of used RunXs in the sample. [We] Calculated the mean and [I] showed how it is influenced by the tail of prices and therefore less representative of this data. But it seemed I was forcing this on them, from the stats side. What is in between average and mean? Why do they not need a representative value from everyday reasoning side???? How does the calculation manage to produce a representative value anyway? Let me ask them!!! (Journal entry, 18 August 2008)

Since they knew how to calculate the mean, I decided to challenge the students' understanding of the mean algorithm. I wanted us to develop a disobjectified narrative about the mean algorithm and a way to connect their realisations of average with the calculation of the mean. Watson and Kelly (2003, p. 2) emphasise the literacy aspect of statistical literacy which is even more imperative for teachers of statistics:

> Statistics requires the basic understanding of statistical concepts, such as random, whereas literacy requires the ability to express that understanding in words not in mathematical formulas. If statistical literacy is to have any prospect of attracting attention from the multiliteracies community, then it must demonstrate an expectation for literacy skills within the field. To do this there must be an expectation that students can describe the meaning of the concepts at the foundation of statistical thinking.

Current perspectives on understanding the mean draw on distinctions between conceptual and procedural knowledge (Groth & Bergner, 2006). Conceptual knowledge is rich in relationships and represents a connected web of knowledge, including relevant procedures, which is constantly realigned and reconfigured (Hiebert & Carpenter, 1992;

Hiebert & Lefevre, 1986). Several different avenues have been followed by statistics education researchers to determine evidence of conceptual understanding of the mean. Some researchers accessed understanding of the mean from the everyday concept of average; others accepted the mean algorithm as given knowledge and researched how the number returned by the calculation is interpreted in context and in relation to data distributions. My research focus was to understand the syntactical meaning of the algorithm itself, hence the reason why addition and division provide a representative measure of variable numbers.

9.2 Discourses on the mean as evident from research literature

Guided by the research literature,[72] I expected the students in my course to realise the mean as *average*, and *average* as *middle* in various ways. In a discussion about the meaning of the mean algorithm, the mean would again serve as the scientific concept at the apex of my developmental pyramid. The base of the pyramid would consist of colloquial narratives about the mean, with *average* likely as a leading realisation. Conceptual movement towards the apex of the pyramid would necessitate saming, objectification and reification. Reflecting on the literature, it seems as if various uses of 'middle' were samed to become the root of *average*. The most pervasive uses of 'middle' described in the literature are *half above and half below,* and *most*. These uses were mapped by researchers on the literate statistical concepts median and mode respectively. It seems that differentiation is needed to develop mean as a literate concept, rather than saming. Downward movement from literate discourse about the mean to its colloquial roots would require disobjectification. I theorised that disobjectification should aim at constructing narratives about 'middle' that tell different stories about middle than the median and mode realisations. Hence, I challenged the students to construct realisations of the mean algorithm that can explain in what way the mean is in the middle of a data set. In Figure 19 I use Vygotsky's (1986) geographic metaphor again, this time to suggest that the bridging discourse, informal statistical discourse about the mean, had to be constructed in the class discussions.

[72] I reviewed the literature about understanding the mean in Chapter 2.

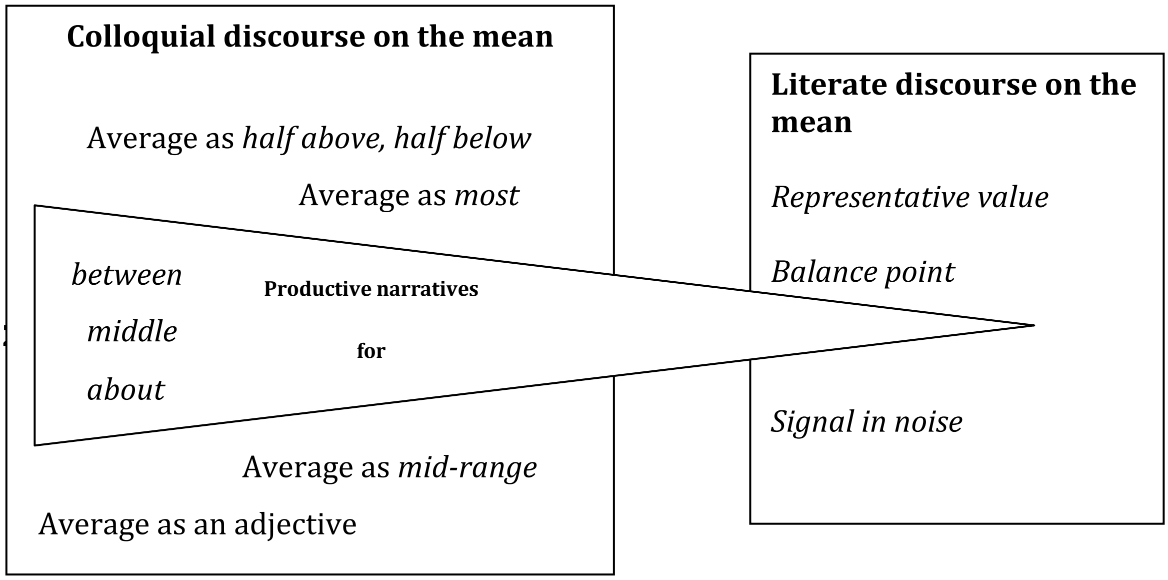


Figure 19: Delineation of discourses on the mean reported in statistics education research

Relevant for my study is Bakker's (2004a, p. 55) reference to Aristotle's notion of the "the mean of the thing" and "the mean relative to us". Aristotle described the mean of the thing as "the same for everybody" while "the mean relative to us" is not the same for everybody, but characterized by one's interpretation of "neither too much nor too little". With this difference Aristotle sanctions the idea that personal factors in context may lead to different qualitative interpretations of the same number as measurement of some property. In Bakker's study as well as in mine, the idea of the mean as neither too much, nor too little was expressed as a self-evident meaning of *average*, which I will argue is a different discursive object than the *mean*.

Since Mokros and Russell's (1995) study, confirmations abound that the arithmetic mean is strongly associated with *average* in colloquial discourse. Average in turn is associated with other aspects such as 'most'; 'large amounts'; 'in between'; 'about'; 'roughly'; 'half above and half below', 'a bit in balance', and 'midpoint' (Bakker, 2004a, p. 97; Makar, 2004; Watson & Moritz, 2000). These associations are evident across ages and education levels. Mokros and Russell refer in their 1995 paper to similar research with similar findings with adults, while the studies by Bakker (2004b) and Watson and Moritz (2000) were done with middle school students. Jacobbe investigated three elementary school teachers' understanding of average and median, and reports their understanding of mean as "average as an algorithm". The meaning these teachers assigned to the mean is average and average is 'the usual' (Jacobbe, 2008, p. 3).

Reading transcripts across a variety of studies where participants were asked what they thought was meant by the *mean*, focused my attention on the almost standard reaction along the lines of "it is just the average" after which they might elaborate on what they mean by average (Groth & Bergner, 2006; Mathews & Clark, 2007; Mokros & Russell, 1995). From a discourse interpretation perspective, Bakker (2004a) agrees that it is not clear how to interpret students' utterances about properties of mean or average such as 'midpoint' or 'middle'. Do they refer to the midrange, the middle-most number or to a point such as a centre of gravity? Even when interpreting students' accompanying gestures the researcher cannot be sure. The interpretive possibilities are provided by proposing mappings between students' colloquial utterances and the discursive statistical objects.

9.2.1 Subjective colloquial discourse on the mean

As I indicated in Chapter 2, idiosyncratic responses about the mean and average populate the prestructural, unistructural as well as the multistructural levels of Watson and Moritz's (2000) SOLO based model of understanding the mean. The levels in the model are structured according to the number of objective elements of the mean realised by participants in their interpretation of the term average or the calculated mean in context. Yet, a commognitive perspective hints at the existence of different discourses, rather than only lack of elements of the mean in responses.

Free responses to the question "If someone said you were average, what would it mean?" from participants of different ages yielded personally meaningful responses like the following:

a) *Prestructural*: You are not the best friend for me (Grade 5)

b) *Unistructural*: That you were okay (Grade 5)
c) *Multistructural*: Not really good and not really bad; in between (Grade 9)
d) *Relational*: Because it [the media] shows a fair representation of the [house] prices. If the average was used, a particularly cheap or expensive house would muck up the fair representation (Grade 9)

(Watson & Moritz, 2000, p. 15)

As discussed in Chapter 2, Watson and Moritz map these responses at different levels of their SOLO-based model. Yet, together, these examples of realisations of *average* as an adjective in context all reflect a judgement on a subjective scale. The prestructural response could plausibly form part of a social evaluation narrative such as "If you think I am only average, you are *not the best* friend for me." It could also be interpreted as "if I said you are average, then I mean you are *not the best* friend for me, only average." Indeed without access to further elaboration by the participant, these are only plausible discursive interpretations. Assuming that average is realised in contrast to *best*, whatever the specific intention of the respondent, indicates a personal qualitative measurement of average on a scale where *best* exceeds *average*. Similarly, the responses at the unistructural and multistructural levels may be more elaborate and perhaps more alienated since they did not refer explicitly to specific contextual properties (like friend), but they are indicative of subjective evaluation scales, rather than objective comparison of measurements of properties. Hence they belong to colloquial discourse and might be realisations of social evaluation routines. Clearly the relational response belongs to literate statistics discourse. The narrative included a meaningful disobjectification of the abstract discursive objects mean and median. Endorsed properties of mean and median were realised as well as a colloquial verification that used examples of primary discursive objects (prices of houses) to exemplify the literate narrative.

Watson and Moritz do not report on possible further realisations of the concept "fair" and therefore the point is moot whether fair is judged by comparison of objective prices or on an overlapping personal scale of cheap to expensive. Indeed, Mokros and Russell (1995, p. 31) comment on the initially egocentric and undeveloped reasoning even of students who have representative and relational understanding of mean.

The discourses that I mapped onto the pre-, uni- and multistructural responses and on the relational response respectively, imply that participants performing social or practical evaluation routines in colloquial discourse, would not understand participants in literate statistical discourse. Similarly participants of literate statistical discourse may not understand the subjective meanings attached to the mean. To argue the point beyond that of the framing role of tasks and questions, a participant of literate statistics discourse is likely to ask a counter question like "average in what?" when posed the same question "If someone says you are average what would it mean?"

9.2.2 Objective colloquial discourse on the mean

Meaningful applications of the mean are not intuitive and not easy to develop. One such application is the comparison of distributions. Makar (2004, p. 118) reports that students who did use the mean to compare data sets used it exclusively, not relating the mean to other statistical properties of the data sets (e.g. the shapes of the distributions or measures of spread). Contrary to Makar's expectation the same students did not develop an interpretive narrative in relation to the concomitant contextual factors they had discussed in class. Students who did realise contextually interpreted narratives [73] expanded their descriptions by remarking on *how much* the enrichment group had improved, rather than just reporting that the enrichment group had improved. Such quantitative comparison between groups sounds promising. However these students' interpretation of the amount of improvement was not based on comparison of the means, but rather on the difference between the higher outliers of the two distributions. The mean values were therefore used to represent the data sets and the participants compared the data sets by values that are *not* in the *middle*. This asymmetry in the comparison of groups is widespread and begs an explanation. Why do participants who disobjectify mean as 'average as middle' not use the middle value to compare data sets? I propose a plausible explanation from a commognitive perspective: In Makar's (2004) study the task was to describe and compare *improvement* of groups. In everyday

[73] The context of comparing student performance between a group that attended an enrichment course and a group that did not (K. Makar, 2004).

discourse improvement is evident from *better* performance of students in each of the groups, and surely an intervention would not be expected to yield negative results. It is plausible that the participants in this research used the ideal, namely the *best improvement* as a judgement of real improvement. Practical wisdom plausibly suggests that if you are not good, any practice will help you to improve, but if you are already good and you still improve the program must be good – ideally the best should also improve. Hence they could have used the measures of the best[74] students for comparison. Commognitively, the students might have solved a different discursive task than intended by the researcher.

Idiosyncratic narratives from participants in statistic education research are indications of the divide between colloquial discourse and literate statistical discourse in a data-context. Despite colloquial use of noun phrases like 'the mean is in the middle', the mean is realised as an adjective in contextual application, in routinely subjective narratives. Statistical measures (e.g. mean, median) as signifiers in literate statistical discourse seem to frame narratives strongly toward abstract reporting, where the comparison of distributions by their means is a routine procedure and where a statement like "2.3 children per family" is assumed to speak for itself. Alongside this literate discourse the competing colloquial discourse remains characterised by attempts to make practical sense by mapping the numerical value of the mean onto concrete objects (e.g. .5 of a family is a child); or where typicality among variable measures is judged by reference to individual, extreme cases or by frequencies such as the number of individual cases that actually improved.

The problem of how teachers should talk about the mean is exacerbated by the paucity of formal language in definitions of *mean.* In general dictionaries as well as statistical dictionaries the definitions are usually circular, caught between a notion of average and the mean algorithm. For example, Collins Dictionary of Statistics (Porkess, 2004) defines mean and average as follows:

---

[74] In Chapter 3 I reviewed Rein, Goldwater and Markman's (2010) notion of an *ideal* as typical of a rule-governed category like *diet*.

Mean, 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” (Porkess, 2004, p. 50).

(The dictionary then proceeds to give the mean algorithm in algebraic notation).

Average. 1. In technical use, average usually has the same meaning as mean or arithmetic mean.

…

3. 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 (Porkess, 2004, p. 14)

As a statistical definition Porkess’ definition of the 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. It may suggest to the colloquial ear that the average value actually exists as an object in a context independent of its measurement. The process definition, provided by the algorithm then explains how to find this object.

Merriam-Webster Online Dictionary does not provide a definition for mean as a statistical object (it does provide the standard process definition for the arithmetic mean). However, this dictionary is more descriptive in its use of average in relation to mean:

Average:

1 a: a single value (as mean, mode or median) that summarizes or represents the general significance of a set of unequal values

2 a: an estimation of or approximation to an arithmetic mean

b: a level (as of intelligence) typical of a group, class or series < above *the average*>

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.

The Merriam-Webster definition of average assigns formal statistical properties of the mean to average, such as representative, typical, or a ratio of which the denominator is "the number of opportunities for successful performance." Both dictionaries implicitly use mean and average as synonyms, perhaps with a little leeway for rounding and estimation. While Bakker (2004a) indicates that the hypothetical nature of the mean and its role in error theory is an important concept which relates the mean to the statistical notion of spread as dispersion from a centre, these dictionary definitions do not make the hypothetical nature of the mean explicit.

9.3 Literate discourses on the mean

Alternative semantics for the *mean* that focuses on its statistical use and properties can be found, such as *mean as a measure of central tendency* or mean as a *balance point of a distribution.* These phrases are mostly encountered in statistics education texts and endeavour to provide cues for conceptual understanding of the statistical object, rather than the colloquial object *average*. It is rare to obtain such descriptions of the statistical mean from participants in statistics education (Watson & Moritz, 2000). Groth and Bergner (2006) report that only one out of 46 pre-service teachers related the mean as average to a balance point and only one related mean as average to equal distribution of data (as in the sense of fair sharing). In terms of making sense of the mean algorithm Mokros and Russell (1995) argue that division by the number of cases in a data set as a fair share calculation does not lead to enhanced understanding of the result of the calculation. They suggest focusing on evening-out strategies[75] and balancing deviations around a given mean as promising didactic approaches.

9.4 Historical discourses on the syntax of the mean algorithm

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

[75] Evening-out strategies refer to informal methods to allocate differences between measurements between the measured cases so that all cases end up with equal measurements. Balancing deviations refer to strategies to ensure that differences between the mean and points above the mean balance differences between the mean and points below the mean.

mathematics, the harmonic, geometric and arithmetic means. Statistical use of the mean can only be traced back to the nineteenth century. In this section, I will only refer to historical development of the mean algorithm. Bakker (2004a) 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 'algorithms' provide insight in the uses and therefore the concepts that underpinned *average*.

9.4.1 The mean as a multiplicative unit

Bakker (2004a) gives two historical examples where one representative value was used multiplicatively to estimate a large total number. In the first example[76] 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[77]. In these early historical examples, the term average doesn't appear, rather, the method or process of calculating some practical quantity was described in words. Both these examples involve practical reasoning in order to achieve a practically verifiable goal. The goal was to determine a measurement for a primary object. Hence, in commognitive terms, this pre-concept of the mean was developed to act on simple discursive (concrete) objects rather than abstract discursive objects. Bakker (2004a) interprets the relevance of these examples as incorporating notions of the arithmetic mean in relation to representativeness. In order to estimate the totals in the examples, the mean had to exist as a number (hence the estimates) and the mean algorithm would be transformed to

(A representative object) x (number of objects) = total number of objects

It is important to note that in this historical use the mean was not unknown or hypothetical. 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 easily available conceptual step to

[76] An ancient Indian story reported by Hacking (1975).
[77] From the history of the Peloponnesian War. See Bakker (2003).

‘creating’ the mean by equal sharing. It does not make practical sense to ask how wide a brick must be[78] to build a wall of a given height with a given number of rows. The concept of the mean as a multiplicative unit is used in statistics to report data such as the number of cell phones per 1000 inhabitants in a country. This use is not recognised intuitively as involving the mean, and is used to report on numbers that can be read and compared easily, rather than an actual multiplicative unit. Bakker indicates that he found no indication of an early discourse on equal sharing of objects that mapped onto average as a concept.

9.4.2 The mean as a geometric object

The geometric concepts of arithmetic, geometric and harmonic means existed long before the statistical concept of mean. Bakker (2004b) explored the Greek notion of average as average lengths, that is, a concept related to iconic realisations of simple discursive objects[79]. The formula that was used to express this average or middle length of two lengths a and c was

$$\text{“}a - b = b - c\text{”}.$$

In this equation, all three objects existed symbolically and are represented such that it is clear that the mean length 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. Bakker (2004b, p. 56) cites the theorem of Pappus where the arithmetic mean, the geometric mean and the harmonic mean of two line segments were indicated in a single construction. The construction placed the two line segments AB and BC as extensions of each other, so that the combined length was a + c and forms the diameter of a circle. Hence the arithmetic mean was half of the diameter (the total length), which is the radius.

---

[78] In practice bricks are made to standard specifications, while walls vary in height.
[79] Lengths were constructed with the use of compasses and straight edges and hence treated as concrete objects (to the extent that numerical discourse on square root lengths was problematic).

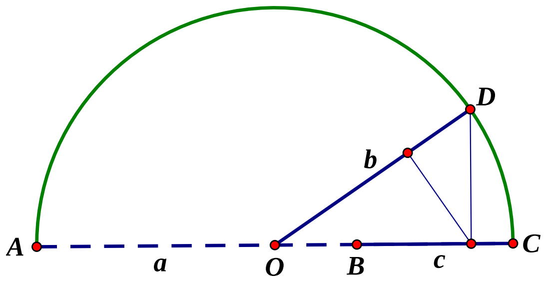


Figure 20: Theorem of Pappus. OD is the arithmetic mean of AB and BC (Bakker, 2004b, p. 56).

Through the construction of Pappus the arithmetic mean existed as an object with a measurable length. The construction belonged to literate geometric discourse, since it used the tools and axioms of Euclidian geometry. As in the previous examples of the mean as a multiplicative unit, the shift from a geometric concept of the arithmetic mean to the statistical concept of representative value or balance point of a data set is not immediately clear. Only in the sixteenth century and possibly enabled by the development of the decimal system, was the arithmetic mean generalised to more than two cases. Even the generalised mean was mostly used to approximate "a real or best value … [such as] the diameter of the moon" and only in the late eighteenth century was the mean used as "the representative value for an aspect of a population" (Bakker, 2004b, p. 60).

9.4.3 The mean algorithm as a tool for eliminating differences

Bakker (2004b) identifies other possible early historical precursors of the mean, namely processes to eliminate error from measurements. These uses of mean supports the suggestion that the mean concept belonged to practical discourses in widely differing contexts, aimed at measuring properties of concrete objects.

Astronomers of the sixteenth century endeavoured to determine the positions of the moon and the planets in relation to the earth. Their world view was such that they believed the planets were placed in position by a divine hand. The goal of their measurements was therefore to determine the *true* positions. Variable measurements were seen as approximations and caused by measurement error. The concept of variation and a possible distribution of measurements were not available in early astronomy discourse and the mean was used to get as close as possible to "a real or best value" for a measurement of a physical object (Bakker, 2004b, p. 60). In current statistics education discourse the concept of the mean as "a signal in noise" (Masnick, et al., 2007, p. 25) is used to convey a similar idea, although not with the intention to evoke physical interpretation of the mean as an object. Using the mean as a measure of true value persists in quality control contexts. In this modern use the objects that the mean represent are members of a relational category. For example, bread baked to a mass specification. Hence the mean does not measure the true mass of a loaf of bread in a sample, but the extent to which the baker was true to the specification – deviation from the mean is seen as unfair practice.

The first historical moment where the relationship of the mean to *differences* between measurements, rather than to the measurements themselves was foregrounded may provide the first meaningful disobjectification of the syntax of the mean algorithm. Bakker (2003, p. 9) describes the use of the term "averij" in ancient maritime law (around 700BC), where the purpose of calculating the "averij" was to share losses at sea fairly between customers. In the event of a storm at sea cargo was thrown overboard to save the ship. The question was how to share the individual losses incurred for the benefit of all. Bakker (2003) reports that the mean algorithm was employed for this purpose, but how exactly it helped to share losses fairly, remained opaque. Examples of "averij" calculations are restricted to the "averij" of two amounts at a time. Indeed, only in the sixteenth century was the mean algorithm extended to n cases.

I pondered possible disobjectification narratives about average loss and came to the conclusion that this historical use of "averij" is an important example of a discourse about varying *differences* rather than varying initial amounts. Averaging losses involves a process of sharing in order to eliminate *differences* between losses, so that all have effectively lost the same amount. However, my initial discursive access to sharing in this imaginary context was not sharing by division, but rather by addition and subtraction. I drew the following sketch to think through the process of reaching the "averij":

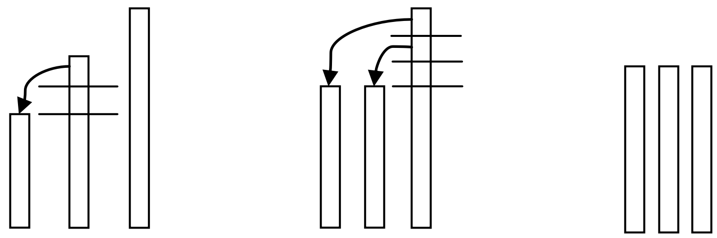


Figure 21: Averaging differences additively

First I ordered the losses (represented by the bars). I then averaged the smallest and the second smallest loss by taking away half of the difference between the losses and adding it to the smallest loss.

Then the difference between the biggest loss and the two equal losses is shared equally among all three bars to achieve the average loss. The losses need not be in any specific order, although ordering is heuristically helpful. Algebraically, modelling the action discourse closely, averaging proceeds as follows:

Let $x_1 < x_2 < x_3$ be losses incurred.

Then average the difference between x_1 and x_2:

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

$= \frac{x_1 + x_2}{2}$ $= \frac{x_1 + x_2}{2}$

Now average the difference between the biggest loss and the equal losses:

Each of the equal losses is increased by $\frac{1}{3}\left(x_3 - \frac{x_1 + x_2}{2}\right)$

The averaged losses then are

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

The largest initial loss has also been reduced to the same amount as the other losses, and hence $\frac{x_1 + x_2 + x_3}{3}$ is the average loss.

I shared the losses of more than two imaginary merchants, proceeding in pairs, without targeting an estimated average, or a true value that somehow pre-existed the discourse.

With hindsight the abstracted discourse of the “averij” as the fair share of the combined *losses* is unproblematic. However, the shift in my reasoning was enabled by adopting a narrative about *additive* sharing of *differences* between the losses rather than the losses themselves. Bakker (2004a) relates young students’ levelling out narratives to find an average length of many bars that represented life spans of batteries[80] to the Greek understanding of the average length of two line segments, rather than to the “averij” process. In the battery problem, students worked with a computer tool that allowed them to place a vertical value bar across the horizontal bars that represented the life spans of two samples of batteries. The placement of the value bar was based on a subjective value (100 hours is a good life-span) or on the midrange of the life spans of one brand of battery (the shortest life span was about 80 hours and longest life span was about 120 hours). I assume the compensation strategy to level out bar lengths was also pair-wise in its execution, but since the bars that represented the life spans were not ordered by length, systematic compensation and abstraction of the mean algorithm was not a likely outcome of the task. The contrast between my practical reasoning and the reasoning of the young students on the battery problem, lies in the fact that I constructed the mean algorithm as an object, while the young students constructed an estimated position on a graph.

9.5 Variation discourse on the mean

The use of mean in discourse on variation developed late in history. The term *l’homme moyen*, the average man, was invented by the Belgian statistician Quetelet (1796-1874). According to Bakker (2003, p. 12) Quetelet was one of the first scientists to use the mean as “the representative value for an aspect of a population”. Quetelet’s use of the mean as a representative value rather than a real value as in astronomy, was an important step in the development of variation discourse on the mean. Bakker (2004a) cites Charles Peirce, mathematician and philosopher, who wrote in 1877 how problematic it was to map continuity of measurement onto situations where

[80] For a full analysis of the batteries problem, see Cobb (1999).

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”. Peirce preferred “most men” instead of “the average man” (Bakker, 2004a, p. 61).

Statistics education researchers concur that statistical understanding of the mean is contingent on understanding variation and distribution. In general participants in statistics education describe mean or average as ‘middle’ or ‘centre’ in ways that can be related to statistical concepts like the arithmetic mean, the mid-range, the mode and the median, but far less often to the midpoint as a centre of gravity (Watson & Moritz, 2000). When graphs are used as visual mediators, participants tend to use colloquial descriptive terms such as clump or hill to indicate which collection of values they consider as average or typical, rather than a specific value. As Bakker (2003) has shown, historical pre-cursors of the mean as arithmetic mean, the mid-range, the mode and the median had practical action-based uses, while mean as a centre of gravity is clearly a highly abstract discursive use. The divide between colloquial and literate statistical discourse on the mean remain un-bridged.

9.6 Problematising the shift from colloquial to informal statistical discourse on the mean

Discourse on the representativeness of average in context does not easily map onto discourse on the statistical mean. Even though we are now used to expressions like ‘the average man’ and in my course we easily spoke of the price of the ‘average car’ the term is still very abstract and the meaning of the number returned by the average is difficult to interpret. Bakker (2004a, p. 99) reports that only a few students in his study understood what was meant by “average family size of 2.5” while many interpreted .5 as a child, i.e. something like half a family member. On the other hand they understood the meaning of an average of 1.5 hours TV time. The didactic recommendation that Bakker draws from this problem is to differentiate between continuous and discrete measurements. However, the difficulty of working with numerical quantities that are hard to visualize, is not particular to understanding the mean, and not a novel observation for a mathematics teacher. For young children, mathematically dividing 2 litres of cooldrink between 5 children is harder than dividing 2 chocolate bars between 5 children.

From a commognitive perspective I suggest that the mathematical nature of the numbers as continuous or discrete is not the only problem. What the numbers represent is a more pressing discursive issue. In my narrative about the meaning and process of “averij” the average amount referred to a primary object such as an amount of money that would be counted. The mean in the sense of the average person or car, has no such practical logic connotations. The statistical mean is a hypothetical measure, a very different concept from the ancient “averij” even though the algorithms are identical.

By this stage in my deliberation of discourses on the mean, the concept ‘middle’ seems to have potential to provide a bridge between the colloquial concept *average* and the statistical concept *mean*. In colloquial discourse, average as middle is likely to refer to a typical concrete object (for example an average car) for which some attribute measurement is judged as not extreme. In literate statistical discourse three measures of middle may act as measures of central tendency of a distribution, namely the arithmetic mean, the median and the mode. Of these, the median and the mode can be disobjectified easily. For example, one can imagine standing in the middle of a row of people in the median sense. If one stands in the median position and the people on either side are untypically different from one self, it would be difficult to accept oneself as representative of the group. The mode as a colloquial ‘middle’ has more potential than the median for a discourse on representativeness.

I will now report on my commognitive analysis of the class discussions of the meaning of the mean algorithm. I will show that disobjectification of the mean reached back to colloquial action and evaluation routines, where the notion of middle was paramount. I will argue that the use of *most* in the students’ narratives were not realisations of the mode, but rather informal realisations of an acceptable, typicality preserving, deviation interval around a central value. The central value remained the vague notion of average in context, but discussion about deviation from the value returned by the mean calculation was productive in order to define the mean as a hypothetical object.

9.7 Introduction of the task

In order to give my students the opportunity to reflect on the concept of the mean and to afford myself the opportunity to learn how my students think about this well known literate concept, I gave the following written prompt for discussion in the fourth session of the course:

Some things to reflect on:

> What is the logic/common sense behind using the mean as a measure of centre? What is the meaning of the standard deviation on a common sense level? That is, can you motivate the use of the SD? Refer to the common sense of the formula and what the computation at each step means.[81]

At the time I posed these questions, the class had already discussed the mean and the median in relation to graphs of symmetric and skewed distributions. They were able to calculate both measures of centre correctly and accepted the rule of thumb to use the median as a measure of centre for skewed distributions, since the mean is affected by extreme values.

In the rest of this section I analyse the video transcripts of group discussions and the subsequent whole class discussion.

9.8 Commognitive analysis of the disobjectification of the mean algorithm

As in the previous chapter I will map the students' realisations onto three discourses:

a) *Colloquial discourse* consists of narratives aimed at practical action and subjective decision making in context. The objects of colloquial discourse are images of simple discursive objects and their observable properties.
b) *Informal statistical discourse* consists of narratives aimed at exploring the numerical properties of the mean. The objects of informal statistical discourse are collections of measurements alienated as abstract numbers.

[81] We did not reach discussion of the standard deviation in the same session in the course. I will not report the discussion on standard deviation here.

c) *Literate statistical discourse* is substantiated by the larger community of users of statistics. Formal terminology is used and properties of the mean as a measure of central tendency are logically related to the mean algorithm.

9.9 Ontological collapse: the mean is the average

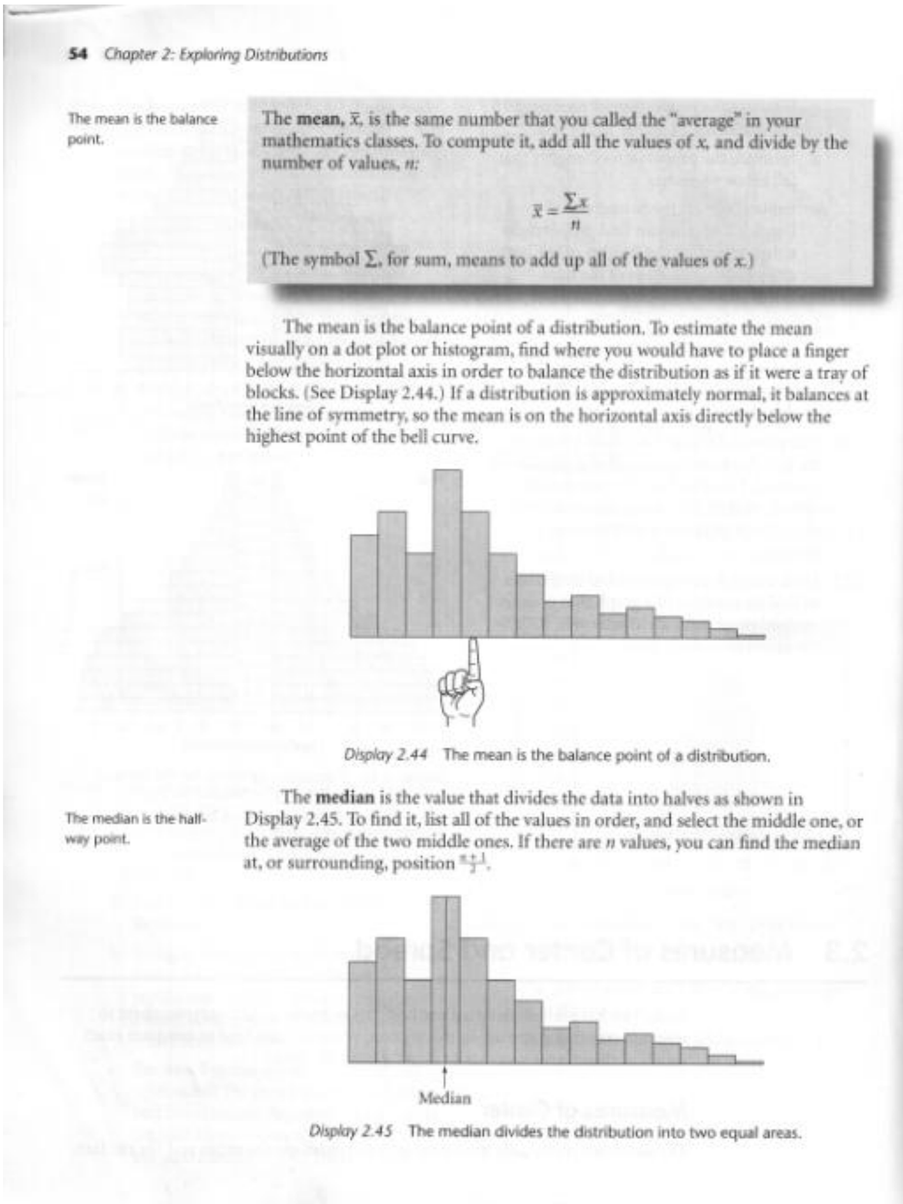
As suggested by the literature, the students approached the task of finding the logic behind the mean by describing the calculation “ritual” (Sfard, 2008, p. 255) that would generate the mean, rather than to explore the syntax of the algorithm. I call their leading realisation a ritual since it is automated as social knowledge, done because others sanctioned it. Once they had explained how the mean is calculated the students seemed to experience closure. I had to provide a further prompt to set the process of disobjectification of the mean in motion. In the excerpt in Table 27[82] I present the ensuing discussion among Group A.

Table 27: Excerpt 1. Session 4: Discussion of Group A

Excerpt 1. Session 4: Discussion of Group A. Discursants KH,RK, GK, SM		
Turn	**Discursant**	**Utterance**
1	KH	So the first question is why do we use the mean as a measure of centre, is that right?
2	RK	Mmm.
3	Lecturer	We all know how to calculate the mean, I am reacting to what I hear... We all know how to calculate the mean. I want the common sense meaning of it, why the heck do we do it?
4	KH	(Referring to the textbook) It says here, the mean is the balance point of the distribution, it balances...the mean’s the balance point.
6	KH	We have to get the common sense here.
7	SM	Allright, we can say that it 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...
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)
16	RK	Mm

[82] The full transcript of the discussions in Session 4 is available is Appendix D.

When the verbal description of the mean algorithm was not endorsed by the lecturer (Turn 3, Excerpt 1, Table 27), the students' next recourse was to try and find an endorsed narrative in the textbook. KH (Turn 4) finds a realisation of mean as a balance point in the textbook (Figure 22), but acknowledges that she cannot find common sense in this realisation (Turns 4 to 6, Excerpt 1, Table 27).


54 *Chapter 2: Exploring Distributions*

The mean is the balance point.

The **mean,** $\bar{x}$, is the same number that you called the "average" in your mathematics classes. To compute it, add all the values of x, and divide by the number of values, n:

$$\bar{x} = \frac{\sum x}{n}$$

(The symbol $\sum$, for sum, means to add up all of the values of x.)

The mean is the balance point of a distribution. To estimate the mean visually on a dot plot or histogram, find where you would have to place a finger below the horizontal axis in order to balance the distribution as if it were a tray of blocks. (See Display 2.44.) If a distribution is approximately normal, it balances at the line of symmetry, so the mean is on the horizontal axis directly below the highest point of the bell curve.

Display 2.44 The mean is the balance point of a distribution.

The median is the half-way point.

The **median** is the value that divides the data into halves as shown in Display 2.45. To find it, list all of the values in order, and select the middle one, or the average of the two middle ones. If there are n values, you can find the median at, or surrounding, position $\frac{n+1}{2}$.

Display 2.45 The median divides the distribution into two equal areas.

Figure 22: Textbook page that visually realises mean as a balance point of a distribution. (Watkins, Schaeffer, & Cobb, 2001, p. 54)

After that dead end[83] SM (Turn 7, Excerpt 1, Table 27) realised the mean through reference to a practical situation where he would use *average*, namely in an imagined situation where he would sell, for example, a car[84] to someone who will pay the average price. The realisation *average* is taken up by the rest of the group. Contrary to my intention as lecturer, my "Why the heck do we do it?" and "common sense" prompts signified for the students that the *use* of the mean should be justified, rather than the algorithm explored for sense-making.

At first glance the students are treating *mean* and *average* simply as synonyms, yet in Turns 10 and 15 KH's utterances suggests a primary ontological position for average. The discursants seem to share the common sense of average they suggest "people" have. The mere requirement to further unpack *average* seems ridiculous – the mean is "just" the average (Turns 8 and 9, Excerpt 1).

Excerpt 2 in Table 28 shows that the discussion in Group B, which occurred parallel to the discussion in Group A, started off in a similar way. The students realised the mean with average (Turn 165, Excerpt 2, Table 28), looked for an endorsed narrative in the textbook (Turn 166, Excerpt 2), and then continued with an imagined enactment of the use of average (Turn 170, Excerpt 2).

Table 28: Excerpt 2. Session 4: Discussion of Group B

Excerpt 2. Session 4: Discussion of Group B: Discursants: SDS, NM, GG, MM		
Turn	**Discursant**	**Utterances**
164	Lecturer	You all know how to calculate the mean...why the heck do we do it [See Turn 4 in Excerpt 1, Table 27]
165	GG	Well, it's the average.
166	NM	It's the balance point of the distribution (from textbook)
167	GG	We're supposed to be discussing it.
168	SDS	(To MM) Yes, she said...

[83] In the realisation map I strike through the text referring to this realisation to indicate that it is abandoned. I do not link it with an arrow to the other realisations either.
[84] The discussion of the meaning of the mean was preceded in the session by marking homework on the topic "a reasonable price for a used car." Hence SM had in front of him his report about prices of used RunXs, and I inferred that he was referring to the car context.

169	SDS	The logic behind the mean ... as a measure of centre (pages through textbook)
170	NM	I'm thinking of this in terms of maybe working on the mean of learners' marks. Let's say they wrote a test, then we work out their mean mark. What exactly are you trying to say? What information are you trying to deduct from that? Working on the mean, what does that tell you about the performance of those learners?
173	MM	You are basically saying, ja, you are basically saying I mean all of those kids, I mean they got sort of like that centre in your mark (hands gesture a small cluster)
174	NM	It's not a centre mark. The mean the way usually I look at it, is…

In both groups *average* was realised as a self-evident term, and utterances that realised the mean as a measure of centre or a balance point were ignored or rejected outright (see Turn 174, Excerpt 2, Table 28). The initial realisation map (Figure 23 below) shows that the literate statistical narratives of the mean as balance point and measure of centre were not intuitive and did not evoke visual mediators that were productive in the discussion. I mapped the only realisation of the mean as a numerical result onto informal statistical discourse, yet as the discussion in Group B continued it was evident that "the mean mark" was embedded in an evaluation routine that belonged to colloquial discourse. A realisation map which combines the utterances in both groups is presented in Figure 23.

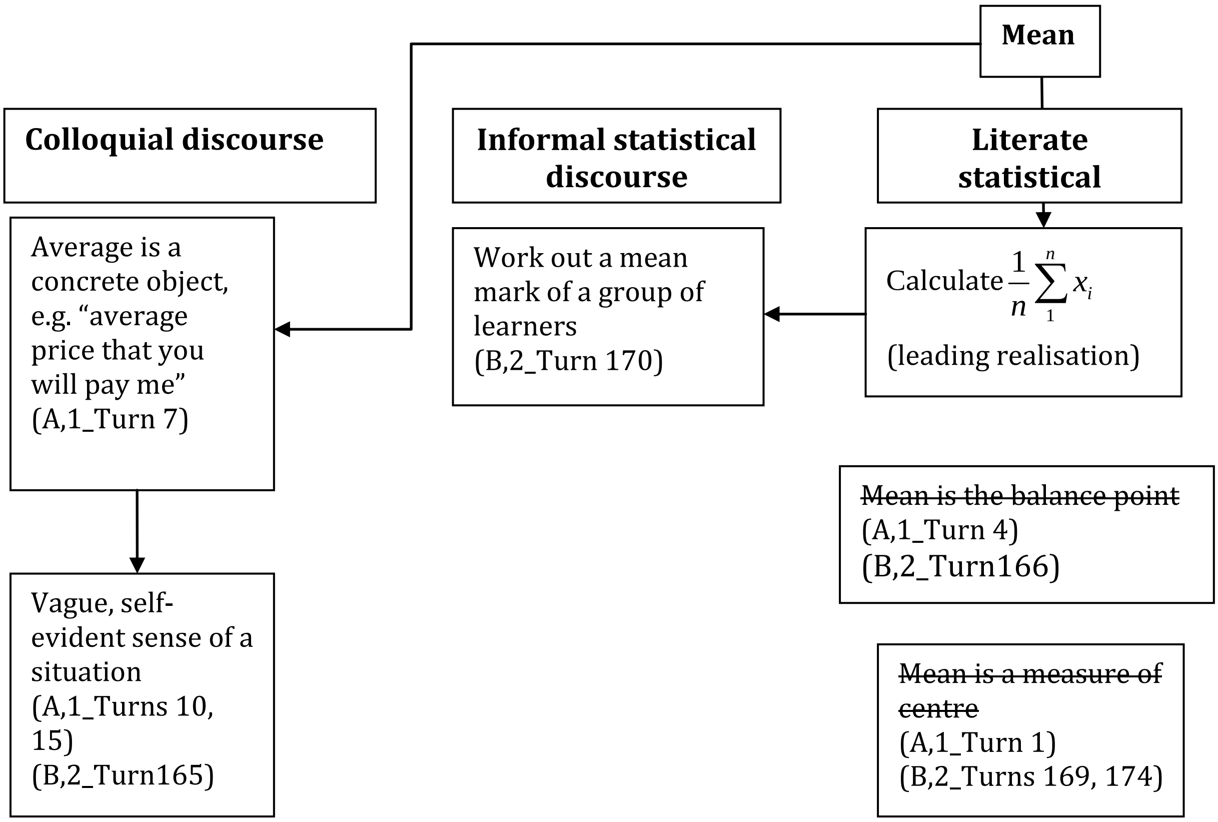


Figure 23: Stage 1 of a Realisation map of the signifier 'mean' in Groups A and B[85]

At this stage, the narratives on the mean and average suggest ontological collapse. To use mean was the same as to use average; to calculate the mean was the same as to find the average object. The consequential omission of this collapse is complete absence of references to objective properties of average. Average is experienced and directly observed in the extra-discursive world, and mean is just another name for average.

After the initial realisations of the mean as average in both groups, the discussions in the two groups developed differently. Group A proceeded to haltingly objectify *average* through discussion of its use in contexts. The concept 'middle' emerged and lead to commognitive conflict, which was resolved by exploration of various contextual examples. Group B's narrative remained evaluative and commognitive conflict around

---

[85] To help the reader to relate the utterances to the two groups of discursants I used A,1_Turn x to indicate that the utterance occurred in Group A, Excerpt 1, Turn x.

the subjective evaluation remained unresolved. I tell my research story about the shifting narratives on the mean by following Group A through their discussion, since their discussion was more productive in terms of developing informal statistical discourse. I then conclude my analysis by discussing the constraints of the evaluative routine that dominated the discussion in Group B.

9.10 Disobjectification narratives about the mean algorithm

9.10.1 The problem of what is the middle

In Excerpt 3[86] Group A realises the mean as middle, and middle as median. The median algorithm is endorsed, but its contextual meaning was queried. Narratives on the mean shifted back to colloquial discourse.

Table 29: Excerpt 3. Session 4: Discussion of Group A

Excerpt 3. Session 4: Discussion of Group A. Discursants KH,RK, GK, SM		
Turn	**Discursant**	**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.
19	KH	So then you know that half the data is above and half below, exactly half and half (looks at SM).
20	RK	I want to say, if you say the average height of kids...let's say one meter two [1.2 meter]...you say that generally...you find kids *around* that (moves hand horizontally at the same height at which he indicated 1 meter 2).
21	KH	Yes.
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.
25	GK	Sorry I missed that one, what did you say?
26	KH	(to RK) Why doesn't the median do the same?
27	SM	Inaudible.
28	RK	I think the median is like...if you have it ordered.
29	GK	Data, mm.
30	RK	You take the middle value.
31	GK	The middle value is the median yes.

[86] Utterances in italics letters were pronounced with more emphasis than the rest. Round brackets indicate gestures and other physical actions; in the square brackets I, the researcher supply clarifying text. Bracketed words and phrases are my interpretations of the spoken words in the videos.

32	SM	That is the mean.
33	KH	So if you have very tall kids in the class you're not going to get the mean there...you get the median.
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...
35	GK	Yes, you're given all the names.
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.
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.
45	KH	Yes, yes an impression.

In Excerpt 3 RK[87] explicitly refers to the mean only twice: first in Turn 18 and then in Turn 44. Between these two utterances, RK consistently disobjectifies the *mean* as the *average*. RK's narrative straddles colloquial discourse and informal statistical discourse. The objects he refers to are simple discursive objects: people with observably different heights. RK's realisation of the function of the mean as "encapsulating" (Turn 18) belongs to informal statistical discourse, since it signifies a single number "let's say one meter two" (Turn 20) which would hold the many real measurements. However, he flanks this realisation by utterances which belong to practical discourse: "gives them a picture" and "generally...you find kids *around* that" (Turns 18 and 20). The abstract mean is given agency in RK's narrative: it gives a picture and it encapsulates. The mean gains this agency by being the contextual average. Hence, average is an embodiment of the mean.

In contrast, KH's utterances realises the mean in literate statistical discourse. She refers to "data" rather than people (Turn 19, Excerpt 3), alienating height measurements as numbers existing separately from the people measured. Woven into the discussion about why we use the mean, are signs of conflict. RK's reference to "a particular number

[87] A quick glance back to Excerpt 1 shows that RK was silent when the other discursants in Group A referred to *average* as self-explanatory (Excerpt 1, Turns1 to 16).

encapsulating” in Turn 18 seems to signified for KH a sense of ‘middle’ which, in turn, signifies the median (Turn 19, Excerpt 3): “So then you know that half the data is above and half below [the mean], exactly half and half”.

KH starts her utterance with ‘so’ which indicates that her utterance is a conclusion following on RK’s description of the mean. Since RK did not use the term ‘middle’ KH’s utterance is unexpected. In Turn 15 (Excerpt 1, Table 27), KH indicated that median is different from average. She resorts to her conclusion again in Turn 26 (Excerpt 3, Table 29) questioning the role of the median: “Why doesn’t the median do the same?” Why doesn’t the median encapsulate and give a picture? Notice that KH glanced in the textbook (Turn 17, Excerpt 3) where it was open on the page in Figure 22. In the textbook, the mean and the median were visually realised as markers on similar histograms, produced vertically in relation to each other. On the top histogram a finger indicated the mean as a balance point and on the bottom histogram an arrow indicated the median. Visually the arrow with the caption ‘median’ replaced the hand. That the arrow and the hand are not positioned at the same point on the horizontal axis needed closer scrutiny, and was presumably not gleaned by KH. Yet, through her question in Turn 26 (Excerpt 3, Table 29) KH chained median and mean first by the properties “encapsulate” and “give a picture” assigned to the mean by RK, then with “half above and half below” which is the property yielded by her surface interpretation of the textbook diagram. As result the median became available as a discursive object in both the colloquial and informal statistical discourses. As to which discourse would be taken up rests in the data.

In the argument that developed between RK and KH about the use of the mean or the median, the complexity of the concept ‘middle’ is evident. RK realised two properties of average as *middle*:

> Property 1: Just as actual ‘kids’ with observable heights are around the ‘kid’ with average height, actual values are around the mean, so mean is implied to be in the middle of the actual values (Turns 18 and 20, Excerpt 3, Table 29)

> Property 2: Just as the average 'kid' is not the tallest or the shortest, the number that is the mean is not an extreme value, but between or in the middle of extremes (Turns 18, 36 and 38, Excerpt 3, Table 29)

KH's question in Turn 26 (Excerpt 3, Table 29): "Why doesn't the median do the same?" can be interpreted as a challenge about the meaning and use of mean as a middle value. RK counters the implication that the mean and the median are both middle values in the same sense, by emphasising that the median is derived from an ordered set of data (Turns 28 and 30). KH (Turn 33) continues the argument by introducing a disabling condition for the mean to be the middle value: "So if you have very tall kids you're not going to get the mean there…you get the median." KH did not say 'So, if you have *some* very tall kids…', yet it is implicitly understood at least by RK that she implies the existence of outliers in her imagined data set. RK falters. Acknowledging the inevitable variation in heights of kids in a real classroom, RK then realises *average* emphatically as a *number* that 'gives the middle' (Turns 34, 36 and 38), rather than indicate the 'kids' at the extremes, but he fails to explain how this middle is different from the median. RK's verbal realisation of measurements as *the height of kids* and his gestural realisations of *middle* are indicative of imagery related to concrete objects *kids of different heights*. His sharp gestures of high and low accompanied by his reference in the singular to "the tallest one or the shortest one" (Turn 36), stands in contrast with his horizontal gesture with "the average gives you the middle" (Turn 38). Hence, despite RK's reference to *the* average height (Turn 34), it seems that RK's middle is a range, while KH's middle is a point. I infer that KH's use of the median as middle would have the practical effect of separating the imagined kids into a tall group and a short group, with no one being not too short and not too tall. In RK's narrative this use of the median would not yield anyone of average height. From a literate statistical perspective, mean and median are both measures of central tendency, but in the colloquial discourse between RK and KH average as 'mean as middle' seems to indicate spread suggestive of the standard deviation. Both discursants are referring to a single number, but their narratives about the uses of this number are incommensurate.

The argument between KH and RK with regard to the meaning of mean and median is a key episode in the discussion, since it created conflict around the meaning of the mean in relation to middle and cracked the hold of colloquial discourse based on concrete images. Using Toulmin's (1958) argumentation scheme I can interpretively model the argument between RK and KH as follows:

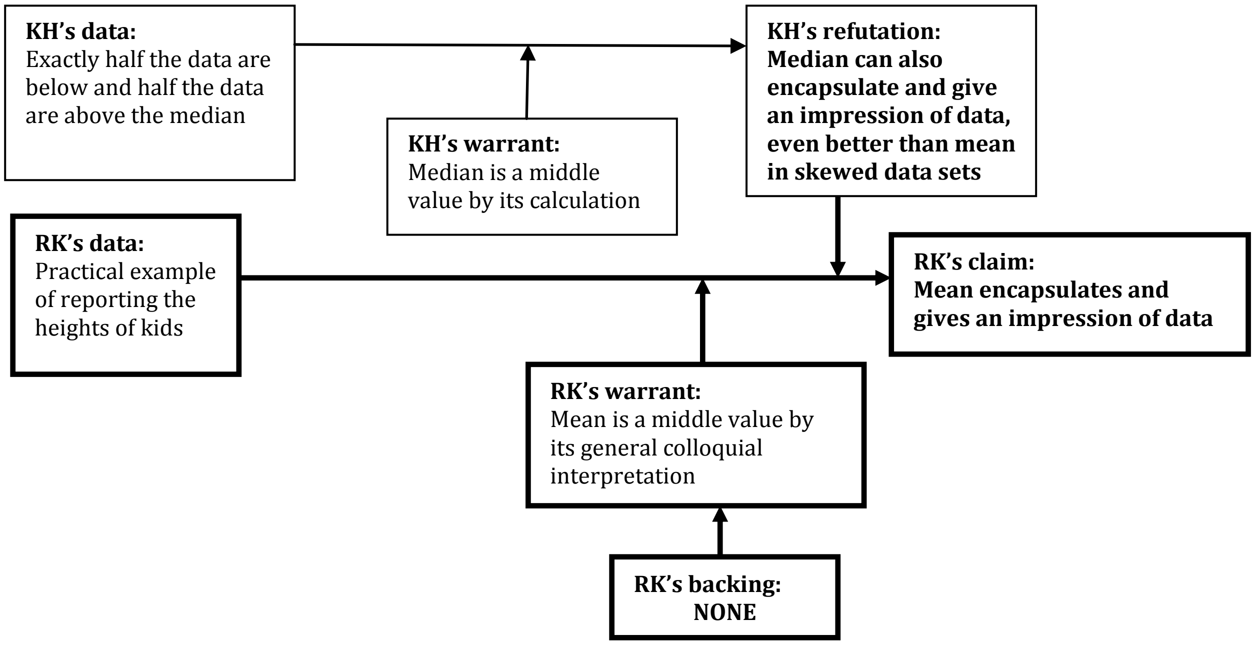


Figure 24: RK and KH's argument modelled according to Toulmin's (1958) argumentation structure.

RK's leading argument for the mean as a "middle value" is represented in bold textboxes. This argument was challenged by KH's refutation. In this argument the discursive status of the data, warrants and backings are decisive. RK's main argument was within colloquial discourse, with data and warrant consisting of practical examples of simple discursive objects. KH's refutation on the other hand, was taken from literate statistical discourse, and the data and warrant are strongly integrated with the operational definition of the median. Since KH's data and warrant were formally endorsed narratives, the claim seemed to have similar status. In order to withstand the refutation RK was in dire need of a strong, formally endorsed backing. The backing had to explain formally how the mean comes to be this middle number. The well known definition of median had enough authority and won out (Turns 40 to 43). Since the rule for determining mean as middle was not clear, the discursants were unable to endorse or reject each other's narratives on the mean and the median as middle values. Unable to resolve the commognitive conflict around mean as middle, RK and KH (Turns 44 and 45) retreated to the initial realisations of mean as “the general picture” and “an impression” of what is going on in a situation where it is used.

KH’s realisation of the median procedure, as well as RK’s realisations that assigned active roles to mean and median were not sufficiently alienated to be assigned to informal statistical discourse. The verbal realisations and the gestures collaborated to tell a story about evaluation of primary discursive objects. Median and mean competed as tools to identify *the average kid or the average kids*. Yet, the students’ verbal realisations were shifting toward exploration of the properties of mean and median and as such were shifting to informal statistical discourse. This was not the case with all the discursants in the group.

RK’s use of “middle value” (Turn 30) to identify the median signified in turn *mean* for SM (Turn 32). SM knew that the mean was calculated differently from the median, yet, *middle value* was such a strong signifier that the difference between the algorithms for mean and median faded in importance: for SM *middle* was a property of *mean* and median was allowed as a complementary object, since it also provided the middle. SM had not been a very vocal participant in the discussion so far. He offered the initial contextual realisation (Turn 7, Excerpt 1, Table 27) of mean as “average price” and in Turn 9 indicated closure with the conclusion that the mean is “just the average”. In the confusing and conflicting narratives on ‘middle’, the properties “encapsulate” and “give a picture” had moved into the background. In Vygotsky’s (1986) terms, in the group discussion (not necessarily so for individual discursants), *average* and *median* was now chained to *mean* by the property *middle value*. This was progress, since the discussion was turning to properties of the discursive objects *mean* and *median*. Mean was no more “just the average” it was the average because it was the “middle value”. ‘Middle’ ontologically preceded average. The realisation map (Figure 25 below) shows how this attention to statistical and colloquial properties of the discursive objects *mean*, *median* and *average* raised the complexity of the discourse and its generative potential. SM was drawn into the discussion of properties of the mean and median, even if he did not contribute regularly.

In Figure 25 I present an updated realisation map of the discussion about the mean.

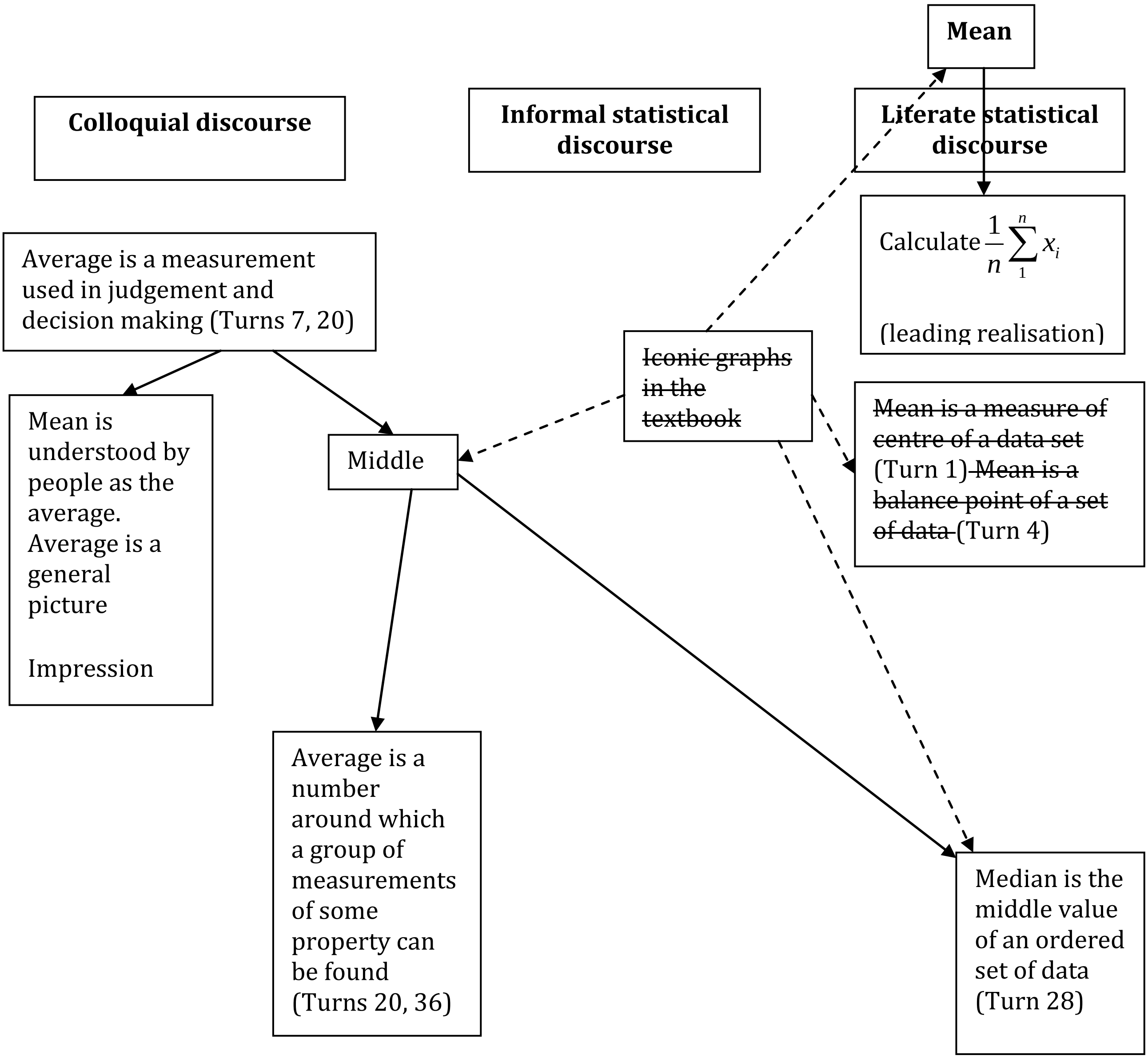


Figure 25: Stage 2 of a Realisation map of the signifier mean in Group A

This map was co-realised by Group A, rather than by any of the individual discursants in the group. The analysis so far shows that the problem of what the meaning of middle was, was far from sorted out. Diagrammatically presented, the realisations of the signifier 'mean' mapped onto two disjunct discourses (Figure 25): colloquial and literate statistical. Literate statistical narratives about mean and median as measures of central tendency were available from the start through the leading realisation of the mean algorithm, as well as through KH's reference to the graphs and descriptions in the textbook, but they were not adopted by the students in literate discourse. Similarly, the leading realisation of the mean as a representative number yielded by the mean algorithm had been abandoned completely. I do not doubt that all the students would have affirmed that the mean is a number, had I asked. However its contextual role as a number (through what this number does) was in focus, rather than its numerical statistical properties. Contrary to the colloquial use of mean as a centre of average values, colloquial narratives about the median were unconvincing. The median seemed not to be accepted as a representative number. The students seemed unable to disobjectify the median or the mean in narratives about the numerical properties of a middle number, which left the informal statistical discourse layer in Figure 25 empty.

9.10.2 The relationship between average and most

In Excerpt 3, Turn 20, (Table 29) RK pointed out that the mean is such that "generally...you find kids *around* that", therefore a centre within an interval. In Turns 46 to 50 (Excerpt 4, Table 30) another property of average is realised in colloquial discourse.

Table 30: Excerpt 4. Session 4: Discussion of Group A

Excerpt 4. Session 4: Discussion of Group A. Discursants KH,RK, GK, SM		
Turn	**Discursants**	**Utterances**
46	RK	An impression. I just get the right words.
47	GK	To express it ,mmm.
48	RK:	No but I want to get the general idea...
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?

In Turn 49 (Excerpt 4, Table 30) 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 brackets. In comparison to the gestures that accompanies RK’s description of the mean in the context of height (Turns 36 and 38, Excerpt 3, Table 29) GK’s gesture suggested grouping together of objects. In Turn 49 (Excerpt 4, Table 30) 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 interpreted GK’s combined verbal and gestural realisations as *most will be around the mean, because they are average* (see also RK’s utterance in Turn 20, Excerpt 3, Table 29). GK did not refer to a measurement that occurred most (the mode), but to the majority of cases that were grouped together as average. I interpreted this as signifying a spontaneous concept of a representative interval around the mean that can be didactically developed to an interval like that created by standard deviation. RK did not explicitly take up the notion of average as an interval, but his horizontal gesture in Turn 38 (Excerpt 3, Table 29) suggests that ‘middle’ referred to more than one case, while his emphasis on “general” together with a sweep of the hand (Turn 50, Excerpt 4,Table 30) supported replacement of many measures by one. I summarised the realisations of ‘mean’ that are endorsed by the group and the possible abstractions that may emerge from these endorsed narratives. Figure 26 shows the three narratives that emerged in the group’s endeavour to disobjectify the mean algorithm:

> The mean is in the middle of all the cases;
>
> The mean *is* an interval of measurements;
>
> The mean is a measurement in an associated interval that represents what is to be expected in the context.

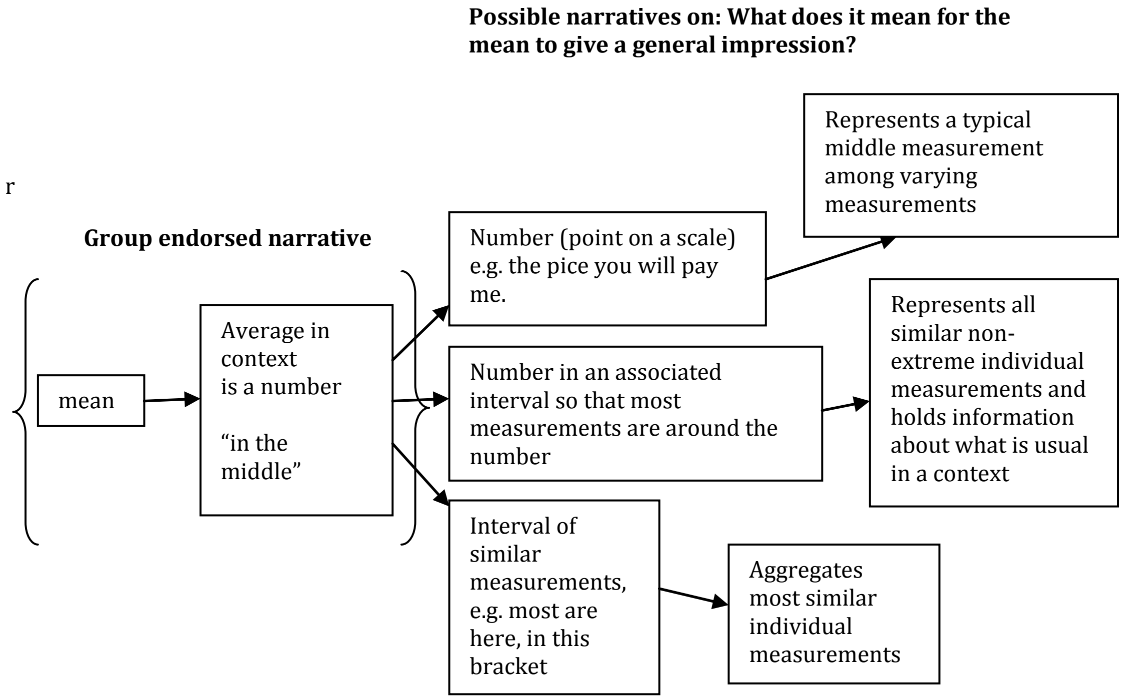


Figure 26: Three colloquial narratives about the mean as a middle number

Of the narratives, the mean as the *middle of all the data cases* is the most concrete and the narrative that emerged at the inception of the discussion. The narrative about the mean as an *interval of measurements* is statistically incorrect, yet contextually reasonable: colloquially there is not just one object or measurement that is described as *average* in some sense. The ‘average man’ represents all of us that are not extreme. The narrative about the mean as a measurement in an associated interval is rich in generative possibilities. Statistically it reminds of the standard deviation associated with the mean as a measure of spread.

The parallel discussion in Group B brought another interpretation of average into the discussion, namely average as a subjective, contextual judgement of *not good*. This judgement suggested a deed routine in colloquial discourse and prevented the discussion to shift to narratives about objective properties of the mean. The discussion in Group B serves as an example of the incommensurability of literate statistical discourse and colloquial discourse.

9.10.3 Commognitive conflict: the mean as an external norm or a measure of central tendency

Understanding the contextual use of the mean as a calculated value is the focus of the discussion in Group B.[88] The mean is calculated with a purpose, and it tells a story about a context (Turns 170 to 172, Excerpt 5, Table 31). Similar to the discussion in Group A, MM realises the mean as a central value in an interval of similar marks, and SDS (Turns 175 and 177) raises the difference between the median and the mean. But where the conflict in Group A arose about the need to explain how the mean is a middle value, in Group B the conflict is about the use of the mean as a contextual judgement.

Table 31: Excerpt 5. Session 4: Discussion of Group B

Excerpt 5. Session 4: Discussion of Group B: Discursants: SDS, NM, GG, MM		
Turn	**Discursant**	**Utterance**
169	SDS	The logic behind the mean ... as a measure of centre. (Pages through textbook).
170	NM	I'm thinking of this in terms of maybe working on the mean of learners' marks. Let's say they wrote a test, then we work out their mean mark. What exactly are you trying to say? What information are you trying to deduct from that? Working on the mean, what does that tell you about the performance of those learners?
171	GG	So what are you actually looking for.
172	NM	What are you looking for.
173	MM	You are basically saying, ja, you are basically saying I mean…all of those kids, I mean they got sort of like that centre in your mark (hands gesture a small cluster).
174	NM	It's not a centre mark. The mean the way usually I look at it, is
175	SDS	The mean gets affected by all values.
176	NM	Uhhuh.
177	SDS	The median, it doesn't really matter what the amount of the value is, it is the number ... of them. So if you have an outlier (gestures a point far to her right with one hand) that will affect your mean, that over here is quite large so your mean will be quite high (left hand indicates a shift towards the right). If it's still the median it will still be at the same position (left hand moves back to original position) even if your outlier was closer in (right hand indicates a point moving left).

[88] I do not report as extensively on the discussion in Group B as I did for Group A, since their narratives were similar, apart from the episode in Excerpt 6. The full transcript of the discussion in Group B is available in Appendix D.

178	NM	But now if I go back to the story about learners' marks for example, what exactly (inaudible.) Is there a specific value you are looking for or what? Cos the way I see it is, if the mean is *low* for that group then I get the sense my group is not performing well.
179	Group	Yes.
180	NM	And if the mean is just maybe...fifty percent (shakes hand horizontally at the same level) still I wouldn't be happy about it (weighing movements with hands). So, that *mean* for those marks, there *is* a certain number that we are looking at as a result. That I need my learners to get this mean and if they are not getting this mean, it means they are performing below it [this mean] on average
181	NM	If I use it [this mean] on the marks then I would know their mean is low, meaning *most* (gestures swirling movements with both hands as if including all) of them didn't perform well, that's why their mean, maybe, would be *low*. If they performed well, their mean would be higher, that would mean generally (hands swirling to the outside indicating ambivalence, uncertainty) they are performing...better.

NM takes the lead in the discussion with an example from her own teaching practice. What is the mean test mark of the class telling about the "performance of the learners"? (Turn 170, Excerpt 5) In contrast to the hypothetical example of an imagined situation at the start of Group A's discussion, NM's question is firmly embedded in a lived-experience context which compelled her own involvement and judgement. For NM the mean does not say something about abstract marks, but about the performance of the learners. Where marks may be high or low, performance is bad or good. The property of the mean as a central value is explicitly rejected by NM (Turn 174). Instead she uses the mean as a target value, independent of the data (marks) that she bases her argument on. SDS (Turns 175 and 177) joins the conversation and offers properties of the mean and the median as summary values. Her realisations seem to have been prompted by the question of the task, rather than NM and MM's utterances. SDS refers to "values", an abstract reference to data and she realises properties of the statistical objects median and mean. There was bound to be commognitive conflict, since SDS's discourse was literate statistical with a strong sense of avoidance of context, while NM's discourse was completely contextual and practical. NM listened politely to SDS but continued to realise her narrative of the mean in context.

NM was persistent about the relationship between the mean as a number against which to make qualitative judgements in context. In Turn 178 NM asks: "Is there a specific value you are looking for or what?" and follows up with a mapping: 'low mean' indicates 'poor performance' of her learners. Her rejection of the mean as a "centre mark" (Turn 174, Excerpt 5, Table 31) in conjunction with her realisation of the mean as being "*low*" (Turn 178) indicates to me that she placed the mean on a subjective, qualitative scale which she mapped onto a qualitative judgement of performance: Mean: "Low" → Average→ High can be mapped onto Performance: "Not well" → Average → Good. As I show in Figure 27, in NM's realisation of the mean, the mean itself could have an average value.

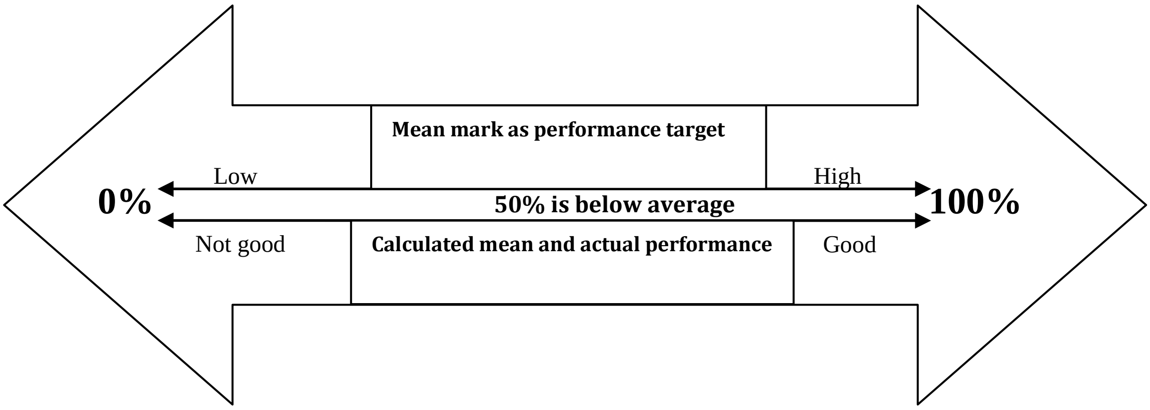


Figure 27: NM's association of mean and average

The scale on which the mean could take different values was not the range of the learners' marks, but an external scale derived from the possible marks a learner could get in the test: From 0% to 100%. In Turn 180 NM indicates that 50% as the calculated mean for her class would be below a certain targeted mean value. NM is ambivalent about whether the mean 'belonged' to the data set 'the learners' marks'. She compares "their mean" to "this mean" (Turns 180 and 181). From a statistical viewpoint, NM is confused about the mean as a property of the dataset, used to compare intra-data variation, and the use of the mean to compare variation between data sets. The only reference class evident from her narrative is that of her learners. The reference class for the targeted mean value is invisible.

Another ambivalence in NM's utterances is her frequent switch between the plural form 'learners' and the singular form 'group'. According to Sfard (2008, p. 170) assigning a signifier that is used in singular when talking about some property of all the group members together, is indicative of encapsulation and the formation of a new compound discursive object. One can speculate that NM was not comparing individual learners in her class to the class as an aggregate. From a statistical viewpoint, NM teetered between a collection of individuals and an encapsulated aggregate. NM might even have implied that her many learners had to individually perform better than "this mean" (Turn 181, Excerpt 5, Table 31) so that "their mean" could be higher.

NM's concern with performance gave rise to an extreme evaluation narrative where the marks that learners scored could not be alienated and objectified in order to yield a representative value like the mean. In Excerpt 6, Table 32 the incommensurability of NM and SDS's discourses leads to NM's statement of loss of meaning (Turn209).

Table 32: Excerpt 6. Session 4: Discussion of Group B

Excerpt 6. Session 4: Discussion of Group B: Discursants: SDS, NM, GG, MM		
Turn	**Discursant**	**Utterance**
181	NM	If I use it [this mean] on the marks then I would know their mean is low, meaning *most* (gestures swirling movements with both hands as if including all) of them didn't perform well, that's why their mean, maybe, would be *low*. If they performed well, their mean would be higher, that would mean generally (hands swirling to the outside indicating ambivalence, uncertainty) they are performing... better
182	SDS	You see I don't know if you can say that ... because if
183	NM	Is, if it's at school
184	SDS	No, I don't mean that, I understand the point you're making, but I'm saying we can't say *generally* (gestures with both hands scooping together) they are performing...because there could be ten of them that have quite high marks, and then you have...twenty of them that have really really low marks, and it could give you...
185	NM	Even in a class of thirty or fifty, it is an indication that they are not performing, if that's the case
186	SDS	(Nods) Ten out of thirty?
187	NM	Ten out of thirty, and yet when the mean is fifty you will say they are performing well, and fifty again, is it not that
188	SDS	So should we say that the mean is more...indicative of....
189	NM	The group's… performance
190	SDS	Performance

191	SDS	…more indicative of the values of your cases, whereas the median isn't
192	NM	(Shakes head in disagreement). The median is just telling you about the...
193	SDS	Where half the dots is...here or there (Gestures with hand along a horizontal line)
194	NM	(Nods head in agreement)
195	SDS	So mean is indicative of the
196	NM	Performance
197	SDS	Of values of the cases. The actual *values*
198	NM	Can we not go back to to to stats alone, just try and keep the conver... the minute we use the terms again somehow I'm lost ... somehow I'm mystified
199	SDS	Well we can't say performance, because that's only for marks. The average for anything
200	NM	OK I see where you're going
…	…	…
207	NM	After calculating the mean, do you then set your own standard, like you ...calculate the mean for this particular group, then there is the ideal value that we would like to get, compare with your ideal values? (moving hands up and down)...Are they performing according to this ideal, or are they below this ideal, or... maybe we are running away from the question
208	MM	I get you very well
209	NM	What is it that we've been saying…

SDS's utterances in this discussion consistently serve to disobjectify the mean through reference to its statistical properties. In Turns 175 and 177 (Excerpt 5, Table31) SDS used literate statistical terminology to react to the conflict around the realisation of the mean as a "centre mark". She explained that the mean is not always the centre mark, since outlier values could pull the mean off-centre. In such situations the median is the centre mark. This contribution was lost in the discussion until NM (Turn 180, Excerpt 6, Table 32) provided another opportunity for SDS to realise the lack of robustness of the mean as a centre mark by attempting an example in the context provided by NM. In Turn 181 (Excerpt 6, Table 32) SDS reacts to the realisation of mean as indicative of the performance of "most" learners "in general". Her counterexample attempts to show that a few good marks can skew a data set, so that the mean would not reveal that most learners actually did very, very badly. But the communication between SDS and NM breaks down. To NM it is obvious even without calculating the mean that if twenty learners have "really, really low marks" and only ten have "quite high marks", most learners are not performing well (Turn 185), irrespective of whether there were 30 or 50 learners in the class. Even if SDS wants to say the mean is 50%, NM knows on average the learners are performing really, really badly (Turn 187). The commognitive conflict seemed to arise from the rift between the literate statistical discourse of SDS where mean signified a property of a data set and the colloquial discourse of NM, where the mean signified an external target on the scale of possible marks for a test and "on average" signified either a general impression or performance halfway between 0% and 100%.

In Turns 189 to 191, and again in Turns 195 to 197 (Excerpt 6, Table 32), NM and SDS object explicitly to each other's word use. NM insists that the mean was indicative of "performance", while SDS rejects the use of performance in favour of the more alienated "values of cases". With this phrase SDS endeavours to separate case (learner) and value (mark). SDS's exploration narrative about the mean, median, data, and values of cases is incommensurate with NM's evaluation narratives about her learners' performance. NM indicates that she is not able to communicate in SDS's discourse (Turn 198). I represent the discursive dilemma graphically in Figure 28.

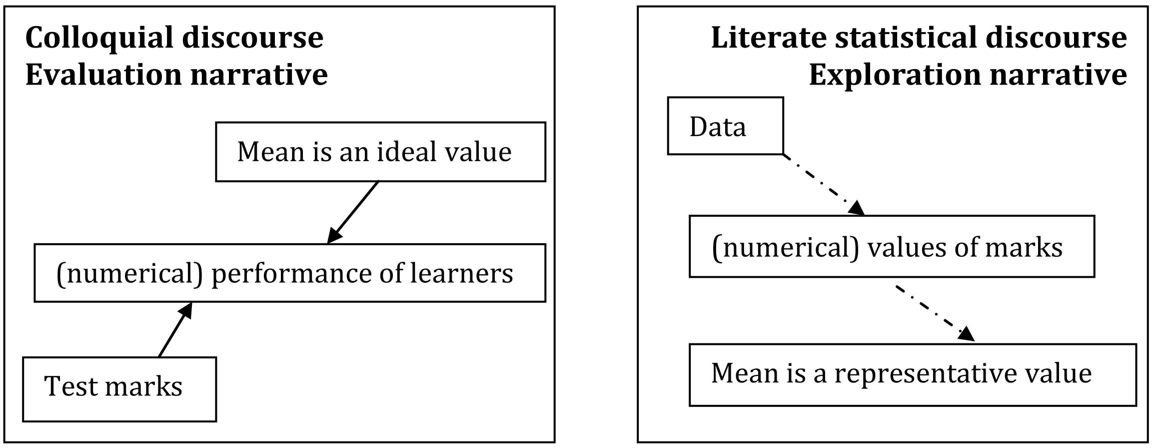


Figure 28: Incommensurable discourses.

In their discussion, SDS and NM are not talking about the same discursive object, although both talk about marks. To NM the class marks indicates performance, and the mean is subjectively judged as a value-for-her. In Turn 207 (Excerpt 6, Table 32) NM doubts the mean as a number that has contextual value in relation to performance: as her learners' mean performance was not up to standard or ideal, she asked for another standard or ideal value against which to judge the mean performance. In contrast, for SDS the learners' marks are alienated as values of cases, and the mean is a descriptive, rather than evaluative, value.

During teaching my interpretation of the discussion in Group A at the end of Excerpt 6 was that the students accepted the fact that the calculated mean represented the data as a ritual – because statisticians say so. Alternatively they called the mean the average and therefore accepted uncritically that the mean was a middle number in relation to the data in context. In Group B the mean was also realised as the average (e.g. Turn 199, Excerpt 6, Table 32), but contrary to Group A, the notion of the mean as a middle value was disputed in Group B. In both groups something was missing in the mapping between the two realisations: *mean* as a calculation that yields a number, and *average* as a sense of the general situation in a context. I propose that the missing narrative must tell how the calculation of the mean yields a middle number.[89]

9.11 Narratives about the syntax of the mean algorithm

The add-and-divide syntax was also the leading realisation in the group discussions. However, none of the students explained the logic of add-and-divide as fair sharing. I propose that fair sharing did not make intuitive sense to them in the context of variable observations and measurements. They were not able to construct narratives that explained the semantic logic of the mean algorithm, and they struggled to understand my prompt to do so.

In this section I report extensively on the discussion in Group A, and point out similarities and differences to the discussion in Group B. In both groups narratives emerged about the mean as a norm or a standard for comparison. In Excerpt 7 (Table

[89] I proposed in Figure 26 that RK was unable to counter KH's warrant that explained how the median achieved the middle position, by explaining how the mean calculation provides a middle number.

33) Group A could not explain a logical relation between the statistical mean[90] and sharing or grouping.

Table 33: Excerpt 7. Session 4: Discussion of Group A

Excerpt 7. Session 4: Discussion of Group A: Discursants: KH, SM, RK and GK.		
Turn	**Discursant**	**Utterance**
56	Lecturer	I'm after, what's the common sense behind it. What's the common sense of mean
57	KH	It gives you a kind of general impression of whatever the situation is
58	Lecturer	Does it?
59		(KH looks at RK, they mumble it does)
60	Lecturer	Suddenly you add everything together, all the prices of the cars, and you *divide* by the number of cars. What does that *mean*, if you do that? What does it imply?
61	RK	Generally the cars cost...(hand movements indicate a horizontal group/cluster)...an amount
62	Lecturer	Think of the calculation. What does that calculation mean? Go back to grade one. When you teach children to divide, what are they doing?
63	KH	Breaking up into equal parts...
64	RK	Or sharing...
65	Lecturer	Yes. *Sharing*
66	RK	Dividing (gestures vertical cuts)
67	KH	Equal portions...so if you divide by eight, you divide whatever it is into eight equal portions
68	Lecturer	So now here you are saying, take the money, share it equally between all the cars…
69	KH	If you do that ,but you don't do that
70	GK	Which means the average price you can get for a car, it falls among, arranges around a particular number, isn't it?
71	RK	Unless… it is not an exact value
72	GK	Mmm, just around a particular value
73	Lecturer	And that *value* is the one that you get when you say, oh, let's pretend they all cost the same
74	KH	Yes
75	Lecturer	What is the common sense behind that? What does it help you to pretend they are all the same? They are not the same!
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)

[90] With the distinction between the arithmetic mean and the statistical mean, I want to emphasise that the statistical mean is a measure of central tendency for distributions of data, while the arithmetic mean yields a "fair share" value in many non-statistical situations.

The realisations of KH and RK (Turns 63 to 67, Excerpt 7, Table 33) refer to actions on concrete objects that can be cut into equal portions. KH objects to my prompt that the money represented by the prices of cars was shared equally among the cars (Turns 68, 69). There was nothing in the situation about prices of used cars that logically signified such actions of collecting and sharing to KH. And indeed, the metaphor was not directly applicable. Yet, in statistics education literature equal sharing is acknowledged as an appropriate and even introductory realisation of the mean algorithm (Shaughnessy, 2007, p. 971). KH (Turns 73 to 76, Excerpt 7) acknowledges the hypothetical nature of the sharing metaphor that I proposed – the sense that one *pretends* that all cars have the same price if one shares equally when calculating the mean. Yet, she does not map this colloquial mathematical (not statistical) narrative onto a narrative of the distribution of variable prices of cars.

In Turns 70 and 72 (Excerpt 7, Table 33) GK confirms that the mean as the result of a calculation is not the same as the average. She realises average as an interval which contained the mean, possibly as the middle number. For GK the average exists in context, while the mean is calculated into existence and then fitted to the average values. Slowly an objective relationship between the mean and the imagined data values emerges. While RK's imagery still indicates decision making in context, he realises the mean as a contextual expected value[91], "…I know what to prepare" and indicates that the actual values could be obtained by addition to or subtraction from the mean (Turn 77, Excerpt 7, Table 33). In comparison to SM's initial realisation of the mean as "the average price you can pay me" (Turn 7, Excerpt 1, Table 27), and RK's own repeated realisations of the mean as an average which gives "an impression" of a context, RK's

---

[91] The expected value is the mean value of a random variable. For a discrete variable X, taking values x_1, $x_2,\ldots,x_n$, with probabilities $p(x_1)$, $p(x_1),\ldots,p(x_n)$, the expected value of X is given by

$E(X) = \sum_{i=1}^{n} x_i p(x_i)$ (Porkess, 2004, p. 58). Watson, Callingham and Kelly (2007, p. 84) state "For school students, however, expectation with respect to chance and data is likely to be experienced in terms of probabilities, averages, "caused" differences, and random distributions of outcomes, whereas variation is likely to be experienced in relation to uncertainty, anticipated change, unanticipated change, and outliers (Watson, et al., 2007, p. 84).

narrative about the mean in Turn 77 (Excerpt 7, Table 33) is exploratory. RK has taken up the challenge to disobjectify the mean, rather than to objectify the average.

In Figure 29 I relate the realisations in the narrative about the mean as a fair share calculation to Group A's previous realisations of average. I omit the realisations about median, since there was no reference to the median concept in the fair share discussion.

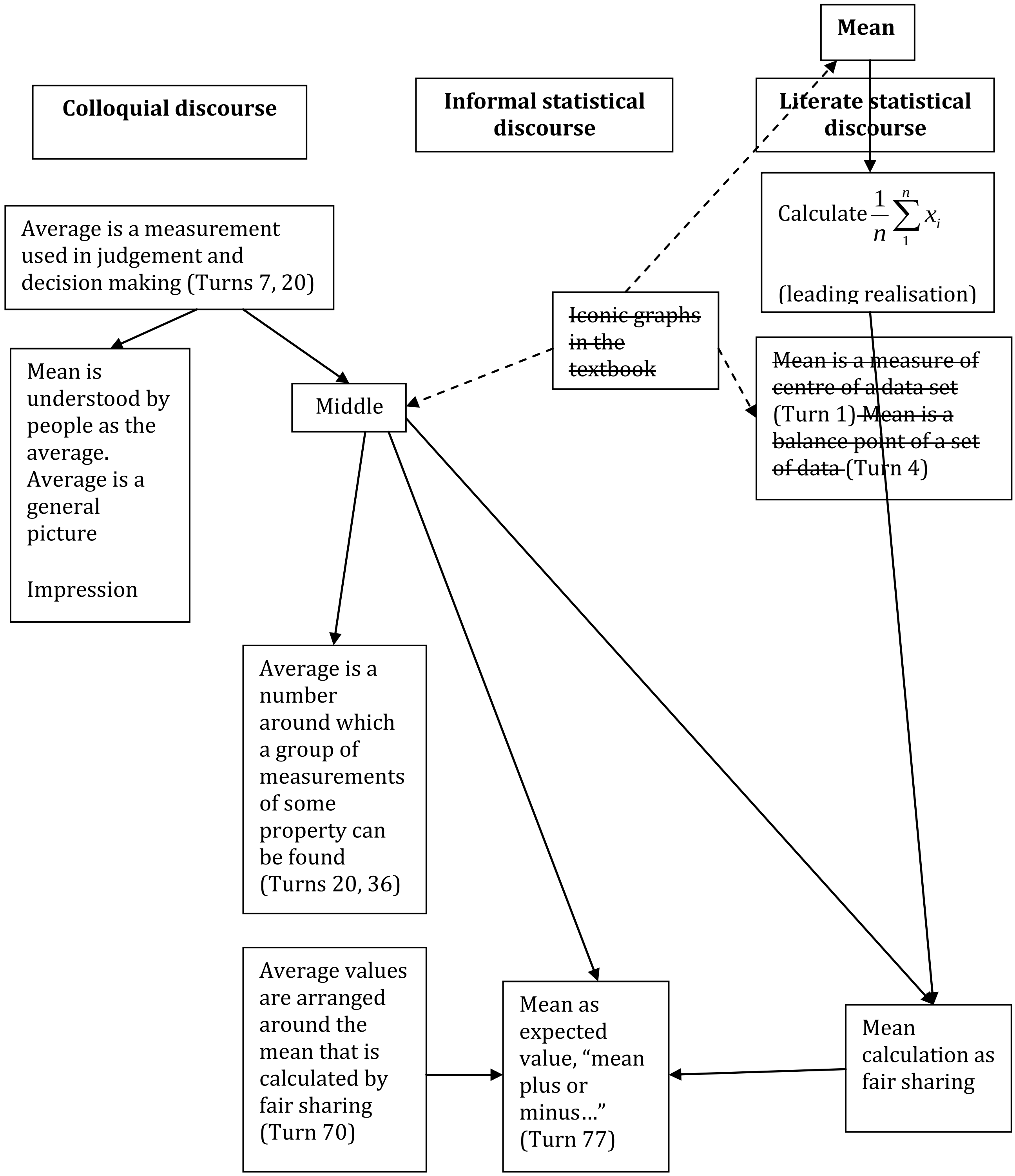


Figure 29: Realisation map: mean as a fair share

The vague impressions of mean as average and middle are now replaced by narratives 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. Since the narratives are about the measurements (e.g. prices) as numbers, the narratives are shifting to informal statistical discourse. However, the shifted discourse is not stable yet.

Between Turns 77 and 113[92] the discursants in Group A repeated their realisations of the mean as the average that falls within an interval of similar values and gives an impression of a situation. They also searched the textbook for references to sharing, but failed to relate the mean as a balance point to the sharing metaphor. In Turn 113 (Excerpt 8, Table 34) I asked the discursants to background their narratives on the average.

Table 34: Excerpt 8. Session 4: Discussion of Group A

Excerpt 8. Session 4: Discussion of Group A: Discursants: KH, SM, RK and GK.		
Turn	**Discursant**	**Utterances**
113	Lecturer	(Returns to Table A) Pretend you didn't know the word average
114	Lecturer	Does it *help* you to see what is going on in your data set if you think about...you're sharing the price, the total price there? Does it help you to understand?
115	RK	(Nods in agreement)
116	GK	It tells us, if we talking about sharing the money, it tells us, how much... how many of us are going to get that ah amount. You know what I'm saying. It's how many (swirling movement with hand, palm up) of us are going to get that amount. Even if there are others...are going to get a little bit more, but most of us will get... that amount
117	RK	Generally
118	KH	Does it always, does it tell us that?
119	GK	It does!
120	GK	Look at the median
121	RK	I think what GK is trying to say, it gives us a sense of what each person is going to get
122	GK	Mm, that's what...
123	KH	Yes, yes as a group, in the group
124	GK	Mm...but it doesn't mean that all of us are going to get the same
125	RK	Not exactly yes, but more or less
126	GK	...just giving us a picture

[92] See the full transcript in Appendix D.

127	KH	If it was an equal distribution within the group (gestures)
128	GK	That's what we get
129	KH	OK
130	Lecturer	(Returning to Table A): I noticed when you said that you said *most* of us are going to get that. I want to challenge the word *most*. Does the mean tell you what *most* of you *are* going to get?
131	KH	No
132	RK	I think, I think *all* of us
133	KH	Yes, if it was distributed fairly... (gestures) then it would be all of us
134	GK	Even if it is all of us, you might find that one is out, or even two is out
135	KH	What if you take the mean though, if you take the mean...and you work that out...and you multiplied it, you get back to your value. It's like with the cars here, if you added up all the prices and you divide it by the cars you get your mean, if you multiply that by the cars you get back to the total price
136	GK	Mm
137	KH	We get the total price
138	GK	That is what I'm saying. It tells us of how much is each going to get in the group. But at the end of the day it doesn't tell us *exactly* that we all gonna get the same. There might be one or two that is out.
139	Lecturer	That's now when you go back to the original distribution
140	GK	Yea
141	Lecturer	And you compare to the mean, you say, no we didn't all get the mean.
142	GK	Mmm
143	Lecturer	So it's a ...[the mean] is something that we do in our heads to get a, seems to me a number that says, what *would* we all get *if* we all got the same. Right?
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 (gesture shaking hand) an approximation
145	KH	Yes
146	Lecturer	An approximation of *what*
147	RK	Ah, like ah, I'll be taking an example. There's an amount of money to share between people, then we say, OK we take the amount and we divide it by the number of persons, then we say that approximately...Oh no, let's go over to the prices of cars. The prices vary, so if you took the mean of the prices you'll say that generally in this shop, if you want a RunX, it's going to cost *about* this amount. But it won't be exactly that amount. If, if somebody who wants to buy a car gives that information, that would give him a sense of how he should prepare.
148	GK	Ja
149	RK	But then it's not going to be exactly that amount, it's going to be plus or minus
150	GK	That amount
151	RK	I don't know if it makes sense
152	Lecturer	Ask your friend
153	Lecturer	(to SM) Does it make sense to you?
154	SM	I think it's making a lot of sense...Because my understanding about the

		mean...when you when we talk of the prices of the cars...so for us ...the average says to us, that is when you add the total price of the cars and divide by the number of cars then the amount that you're going to get that is the average *amount* that you're going to pay. It's not the exact amount it's the average amount that you are going to pay.
155	RK	More or less
156	SM	Yes, if you talk of the car that is priced at eighty thousand (80000) and then at ninety thousand (90000) the two cars. Then the average amount is eighty-five thousand (85000)
157	KH	So it's like taking the total amount and share it out equally amongst each individual (hand moves horizontally across at the same height)
158	SM	Yes
159	GK	So it is an indication of how much we can expect to pay.
160	SM	Yes, yes, it just gives us the average [85000]not the *exact* amount..[80000 or 90000]what you're going to pay
161	RK	So if we have raw data, just prices of cars, it will be confusing to consider each individual price like that (gesture going down a list) then you have to start deciding how am I going to buy one...but then…the mean tells you exactly how much approximately you are going to pay

In Group B's discussion (Excerpt 9, Table 35) SDS's literate statistical narratives about the mean algorithm as "evening out" and "balancing" are not understood by the rest of the discursants.

Table 35: Excerpt 9. Session 4: Discussion of Group B

Excerpt 9. Session 4: Discussion of Group B: Discursants: SDS, NM, MM, GG.		
Turn	**Discursant**	**Utterance**
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 (gesture 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
245	MM	OK so what is it we are trying to share...are we trying to share this one mark to all of those kids or what?
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 (gesture)
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
252	NM	So why is it important?
253	SDS	So why *is* it important?
254	GG	The fact that it is sharing amongst...groups
255	NM	... to see how much they will get...compared to the whole. What is the whole, is the whole not getting...full marks? How far are they from

		the whole if they were to share...
256	MM	You are talking standard deviation
257	NM	Ha? No I mean I'm still going back to this interpretation of mine of if they are below average or above average or what...somehow I keep going back to that
260	SDS	We can talk of this in terms of balance
261	NM	Trying to balance...to balance
262	GG	Balancing what?
264	SDS	Balancing the values of the cases, so each case has the same value
265	NM	But why?
266	GG	Ja, why?
267	SDS	Why? So that's.... (gestures, draws funny face)
268	GG	Ja, you see, it's (airy hand movements)
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)
271	NM	Yes, I keep going back to it it's a judgement, judging if they
272	SDS	Judging on assessment criteria
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, 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..
274	SDS	The...cluster
275	GG	Cluster?
276	SDS	It's far away from the cluster of values
277	GG	Ja
289	Group	(All laugh)
290	GG	(To SDS) What did you say about the judging thing?
291	MM	Ja, I mean it's all about judging
292	SDS	(Reads from her book) Balancing values so that each case gets the same value so that you can judge and assess a set of cases
293	MM	Balancing what?

Both MM (Turn 245, Excerpt 9, Table 35) and NM (Turn 255, Excerpt 9, Table 35) are not sure what the object was that would be shared equally. It is plausible to suggest that their mathematical sharing metaphor was that of a single whole shared into equal parts. NM's persisting practical and evaluative narrative about her learners' marks prevents her from objectively describing the total reached after adding the marks. To her, the "whole" was "full marks" (Turns 246 and 255, Excerpt 9, Table 35). Yet, in her colloquial discourse she acknowledges the usefulness of some number that could tell her how far her learners were from the ideal full marks – she was intuitively aware of the need to describe deviation, but not deviation from the mean. SDS illustrates levelling out by compensating the lengths of the bars of a histogram (Turn 247, Excerpt 9) and while this strategy to obtain the mean is acknowledged by the other discursants, the reason for doing so remains unclear (Turns 252, 253, Excerpt 9, Table 35).

To the discursants in Group B the literate statistical concept of balancing was no clearer than levelling out. SDS took the balancing metaphor from the textbook, but could not explain why it was necessary to balance data around the mean (Turns 260-267, Excerpt 9, Table 35). SDS's balancing narrative was about the actual data values, and did not tell the story of balancing the deviations from the mean (Turn 292, Excerpt 9).

9.11.1 Deviation narratives about the mean

The students' action based narratives in colloquial discourse showed awareness of variation between measurements, but little awareness of any pattern in the variation. Awareness of distribution as a pattern in variation (Bakker, 2004a) was limited to a sense that "most" values are expected to be similar enough to the mean to be called average and "a few" that would be too different to be average (see Turn 116, Excerpt 8, Table 34). In the discussion of the mean algorithm, the students did not realise the variation that existed between measurements of car prices or heights as a variable and measurable property itself. Although they were aware of the variation between their imagined contextual measurements, and their gestures indicated that they were working with intuitive notions of distribution of measurements around a middle value, the students did not experience a need to order or quantify the variation between measurements.

In statistics education literature researchers argue that understanding of measures of central tendency of a data set is dependent on understanding a distribution of data as an abstract discursive object (Bakker & Gravemeijer, 2004). According to Sfard (2008) abstraction is based on our previous discursive actions on objects, yet my students' narratives about the act of distributing measurements were limited. RK's utterances and accompanying gestures were the clearest suggestion of images of children ordered according to their heights (Turns 36-38, Excerpt 3, Table 29). However, the ordered distribution of data was only related to the median, not to the mean (see Turn 28, Excerpt 3, Table 29). Two realisations approached the notion of the distribution of measurements. Firstly, in Turn 156 of Excerpt 8 (Table 34) SM created a numerical interval around the mean, which was narrow enough that the mean could in practice be a reasonable approximation of the actual prices. Secondly, in the parallel discussion among Group B, GG (Turn 273, Excerpt 9, Table 35) gave a numerical value in relation to her imagined mean that she judged as unusual: “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, say it’s 130 000 then we can see that the car that was 190 000 is *way* out, was way out of the average ... OK? Not average, was way out.” The judgement of the size of the difference between a data value and the mean value was intuitive and contextual, but has potential for development of deviation narratives.

Such intuitive judgements of deviation from the mean as acceptable or unusual, without intuitive recourse to establishing the centre of the deviations objectively (i.e. a zero deviation value) is at the core of the problem to relate the arithmetic mean to the statistical mean. That the mean algorithm had no roots in the students’ colloquial discourse was expected from the historical overview of its development. However, even deliberate effort to construct plausible narratives about the semantics of the algorithm was a difficult task. Commognitive conflict arose among the discursants in Group A as to what exactly the object of division was and what the result of division would convey. I interpreted narratives about deviation from the mean as progress in the discursive disobjectification of the mean algorithm. In order to eventually link the realisations of average as middle to the mean as the result of addition and division, understanding the mean as a single value from which actual measurements deviate would be a key step. The invariant property of the mean as the number such that the sum of deviations from the mean is zero, was for me the important property to develop.

I noticed in the analysis that my utterances explicitly separated the mathematical calculation of the mean from the ‘real situation’. I maintained that we calculate the mean "in our head" and then return to the data in order to compare the actual measurements to the calculated mean. I summarised the discussion with an explicit subjunctive narrative on the mean: "what *would* we all get *if* we all got the same" (Turns 139, 141 and 143, Excerpt 8, Table 34). I explicitly realised the hypothetical nature of the mean compared to the actual measurements in a data set, and was rewarded by RK's conclusion (Turn 144, Excerpt 8, Table 34):

> It gives me a sense that uh, I think the value of the mean perhaps we should have a constant that we have. [The actual measurements] It's [the mean] plus or minus *something*. You know what I'm saying (bashful), because, it's like (gestures a shaking hand) an approximation.

RK tentatively realised the mean as some constant value compared to the variable measures in a dataset. This realisation signalled a shift in his discourse: without the mean, we are aware of relative variation among actual measurements; with the mean we become aware of deviation from a single a hypothetical measurement. RK interpreted this "constant" as an approximation to the actual values in context. RK's choice of the term constant was meaningful. Merriam-Webster online dictionary defines the noun 'constant' as follows:

> Constant (noun): something invariable or unchanging: as
>
> **a**: a number that has a fixed value in a given situation or universally or that is characteristic of some substance or instrument

Indeed, at this stage of the discussion, Group A had disobjectified the mean calculation in the sense of definition a. The mean calculation yields a number that has a fixed value for a given data set and it is characteristic of the size of the measurements (heights or prices) that make up the dataset. RK's seemingly contradictory utterance in Turn 161 (Excerpt 8, Table 34) clinched the constancy of the mean: "the mean tells you exactly (without varying) how much approximately you are going to pay." The mean is not a vague sense of 'average' anymore - it is a constant, a number that gives a reliable impression of the situation.

As this discussion wound down, SM, who had been a rather non-vocal discursant, offered a specific example to realise his understanding of the mean and its use (Turns 154, 156, 160, Excerpt 8, Table 34): The mean is the average amount, which is the middle value between extreme values. In the literature this kind of realisation is interpreted to signify mean as midrange (Bakker, 2004a). In the context of this discussion, I interpreted SM's contribution to also realise the mean as a constant number from which actual measurements deviate in a patterned way. I cannot claim that he assigned any deeper understanding related to the differences or the sum of differences from the mean, but had I realised in-the-moment where this example fitted into the realisation map of the mean, I would have taken it up to stimulate discussion about mathematical realisations of mean as middle. Sadly I missed the opportunity!

Parallel to Group A's realisation of the mean as a constant, SDS in Group B called the mean a "norm" and levelling out "finding the norm" (see Turn 232, Excerpt 9, Table 35). However, norm might have signified different discursive objects to the discursants in Group B. Merriam-Webster online dictionary provides the following realisations of "norm" as a noun:

> 1: an authoritative standard: model
>
> 2: a principle of right action binding upon the members of a group and serving to guide, control, or regulate proper and acceptable behaviour.
>
> 3: average: as
>
> > a: a set standard of development or achievement usually derived from the average or median achieved of a large group
> >
> > b: a pattern or trait taken to be typical in the behaviour of a social group

NM was likely to understand "norm" as an authoritative standard or a principle of right action in relation to her class' marks. While SDS's norm was derived from the "average" of the imagined group of data, norm might still have indicated a typical car GG (Turn 273, Excerpt 9, Table 35), one that is not "way out" expensive.

9.11.2 Routines realised in the discourse about the mean

According to Sfard (2008), there are three kinds of mathematical routines, namely deeds, rituals and explorations. She describes the difference between deeds and explorations according to their aims. Deeds are practical routines, which produce a change in discourse independent objects, while explorations such as mathematical routines aim at producing narratives about discursive objects, such as mathematical or statistical objects. Rituals are aimed at "alignment with others and social approval" rather than "any kind of self-sustained product, as is the case with deed and exploration" (Sfard, 2008, p. 301). The students' leading realisation of the means as its algorithm was a ritual, aligning themselves with what people do when they refer to the mean. KH and SDS's narratives that realised textbook terminology related to the mean as a balance were also rituals, since they acknowledged that they did not make sense of the concepts. Attaining a literate discourse on some object implies that literate procedures replace colloquial procedures. This process was started with the shift from the mean as the average, to the mean as a middle value and eventually the mean as a standard or a norm. The students realised deed routines with their narratives about average height and average price. NM's narrative about the mean mark of her class is a clear example of a deed routine, because her purpose was to evaluate her learners' performance. Exploration routines developed when the commognitive conflict about the meaning of middle arose, and when I insisted that they embody the calculations in the mean algorithm. Their exploration of what it implies if the mean is in the middle, and what the effect of equal sharing is, lead to the disobjectification of the mean as a standard.

Mathematicians construct new discourses on mathematical objects completely intradiscursively, while constructing a new (literate) discourse on a mathematical object requires of learners and students to bridge colloquial and literate discourses. The result is often a conflation of the two discourses. As Sfard (2008, p. 230) explains: "What the mathematician views as inherently intradiscursive, metalevel activity, less experienced mathematists would often replace by an object-level, quasi-empirical procedure…in which the participants of a conversation refrain from discourse-on-discourse (from manipulatives on narratives) and construct narratives about extradiscursive reality on the basis of what they know from their direct, everyday experience." In the discussions of both groups, the narratives on the mean were about the extradiscursive reality based on everyday experience. When a statistician pronounces "mean is the middle or centre of a data set" the utterance is metadiscursive and filled with relationships between previously endorsed statistical narratives on distribution, deviation, balance, position and representativeness. When my students uttered "mean is average" and "mean is middle" their narratives were based on what they know from everyday experience, to the extent that they could imagine the specific case (car or learner) that would embody a middle position. According to Sfard they bypassed the metadiscourse on mean, and replaced it with narratives on some directly observable situation.

9.12 Summary

By means of summarising the discussion, I present Group A's realisation map for 'mean' as it had expanded to the end of the discussion (Figure 30).

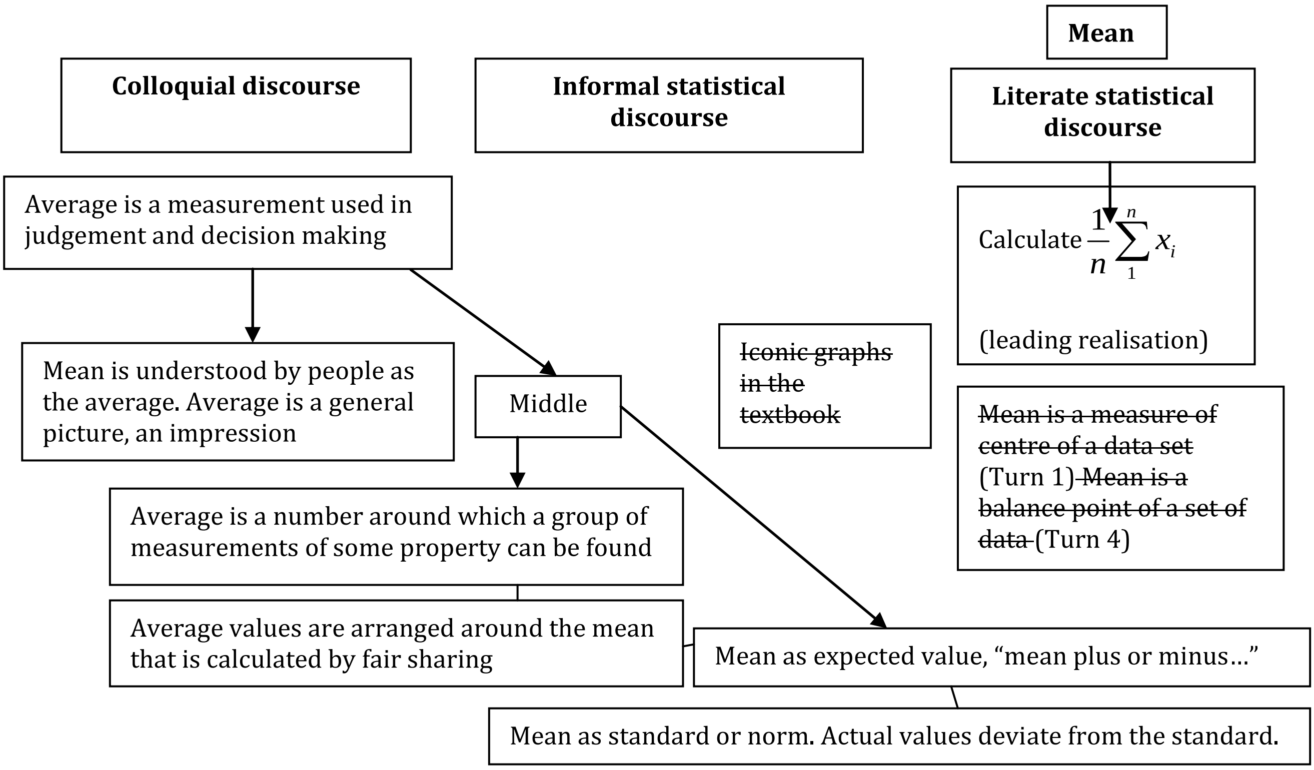


Figure 30: Discourses about the statistical mean and its algorithm

While the textbook provided from the onset an endorsed statistical narrative on the mean, the students were not able to make sense of the endorsed narratives when compared to their colloquial realisations of mean as average in context. Specifically, the realisations of the mean as a balance and a fair share were incompatible with average in context. Average was realised by the students as the discursive object ‘middle’. I have attempted to follow the dynamic process of discursive development of the mean as middle by categorising the realisations as narratives in different discourses. The realisation of average as in some way a middle value that relates mean and median signified to me shift toward informal statistical discourse. However, ‘middle’ was not a stable property and narratives on middle reverted back to colloquial discourse as the students endeavoured to reach shared understanding about the nature of middle. The conflicting realisations were: middle as median, separating the data ordinally in two halves; and middle as where most data are, collecting all measures that are the same. However, I have shown that gestures that accompanied verbal utterances indicate an understanding of mean as the middle of an interval of similar, not equal values. Hence, I interpreted their references to most as an interval of (standard) deviation, rather than a modal interval. The shift of variation narrative to deviation narrative enabled the final realisation of the mean as a constant from which actual values deviate.

Just as Sfard (2008, p. 183) points out that the notion of a process which is its own product is unimaginable in the extra-discursive world, it was accepted, but unexplained in the discourse realised by the students in my study, that the mean algorithm could produce the “impression” which they called “average” in colloquial discourse. The definition of the statistical mean[93] as the calculation algorithm of the arithmetical mean is not conducive to develop a bridging discourse for statistical reasoning, especially not for high school or tertiary students who already know the algorithm. Yet, in textbooks the mean is defined as the result of the calculation (Groth & Bergner, 2006) and then its consequential properties as balance point or a measure of central tendency are illustrated. I propose the following didactical definition for the mean that will make its

---

[93] I use statistical mean to indicate its use in working with a set of measurements as data, rather than with context free numbers

statistical use explicit and from which the calculation algorithm can be deduced, along with the statistical properties of representation:

> The mean is a numerical value that indicates a hypothetical measurement of zero variation among varying measures over all the cases in a data set. As such the mean is the origin for measuring variation in a data set.

Had I the knowledge-for-teaching of this nature of the statistical mean when the discussions started, I would have been able to take up two key realisations to develop in tandem:

Key realisation 1: Mean as 'average-is-middle'.

Key realisation 2: Actual data points are variable and deviate from the mean.

I propose that variation should be abstracted as a measurable property of data in order to develop the mean as a discursive object, rather than a process. Starting from colloquial roots compensation strategies as described by Bakker (2004a and 2004b) have potential to be formalised to yield the mean algorithm. Growing down from literate discourse, the mean can be disobjectified as a standard for comparison between variable data points, and developed mathematically to yield the algorithm. The discursive action that is needed to underpin the mean algorithm is the measurement of variation. For two values, middle is intuitively easy to establish. Middle means the distance from middle to either of two ends is the same. What if we have three values? It is intuitively accessible that on a straight line, such as a number line, the distances from middle to any of three values are not the same, yet we can assume rather directly from the spontaneous term 'middle' that the combined distances to points below the middle must be the same as the combined distances to points above the middle. Mathematically, we derive the formula of the mean from the middle property as follows:

Let x_1, x_2 and x_3 be three measurements and let $\bar{x}$ be a number such that
the sum of the distances $|x_i - \bar{x}|$ for $x_i < \bar{x}$ is equal to the sum of distances $|x_i - \bar{x}|$ for $x_i > \bar{x}$
For example, let $x_1 < x_2 < \bar{x} < x_3$.
Then $|x_1 - \bar{x}| + |x_2 - \bar{x}| = |x_3 - \bar{x}|$

$\therefore\ -(x_1 - \bar{x}) - (x_2 - \bar{x}) = x_3 - \bar{x}$

$\therefore -(x_1 - \bar{x}) - (x_2 - \bar{x}) - (x_3 - \bar{x}) = 0$

$\therefore -(x_1 + x_2 + x_3) + 3\bar{x} = 0$

$\therefore 3\bar{x} = x_1 + x_2 + x_3$

$\therefore \bar{x} = \dfrac{x_1 + x_2 + x_3}{3}$

This derivation of the mean can be generalised to any number of data points. Mathematical concepts specifically related to word use that underlie this route to the mean algorithm are algebraic *difference as a distance*, based on the *spatial distance metaphor* of numbers on a number line. (Lakoff & Núñez, 2000) Considering the midrange interpretations of average and mean cited by Bakker (2004a), Mokros and Russell (1995) and Watson and Moritz (2000), which was also realised by MM in Turn 156 (Excerpt 8, Table 34), the spatial metaphor related to average as 'the middle' is intuitive.

In Chapter 10 I interpret the students' discourse in relation to the literature on informal reasoning and reflect on my research findings.

Chapter 10: Conclusions and Reflections

10.1 Introduction

The research questions that guided my analysis of the discourse of teachers as students in an introductory statistics course were:

a) What are the objects and the narratives of the discourse(s) that are realised during informal reasoning and discussions of statistical objects?
b) What shifts are evident between colloquial discourse and literate statistical discourse?
c) What are constraining and productive narratives in the shift towards statistical discourse?

I framed my study within the process of curriculum change in South Africa and the converging international vision for statistics education at school level. I argued that meaningful statistics education in South Africa is dependent on professional development that deals with the complete cycle of statistical investigation and focuses on developing statistical reasoning. Teachers who want to teach statistics meaningfully despite a narrow curriculum will have to be able to access everyday contexts for statistical deliberation. In addition they need to be able to understand incipient statistical discourse in order to conduct classroom discussions that provide a bridge to statistical discourse. Statistics education literature supported my concern with the initial stages of statistical reasoning in order to grasp the system dynamics of a context. In particular, I drew on Wild and Pfannkuch's (1999) model of statistical inquiry which includes a planning stage that requires structuring the data-context with the view to do statistical investigation (see Figure 1, Chapter 1). I argued that grasping the system dynamics as a learning task is under-researched. Fielding-Wells (2010) shows the difficulty that nine year olds have to relate questions, evidence and conclusions in the planning stage of statistical inquiry, but there is a dearth of work that explore reasoning in and between the Problem and Plan phases of the investigative cycle.

My study provides rich information about the discourse of teachers as students at the very initial stage of statistical reasoning and investigation. I analysed the students' structuring of the data context through the questions they formulated as well as the subsequent discussions based on their questions and on their methods of gathering data from the internet.

The students' failure to consider "reasonable price" as a signal to compare and summarise price data of cars, made me aware of asymmetry between the notion of typicality or representativeness in colloquial discourse and in statistical discourse. I chose the term 'reasonable price' because it was indicated in statistical education literature as a colloquial realisation of the mean. In order to gain better understanding of this asymmetry, I analysed a second episode from the course, namely the students' discussion of the meaning of the mean algorithm. Different from other studies, I probed their understanding of the syntax of the mean algorithm, rather that the semantic meaning.

Through the two episodes I probed statistical reasoning from two directions. On the one hand, the complex everyday context had to be structured statistically in terms of reference classes, comparison of many observations and identification of sources of variation, without the explicit use of statistical tools. The process of grasping the system dynamics of the data-context of car prices required abstraction and objectification of price-relevant features of the context. On the other hand, the syntax of the mean, with which they were familiar had to be disobjectified towards its colloquial use and re-abstracted as a statistical object rather than a mathematical object. The task of structuring the context is one that every statistical investigator has to do, but the interrogation of the syntax of the mean is specific to the kind of work a teacher must do in order to prepare to facilitate reasoning about statistical tools.

I identified similar discursive routines in both episodes, namely evaluation routines and exploratory routines. I interpret my findings against the research I reviewed in Chapter 3, namely research about text comprehension and the notion of 'The Immersed Experiencer' (Zwaan, 2004); research about formation of categories; and research about everyday reasoning. I will argue that persistent idiosyncratic reasoning as reported in statistics education literature, can be explained by discursants' inevitable immersion in the data-context, which tends to lead to the formation of role-based categories as reference classes and consequently evoke deterministic, deontic reasoning.

10.2 The objects and narratives of discourse about a data-context at the start of statistical enquiry

To answer my first research question, I will integrate the findings in Chapters 6, 7 and 8. These chapters follow the discussions about the data-context of used car prices. In Chapter 6 I reported on my analysis of the students' questions based on contextual information (anecdotes from blogs as well as articles in financial magazines) and a multi-variate dataset. I placed their questions on a continuum constituted by evaluative questions and exploratory questions. The evaluative questions were off topic or irrelevant; non-mathematical or non-investigative questions compared to the hierarchical frameworks proposed in statistics education research (Arnold, 2009; Allmond and Makar, 2010). The collection of exploratory questions map onto the types of investigative questions described by Graham (2006) and Pfannkuch and Horring (2005), namely: summary questions, comparison questions, and relationship questions. If we accept that the formulation of questions involve reasoning, the implication of the categories described in the literature is that better questions indicate better statistical reasoning. Commognitive analysis of the discourse that surrounded the questions posed by the students in my study, revealed that focusing on questions alone as intersection between statistics and context does not provide substantial information about the status of students' statistical discourse.

I framed the students' questions as narratives embedded in colloquial reasoning and influenced by the context, rather than objects lacking in some way in anticipation of statistical reasoning. I used Kuhn's (1991) descriptions of different kinds of evidence to classify the questions (as evaluative questions or exploratory questions) based on the kind of answers or the evidence that would reasonably be required for closure.

Evaluative questions were deeply concerned with social relations in the imagined context. These questions can plausibly only be answered by an authority or the authority of personal experience. These questions are clearly not statistical questions, but they are not merely idiosyncratic. Discourse analysis required me to pay close attention to word use, and revealed a strong concern with issues of social trust. Words like "check", "believe", "approve", "bias", "credibility" and "reliability" peppered the collection and foreshadowed subjective evaluation narratives in the subsequent discussions. Reflecting on the implications of these questions for statistical reasoning convinces me that the dispositions of the students who formulated these questions are indeed critical, sceptical, curious and aware, seeking deeper meaning, practically logical, and engaged – all are dispositions required in dimension four of the model of statistical inquiry (Figure 1, Chapter 1). But for these discursants variability and uncertainty in the imagined context seems dangerous, and deterministic certainty seems needed in order to act prudently in the context. Questions about the role of the Automobile Association to verify records of used cars (E1, Table 8), and the existence of a market standard or an ombudsman (E3, Table 8) indicate that such authorities may signal islands of constancy in a variable and confusing context.

Commognitive analysis of the discussions of the students' questions (reported in Chapter 7) enabled me to identify routines that can be described as discursive deeds, aimed at effecting change in discursive objects. I identified two deed routines, namely social evaluation and action evaluation routines. Social evaluation routines embedded the written evaluative questions in expanded narratives about the seeming unfairness of agents and market tools. In particular the application of a statistically derived norm such as the depreciation table in the face of huge variability was not trusted. Within the action evaluation routines, sources of variability like mileage and year of manufacture were tied to subjective judgments of the resulting condition or value of a used car. Yet another round of discussion was analysed, this time interrogating the students' narratives about the relationship between price and mileage (kilometre reading), year of make, and model of used cars. The task context of these discussions was different from that of the discussion of their questions. Where previously the contextual discussion was framed by colloquial information about the data-context in addition to a dataset, this time the students were asked to access a website to gather data about prices of used cars. The search options on the website required decisions to order the importance of concomitant factors. In Chapter 8 I reported that the subjective hold of the context during this round of discussions was still evident from the students' evaluation narratives, although a shift toward exploratory narratives occurred.

The question arises why evaluation routines were so persistent in the formulation of contextual questions, the discussions of the questions and the discussions of concomitant variables. My conclusion is that the students' immersion in the data-context, through the interaction with the contextual information in the support materials, led to the structuring of 'used cars' as a role-governed category. This is the primary discursive object that emerged in the discourse about the data-context.

10.3 Shifts between colloquial and literate statistical discourse

Shifts between colloquial and literate statistical discourse seem to be related to category formation. In Chapter 3 I reviewed literature about the formation of categories and category-based induction. The formation of suitable categories as reference classes is necessary for informal statistical reasoning about contextual aggregates, and typicality and representativeness in data-contexts. The literature review shows that artefactual categories are harder to construct than categories of natural kinds. One reason for the difficulty is that most artefacts are made for a purpose and hence form role-governed categories. I did not know of this distinction and its implications at the time of teaching, or indeed until I wanted to understand why the students in my study struggled to reason about price and its concomitant variables as I expected them to do. Role-governed categories relate a person to constituent objects by means of the role of the objects, as opposed to their objective features. Role-governed and feature-based categories are qualitatively different, but exist on a continuum, since any category of objects also have features and any object can be posited in a role (Rein, et al., 2010).

10.3.1 Constituting reference classes as feature-based categories

The category ‘used cars’ is role-based since it relates a car to a current owner and a prospective buyer. All used cars share this role property while they vary hugely in their objective features. According to research in category-based induction, role-governed categories tend to yield ideals as exemplars for comparison to other category members, rather than a central or typical member. The students’ persistent action evaluation narratives can therefore be explained by concern with finding a used car that is closest to an ideal used car in a role-governed category. In the statistical reasoning literature, reasoning about ideal or extreme cases constitutes a local view of data. The ideal used car was realised in different discourses. The most subjective discourse was a value-for-me discourse where the ideal car is as close as possible to a buyer’s dream car and yet falls within the buyer’s budget. Slightly more objective and open to local endorsement was the identification of an ideal used car in a discourse about value for money. In the value for money discourse features of used cars started to emerge as the basis for comparison as the best car would have the lowest price for the most desirable features. Further support for my conclusion that ‘used cars’ was constituted as a role-governed category, is the students’ resistance to the use of the depreciation table as central values for comparison of prices. In their narratives they repeatedly indicated that the use of such a market standard is unfair in the face of the huge variability in the condition of used cars. For example, they were not willing to pay the same price for a car that was made in the same year, but sold during different times (January and December) of the year, or to accept the trade value as a fair price based on average kilometre reading if the car was driven on difficult terrain. As predicted by the literature about formation of role-governed categories, it seems that a typical used car was a misnomer for them, and hence a reasonable price that is a typical price could not emerge.

Reflecting on the difference between ‘cars’ as a category and ‘used cars’ as a category also holds answers for the social evaluation routines that persisted in the discussion of the data-context. The category ‘cars’ is closer to a feature-based category than a rule-based category. Cars share many distinctive features and it is possible to conceive of a typical car (possibly a sedan) and a non-typical car (like a two-wheeled car). Cars remain cars whether they are offered for sale or not. In contrast ‘used cars’ relate the object car to two social agents, namely a current owner and a prospective buyer. Depending on the context, ‘used car’ may also relate a prospective buyer to a seller that is not the current owner in the same sense as the owner who used the car. The increasing distance between the prospective buyer, who wants to know what she buys and the owner who ‘knows’ the car introduces a layer of social complexity. As indicated by Cummins (1995) and discussed in Chapter 3, social situations tend to evoke deontic reasoning, reasoning about obligations and permissions. I propose that the social evaluation narratives of the students were realised in order to detect violation of social obligations as a consequence of structuring ‘used cars’ as a role-governed category. I want to go further by suggesting that even where social relationships are not salient in a data-context, if the reference class is constituted as a role-governed category, reasoning about social obligations is likely in incipient statistical discourse, and hence local views of data and idiosyncratic narratives.

Categories are discursive constructs and open to change. Rein and his colleagues (2010) showed that contexts can be experimentally manipulated to evoke role-based or feature-based categorising. This implies that contextual framing influences categorising toward either end of the role-governed to feature-based continuum. Bakker (2004a) advises that students need to adopt the role of data analysts rather than actors in the data-context in order to shift to objective comparison of distributions of measurements. My analysis of my students’ contextual narratives confirms this requirement also for the formulation of questions and getting to grasp the system relations at the start of the PPDAC cycle. The role of context analyst is not intuitive, but seems necessary to formulate exploratory questions about the data-context.

Bakker (2004a) further suggests that teachers should look for contexts that elicit objective and precise arguments. While I agree with the argument that some contexts may be more suited to elicit shifts to statistical reasoning, Zwaan (2004) suggests that spontaneous immersion as an actor in a data-context is ubiquitous and an inevitable first orientation toward text. Therefore I suggest that the initial narratives of students in response to ill-structured and context-based problems are more likely to be evaluative than exploratory, despite the specifics of the data-context or the task-context. Despite the adaptations which Bakker made in his tasks to elicit the role of data-analyst, about 43% of the arguments that his participants used to compare two distributions of data about batteries remained evaluative and aimed at decision making in the data-context. Bakker's examples of "low rate" arguments (2004a, p. 117) indicate that such arguments were based on ideal values (e.g. the brand of battery with the highest value is the best) and frequency arguments based on arbitrary intervals (e.g. the best brand has more values in a certain interval than the other brand). He indicated that the assumed role of such reasoners might have been that of buyers of batteries. I argue that this assumed role implies that 'batteries' was constituted as a role-based category, rather than a feature based category. Batteries are artefacts that have the purpose of "making things work" and if one buys a battery the purpose provides criteria for an ideal. In Bakker's study the ideal was either consistency or longevity. These examples fit well with my hypothesis that spontaneous immersion in the data-context seems to elicit the formation of role-governed categories, rather than property-based categories as reference classes. Role-governed categories are not conducive to statistical reasoning, since they evoke local views through comparison to ideals; and judgment instead of feature comparison. A further implication of the fact that the same collection of objects can be constituted as a role-governed or a feature-based category is that re-constitution of categories is possible.

10.3.2 Constituting reference classes as feature-based categories

During the process of teaching and learning idiosyncratic discourses shift toward literate statistical reasoning. My interest was to determine how and when the discourse shifts. Close attention to word use as required for commognitive analysis enabled me to identify shifts within the evaluation narratives of my students. In Chapter 7 I reported on the use of the words *value*, *cost* and *price* as indicators of shifting narratives. The use of 'price' as a subjective measurement of value emerged from narratives about price as value-for-me and price as value-for-money. From a category-based perspective, price in these narratives was about the relationship between a used car and a buyer, and hence about a used car (or used cars) as member(s) of a role-governed category. In Figure 14 I presented a pedagogically desired trajectory for the shift from evaluation narratives to exploration narratives about price. In this trajectory, narratives about price as value-for-money have to shift to narratives about price as an objective indication of the relative worth of a used car in an aggregate. I proposed that the final abstraction of price was dependent on this shift and that the shift in discourse entailed the re-constitution of 'used cars' as a feature-based category. In my study, price eventually emerged as an objective property of a used car in narratives about the relative worth of cars, but not as an abstract measurement. Yet, the narratives about the relative worth of cars were sufficiently alienated and feature based that evaluation narratives started to give way to exploration of properties. I propose that the narratives about price as the worth or value of a used car relative to other used cars in the constituted reference class, show that 'used cars' had been re-constituted by the students as a feature-based category. When features like kilometre reading, year of make and model became the objects of their discourse, the students' narratives became exploratory. One argument in particular occasioned a shift towards exploratory narratives in the discourse. I analysed the cause-effect reasoning in the intense argument to relate kilometre reading to price in Chapter 8 (See Figure 18). The argument started as one about the ideal used car but ended with the conclusion that any car with a lower kilometre reading than another car in the same reference class can be expected to have the higher price. I propose that the argument was initially deontic, because it involved an obligation to determine the ideal used car in a role-governed category. The deontic argument can be framed as: If you buy a used car, you are obliged to buy the best used car for the lowest reasonable price. My analysis of the students' causal reasoning illustrates how reasoning about alternative causes for a used car to be good, if not the best, and disabling conditions that systematically eliminated the proposed alternative causes led to objective consideration of variables that influence the price of used cars. Along with the objectification of the data-context and the re-constitution of 'used cars' as a feature-based category, comes the possibility of direct and objective comparison of prices of used RunX cars to a representative central value.

10.3.3 Narratives about the data-context in the disobjectification of the mean

I discussed the narratives to disobjectify the mean algorithm in Chapter 9. Here I would like to integrate the findings with that of the discussions of the data-context of used cars. The discussion about the meaning of the mean algorithm was a challenge to the students to disobjectify and re-enact the mean algorithm. For this discussion, the students were free to evoke their own data-contexts if they wished to. Three different contexts were evoked: the heights of an imaginary group of learners (RK); the imaginary marks of one student’s (NM) imaginary learners, and the context of prices of used cars (MM). The students used these contexts to construct disobjectified narratives about the use of the mean, elaborating on the semantic meaning of the mean, rather than the logic of its syntax. In terms of syntax, both the median and the mode have the advantage of accessible logical actions as disobjectifications.

Similar to the discussion of the data-context of used cars, the evoked data-contexts influenced the students’ discourse. In particular, I propose that NM’s extreme evaluation narrative in which the mean itself was judged as average in the sense of not good, is the result of her constitution of the marks of her class as a role-constituted category. In her role as teacher, marks relate the quality of teaching to the performance of the learners. This can explain why she remained concerned with an ideal mark, and unwilling to discuss the variability of the imagined marks in objective terms. Had she reasoned about objects in a more salient feature-based category her narratives might well have been more exploratory.

Less extreme, but also indicative of an evaluation narrative about objects in a role-constituted category is SM’s short narrative about the mean as the average price that a buyer will pay him as the imagined seller. In the earlier discussions of the data-context SM persisted with evaluation routines about used cars, realising ideal objects in his narratives, or opting out of the exploration when his ideal seemed out of reach. In the light of his previous narratives in the context of prices of used cars, SM’s disobjectification of the mean as average may simply tell the story of a seller who is willing to fairly negotiate down from his ideal price. In such a narrative, the mean is not a representative price, but an acceptable or good price for both parties.

Narratives in which the mean was disobjectified as the average in the sense of a representative measurement were exploration narratives, with the purpose of telling an objective story about the context. The story about the context was “a vague impression” that endeavoured to exclude extreme observations. As I indicated earlier, I was surprised that the students in my study did not abstract the term reasonable as a statistical central value, since statistics education literature shows that reasonable is a disobjectified realisation of the mean. In hindsight I should have expected the asymmetry. The mean can be judged as reasonable, depending on the perspective of the reasoner. Such judgment may be completely deterministic and subjective and not a disobjectification of the mean in the sense of finding a previous discursive act of meaning (Sfard, 2008). The mean as an abstraction of the colloquial term ‘reasonable’ is even less feasible, since people reason from many different premises about what is reasonable in different contexts and convergence to the abstract notion of a statistical mean value is unlikely in colloquial discourse.

10.4 Constraining and productive narratives in the shift from colloquial discourse to statistical discourse

My reflection on the findings related to the discussion of the mean algorithm serves to support my argument that the lack of objectification of reference classes in data-contexts inhibits the development of statistical discourse. Both in the discussion of the mean and the discussion of a reasonable price for a used car, reference classes as role-constituted categories seemed to prevent alienation of features as variables and the abstraction of measurements. Without the abstraction of measurements the mean has no advantage over the median or the mode as a middle value, and a subjectively ideal value trumps a statistical summary in context. I further suggest that idiosyncratic narratives about the mean as described in the literature are indicative that the underlying data are embodied in role-constituted categories. Productive narratives in both discussions were those that suggested alienation of the reasoner from the imagined context. Along with alienation narratives seem to shift from risks and consequences of imagined action to relationships between objective properties of the context. Sfard (2008) argues for shifting discourses towards new discourse, but also says that there is not a smooth development from practical action discourse to exploration discourses that we know as sciences. So while the semantic meaning of the mean average in relation to visual realisations like graphs seems intuitively close to people's colloquial understanding, the syntactical meaning is more difficult to develop.

10.5 Implications for research

10.5.1 The affordances and constraints of commognitive analysis

Commognitive analysis of the discourse in the course I taught enabled me to suggest why idiosyncratic narratives or non-questions emerge in statistics classrooms. It also enabled me to describe routines that can be expected in the discourse of statistical novices. However, my study was too small to generalise to discourses in other classrooms, and my findings are in need of wider endorsement or refutation. For this purpose I presented extended transcripts in the spirit of the principle of completeness. My role as lecturer certainly influenced the discussions and the emerging discourse in the classroom. In another classroom, where the lecturer has a different focus and frames the learning task differently, the discourse may develop differently. I endeavoured to alternate my perspective between that of insider and outsider regularly through critical consideration of my own word use and reflection on my intended word use during teaching. The principle of operationality requires that the commognitive researcher disambiguates and operationalises her own vocabulary in order to guard against misunderstanding. I found this principle very challenging. Firstly, English is my nine-to-five language and not the language of my "discourse-for-me" (Sfard, 2008, p.297). Hence, formulation of unambiguous descriptions of the narratives and routines I identified was a challenging task. So too was the formulation of narratives to relate my commognitive findings to findings in statistics education literature that are reported in a different discourse. I am aware that my research narrative is dense and in the reading often wordy. This is a result of my effort to be unambiguous in my second language.

The theory of commognition is much richer than I could make use of in my research and provides more analytical tools than I could use in the scope of my study. Some of the tools were not easily applicable to my research. I will discuss the difficulty of distinguishing between narratives and routines, and the problem of constructing realisation trees.

Commognition is a strict theory that does not allow a researcher to hide behind unoperationalised terminology. It demands thorough comparison to definitions and properties of commognitive objects. As far as I am aware my study is first to make use of commognitive analysis of extended contextual discussions in a statistics classroom, so that there are no precedents for the identification of routines. Re-interpretation of short narratives reported in previous studies of statistical reasoning provided pointers to the possibility of routines like “compare data to an ideal value” or “compare data sets by means of an ideal value” or “compare data sets by their middle values”, but the paucity of rich descriptions of discussions of data-contexts means that I had little pointers to routines at the start of the statistical investigation cycle. The repetitive narratives which I identified as routines after they were confirmed in the three sets of analyses of the students’ discussions are therefore local routines. My interchangeable references to evaluation narratives and evaluation routines, or exploration narratives and exploration routines are indications of my awareness that they are in need of wider endorsement. However, to strengthen my argument that they are likely candidates for routines of statistical reasoning, I anchored the routines in plausible causes for their emergence evident from literature about categorisation and everyday reasoning.

The commognitive definition of a concept as a word or another symbol and its uses is productive and enabled me to identify price and value as different concepts in the students’ discourse. However, delineating discourses by means of word use (hence concepts) was a difficult task, as Sfard (2008) warns. Sfard uses realisation trees in order to disobjectify literate mathematical concepts and identify discursive objects. I could not construct realisation trees from my students’ discussions. There may be various reasons. It may for example be that their mainly verbal realisations did not yield enough information for the construction of trees. I think, however, that the discussions about the data-context entailed objectification without the benefit for me as researcher of widely endorsed discursive pathways for the objectification of data-contexts. However, I adapted the construct of realisation trees to produce discourse maps as a tool to delineate colloquial and statistical discourses. My definitions of colloquial and informal statistical discourse are theoretical propositions that need further refinement and specification. In the endeavour to delineate discourses, the different realisations of the same word (concept) were allocated to different discourses. In order to do the allocation I included my interpretation of the purposes of the realisations as family resemblances. For example, if the purpose was experiential comparison of prices, I assigned the realisation to informal statistical discourse, while realisations aimed at decision-making that bypassed a meta-discourse of comparison, it was assigned to colloquial discourse.

10.5.2 Focuses for future research

Shaugnessy (2007) indicates the need for research of classroom discussions from a discourse analysis perspective. My study raised specific questions for the use of discourse analysis in statistics education. More research is needed about the initial stages of statistical reasoning in contexts. Following Sfard’s (2008) metaphor of research as an archaeological excavation, my contribution was to lay out some excavated shards and suggesting that they all belong to a vase. More research is needed to fit the shards together, or to compare shards and suggest the possibility of different pots. As I indicated earlier, more research is needed to identify and compare discursive routines in early statistical reasoning in data-contexts. The routines identified in my study were described in terms of ‘how’ and ‘why’ they emerge, but more research is needed to confirm their applicability conditions, that is, when they are likely to emerge.

I identified the problem of overlapping scales of quantitative and qualitative measurement evident from the use of comparison terms as an important aspect that leads to incommensurable discourses and prevents shifts from colloquial to informal statistical discourse. More research is needed to identify instances of word use that indicate comparison on different scales. Such word use may play a role in the problem of alienation of measurements.

A related issue which I did not address in my research is the quantification of comparisons. For example, when my students reported that female students read faster than male students based on a difference of ten words per minute between the means, the narrative is statistically true, but does not satisfy in the context. Ten words is hardly a line of text per minute, and not a difference that counts in the context. Therefore, I want the students to report that the reading speeds are very similar, rather than different. Comparison of variable data within and between distributions is at the heart of statistical investigation. It therefore is important that the field generates more information about the use of comparison words in data-contexts.

I would welcome more research about category formation in data-contexts in order to engage with my research finding that feature-based categories and role-governed categories elicit different routines. Such research may help educators to select productive data-contexts and it may have a bearing on statistical discourse about sampling.

Commognitive research to chart shifts between informal statistical discourse and literate statistical discourse is necessary to draw together previous research findings. We do not have a clear narrative as researchers and educators about words and their uses that occasion shifts to literate statistical discourse, or responses at the higher levels of SOLO based models of statistical reasoning.

With regard to the well researched concept of the mean average, more research is needed to establish a relationship between the mean algorithm as a procedure and as an object. I proposed in Chapter 9 that the mean as an object can be defined as a zero for the measurement of variation. Further research is needed to establish if this is a didactically productive definition. Commognitive research about the emergence of deviation narratives in relation to the measurement of variation will deepen our knowledge for teaching the mean meaningfully. The work of Lehrer and Kim (2009) in this regard is promising, and may produce data for commognitive analysis.

10.6 Implications for teaching and professional development

My research confirms that structuring a data-context in order to grasp the system dynamics is a complex and time consuming task. I reflected on the benefits of the extended discussion for meaningful teaching of statistics. The most important benefit for me was the realisation that no context is statistically given. I did not expect a prolonged process of discussion and argumentation to reach shared focus about the relationships between price, kilometre reading and year of make. Although I did not analyse the students' views about the experience I can informally report that they said they did not know one could investigate the context statistically when one wants to buy a car. The experience of structuring the data-context was new to them.

My findings can serve as guidelines for engaging with the structuring of a context as a learning task. In the wake of my research I accept that colloquial discourse about a data-context is inevitable and if it is ignored or discounted as idiosyncratic, may influence the complete investigative cycle. Teachers may benefit from knowledge of category formation in contexts in order to anticipate the emergence of evaluation routines. I propose that purposeful discussion to background emerging issues of social trust and personal decision making in favour of exploration and comparison would be a productive didactical move to support shifts from colloquial discourse to literate statistical discourse.

The students in my study did not hesitate to admit that their evaluative questions are not statistical questions. They were able to suggest plans to gather data to answer exploratory questions as can be seen in the full transcripts in Appendices A and B. Yet they formulated the full range of subjective and objective questions. I argue that unresolved subjective questions may influence their analysis of objective data, for example through persistent evaluative routines. Further dedicated discussion of sources of variability in the context, and ordering the importance of those sources, may serve the purpose of emphasing and retaining the sense of uncertainty and variability throughout the process of statistical investigation, so that informal inferences are more likely to be hedged and reported in tentative terms.

My research sensitised me to my own word use when teaching. Knowledge of deed routines and exploration routines specifically made me aware of talking about discursive objects as if they were objects in the world. I am no longer using the term average as a realisation of the mean in my teacher talk, and insist that my students operationalise average in their discourses. This allows me to engage in discussion with education students about the pros and cons of using the median or the mean as a middle value, and the way in which the calculation of the mean produces a middle value.

Teachers who pay attention to their students' word use may become aware of overlapping measurement scales hidden by uncritical use of comparative terms. Such awareness may provide ways to identify commognitive conflict and ontological collapse during classroom discussions. Discussion of the difference between personal judgment and objective description may provide a way to turn such instances into productive discussions.

Through this study, statistical reasoning has emerged for me as a way of seeing the world. As a mathematics teacher educator it sensitised me to the need to develop variation talk in my mathematics teaching. On example will suffice. I am now aware that my discourse on functions lacked narratives about co-variation and the gradient as a *measure* of co-variation. Standard practice in providing realisations of functions as tables of values is to unitise the change the independent variable, so that the only reasoning task to determine the gradient, is to analyse the change in the dependent variable. This constrains the discourse about co-variation to discourse about variation in one variable, and makes the phrase 'rate of change' meaningless. Although not reported in this study, I was confronted with the result of my impoverished function narratives when my students had to interpret the gradient of a regression line in context.

Finally, a meta-view of informal statistical reasoning as embedded in colloquial reasoning may help teachers to grow the new discourse, statistics, by gradual transformation of a discourse in which their students are already conversant. Sfard advises: "One way to preserve discursive continuity is to "grow" new routines in conjunction with familiar deeds that the new routines are supposed to enhance" (Sfard, 2008, p.260). The inevitable commognitive conflict between discursants in shifting discourses provides opportunities for meta-level learning and meaningful teaching of statistics.

Appendix A: The support materials for the discussion of the used car context

Appendix B: Full transcript of Session2: Discussion of questions about the used car market

Full transcript: Session 2: Discussion of questions about the used car market		
Turn	Discursant	Utterance
1	Lecturer	Looking back at the problem of previous week. Ask them to look for situations where stats are used to make decisions.
2	Lecturer	So, the main thing that I think we have to start finding out is what can statistics do and what kind of questions can we answer with statistics…So, I found on the internet, after a a news uhm item this week, information about car values. And I know the men here like cars, and the women here like cars. And we like to have smart cars, don't we and new cars. But the information here says we have to consider: Is a car an investment? So I am giving you the information I found, from a website, car prices…and I want you to spend the time to look it through, scan quickly to see what people on the internet said about prices of cars. After that you will see there is a data set that I got for you, of second hand prices of one type of car, a Toyota RunX. After that, page 11, are official data that says how car prices devalue. So, your first task now… is: get to know the context, what happens to the value of second hand cars? Is it worth buying a new car? I am giving you about 20 minutes to look at the data and you come up with questions that you can ask that we can answer with the data. So, I want a list of questions that is answerable with the data that you have and that is about the situation with second hand cars.
3	Lecturer	And I want the group to work so that everybody in the group stays together. I suggest you talk about what the data makes you think of, then you say OK, jot down your own questions, and then you come back to the group… …So I will give you 20 minutes for that…. and then I will bring it back to the class and we can discuss…
4	Lecturer	I want the data to speak to you now, about what's going on with second hand car prices
Session 2: Discussion of Group B: Discursants KH, SM, RK, GK		
5	GK	What is the difference between the asking price and the selling price? Is it just what they ask?
6	KH,SM,RK	Yes (continue to read through column headings) (KH hesitates and frowns)
7	Lecturer	(After inaudible comment by GK) I want you to think about the situation of buying a car. Here you have information about secondhand cars. Read what the people say. Look at the data and make a problem about it. What are questions you can ask that we can answer with the data? OK? (They read through in silence)

Session 2: Discussion of Group A: Discursants: NM, MM, GG, SDS		
8	Lecturer	(to NM) I don't want you to spend too much time on this. I want you to scan and say, hey this is interesting, what does that mean…
9	NM	This is kilometers (refers to table of cars from website)
10	SDS,GG	Mm?
11	NM	Mileage. This is kilometers.
12	GG	Yes, they obviously don't have mileage information.
13	NM	Are we supposed to get some other information?
14	GG	No I think they just see the mileage and offering a price for it…..But if you know the price, how can you work out the price from it, from the previous ones (referring to the depreciation table)
15	NM	OK?
Session 2: Discussion of Group B: Discursants KH, SM, RK, GK		
16	Lecturer	So what have you learned about the second hand car market?
18	KH	Some models depreciate more than others.
19	GK	The second hand car depreciate less than the new car
20	Lecturer	You mean the older a car is the slower the value depreciates?
21	GK	And then the other thing is what they they they advertise is not how it is (inaudible)
22	Lecturer	In what sense?
23	GK	In the sense …that…ah…they give you the things you want to hear…they'll tell you …ah…about the kilos, about who was the previous owner, and then they don't tell you about…what's that now…
24	RK	If it's been his for a long time
25	Lecturer	Mm, they spin you yarns?
26	GK, RK	Yes
27	Lecturer	Yes, yes I think you are right. Look at the information, be careful, there is an opinion here that wow the car is great value. I think carry on with what you are suggesting, but ask questions now…Maybe on your own, get a range of questions…
28	KH	We must get proper questions, now. Which can be answered with this data.
29	SM	It is not worth to buy a car
30	KH	Mmm [tentative]
31	GK	What do you mean it's not worth to buy a car. You mean it's not worth to buy a new car or an old car
32	SM	It does not matter, if it is a new car or an old car. As long as it is a car
33	GK	Mm?
34	SM	Ja
35	GK	The reason it is (inaudible) the petrol price?
36	SM	The interest rate, service…cars are very expensive, OK? And then, when you buy it it depreciates.
37	GK	Mmm. But is it true the older cars they depreciate slower than the new cars?
38	KH	(Nods affirmatively). They say if you drive a new car out of the showroom, as you drive it out it depreciates.
39	SM	MmMm
40	GK	So the older car, isn't that the same thing?
41	KH	No, because it already has depreciated

42	GK	(Frowns, turns her head away)
43	KH	(inaudible) You can buy an old car, an old skedonk for a R1000, it cannot depreciate much more.
44	GK	Maybe if we analyse this last…we can get a good question…
45	GK	Formula for the trade price…
46	KH	There isn't a standard depreciation. Different cars depreciate different amounts
47	GK	…about when in year a car is bought compared to the trade price
48	KH GK	Who is disadvantaged by this guide?
Session 2: Whole class discussion		
49	Lecturer	(At the white board) I am going to try and group your questions according to what type of question it is.
50	KH	Can I just ask at what level we are aiming these questions? Inaudible.
51	Lecturer	It's for you. I want you to understand what influence car prices, how prices vary.
52	DH	The influence that petrol price has on car prices. And then, does it influence new car sales and secondhand car sales
53	SM	Cost of maintenance of new car and old car
54	Lecturer	(Writing and revoicing questions from the class) How credible is information about old cars? E.g. has it been in accidents? Is the mileage accurate? How many owners? Is there still money owed? Who the owner was.
55	Lecturer	So it is questions, that if I know the car, I want to know extra things
56	RK	Hybrids/Cars that don't run on petrol
57	Lecturer	If a technology gets replaced, what happens to the old cars?
58	DH	All the repossessions that happen. They are flooding the market
59	Lecturer	How does repossessions, flooding of the market influence the price?
60	KH	We were interested in how they work out the depreciation. The depreciation varies from car to car, so in this guide, how do they actually come up with their figure?
61	SM	If you go and buy a secondhand car, do we have all the information
62	GK	About the apparent pitfall…how is the…pitfall….disadvantaging the buyers? You know, that they say the price, uh, they compare the prices of the new car and…uh…the old car within a period of time. Maybe they work out the interest, I mean the depreciation. They say the depreciation ranges from January to January the following year. If you buy the [new] car in between the year, and then you want to sell it for the next year, where are you going to fall there? Are you going to be disadvantaged?
63	Lecturer	Are you saying, if you buy a car in 2005, and you buy it in December 2005…
64	KH, GK	Will we sell it at the same price?
65	Lecturer	When bought new – Jan v Dec 2005, what will be better buy in 2006.
66	GK	Maybe, can we say are the cars worth the same price? Are the value of the cars the same?
67	Lecturer	(Writing): Are those cars still of same value? OK, and I think you have identified here (referring to questions written down in Turn 54), that one thing that plays a role there is the mileage.
68	GG	But, if a person bought it in January, and didn't do much, say, ja, there is

		no mileage on the car, then it is probably better to take the one from January.
69	Lecturer	Why?
70	GG	Well, if…the one in January…doesn't do much driving…and the one in December has decided to do a road trip for example,…so it's mileage, ja…and I suppose it's also the condition of the car as well.
71	KH	They don't look at that. They take a look at the car, take out the table and say this is the trade value of the car.
72	RK	What if the cars are used on different terrains, you know like…the person uses the car on a good surface, and then a person uses the car for a short time…driving you know, not on the road…
73	RK	I am saying if they are using mileage to to to assess the value of…then they might be misleading this customer.
74	Lecturer	Any other questions you want answered?
75	RK	I wonder how reliable the depreciation calculations are. When you say a car depreciated by 15%, how reliable is that information?
76	Lecturer	Reliable, accurate?
77	NM	There is a statement on page one which says...mm… lost up to half of their value in one year. My question about this statement is, is this good news to the buyer or the seller of the secondhand car?
78	Lecturer	OK. Would you say good or bad for the buyer of a new car?
79	All	Bad
80	Lecturer	Bad? Do you agree?
81	SDS	OK, if I was trying to buy a car, and I was getting this data, I would look at specific criteria. Like if I was looking for a specific colour, I would choose all the ones with the colour I want and look at what is an appropriate price. Not too cheap so that there is probably something wrong with it, also not something too expensive, because then I can't afford it. So I'm looking for, I wanna say an appropriate price for each criteria–colour, or model, or area…
82	Lecturer	(Writing) What is an appropriate price for different criteria. Um, what shall we call those? Variables?...you mentioned you don't want to pay too cheap, cos then you would suspect…
83	SDS	I'd be suspicious, ja.
84	Lecturer	…and not pay too much, then you'll think you'd be done in.
85	SDS	Mm!
86	Lecturer	There was a question from SM
87	SM	Why, why does the second hand car depreciate slower compared with the new cars?
88	Lecturer	(Writing) Why, and does, a new car depreciate faster than a second hand car?
89	GG	What is the most reliable brand of car, to buy secondhand?
90	RK	Is it its resale value or more (inaudible) …reliable? 'Cos if it is reliable then it won't break down.
91	GG	Mmm, so probably resale then
92	Lecturer	Which make has the best resale value?
93	SDS	With the two graphs that we dealing [with], I'll ask how does a car's mileage affect its or compare to its price. Because we expect as the mileage increases it will be cheaper, but if you look at those [data]

		tables it's not the case for all cars
94	DH	Uhm, one other thing that relates to where the data is from. Actually, are the people who produce these tables real statisticians, do they work for the car dealers themselves or do they work for the second hand car dealers? Last week we had one set of data and we argued both sides from one set of data. So….who's actually gathering the data and analysing this, are they statisticians, or are they just smiling…because of (inaudible).
95	Lecturer	Who gathers the data and who produces the statistics? And you are saying they do it for their own reasons…
96	DH	Do they get paid by the new car dealers, lots of money to make it look better, or are they paid by (inaudible)?
97	Lecturer	Are you interested in that?
98	All	Agreement and interest
99	KH	It comes back to the question how they are calculating the depreciation
100	RK	Ja, cos I think that obviously they want …(inaudible)
101	KH	If you can't find the validator in the house, and you rely on this one set (of data), how do you know where it comes from, how did they get it…
102	SDS	Is there a board, Mrs Lampen?
103	Lecturer	There is an association of second hand car dealers, yes, and uhm…the second hand car prices are, uhm, fixed within a band… While you are interested, there was a fight in the newspapers about a year ago, exactly this. Second hand cars, new cars were very cheap about a year or two ago, you could get cars with everything that opened and closed, cd-players and so on, for the price of what you would expect for a car with nothing extra. OK? So they tried to sell the new cars, and if they, and then they manipulated the price of second hand cars…said no you can't give more discount than this or this, otherwise people won't buy new cars. Right? So yes, there, there are there is politics, and there are power games for sure. That is why we must be informed.
104	Lecturer	And which of them are now statistical questions? And if they are not statistical questions, what are they? What kind of questions, what shall we call them, OK? Let's quickly think. This one. Does the petrol price influence car sales or the prices of new and second hand cars? Is that a question we can answer by stats?
105	Class	Yes
106	Lecturer	What will you use as variables to compare?
107	DH	You take the petrol price over as long a time as possible; we need new car prices, obviously by some kind of average, initially, you might break it down. And then car sales, how many cars have been sold, uhm, and then split up between new and second hand.
108	Lecturer	OK, so, we can get information about what was the petrol price over a number of years, and then for those same years, look at how many cars were sold, and what was the prices of cars (unclear)….data that have variables.
109	RK	I think that…stuff goes up with inflation, and petrol price goes up and car prices go up, so I don't think you can say it is a direct relation.
110	Lecturer	Ja, OK you say, watch out, any of these may change anyway.
111	RK	Ja
112	DH	You may have to take inflation out of it.

113	Lecturer	OK, but still we can gather data to actually (investigate). This one: The cost of maintenance between an old car and a new car. Can we find data to look at it statistically?
114	Class	Silence
115	GK	Maintenance, uh-uh (negation)
116	DH	Yes. (Some talk about service. Inaudible)
117	Lecturer	OK, so if we look at maintenance, we need to say, what can we measure here to say maintenance. And there is the idea that we can take…what does a car's service cost as…it gets older
118	RC	As soon as you buy a new car, it becomes pre-owned. So will it be the cost of a new pre-owned car or (inaudible), cos you can't have the cost of maintenance of a new car, if it's new, if you haven't bought it. So, it's like, there's a…, you know what I mean…
119	Lecturer	Do you hear what she is saying?
120	RK	Ja
121	Lecturer	The moment you buy it and you've driven 30 000 km to your first service, it's not a new car. Ok, but we have an idea that the age of a car is changing, right? Doesn't matter if you keep the car or it is resold, the age is increasing, and it seems to me you had an idea (pointing to SM) that therefore the maintenance cost is increasing. You had a problem with it, right?
122	Lecturer	OK. So, ja. We can find (statistical) information about that. This one: How credible is information about pre-owned cars?
123	NM	Before you go on, can I just ask a question please?
124	Lecturer	Ja?
125	NM	The thing about financial maths and stats. Ah, somehow that second question [cost of maintenance] to me, it falls into financial math and I'm not sure there is a clear cut. I can't find a clear cut when I am looking at that question.
126	Lecturer	OK tell us more what you think about, what do you think of when you think financial math then.
127	NM	I am thinking…those cars for example, calculating the cost each time, is just calculating just those numbers for each year, and I don't see the…tools from stats that we can use to answer that question…I can't pull out and say, this thing in stats is helping me to answer that question other than the tools from financial maths.
128	Lecturer	Anybody else there? NM is challenging us. She says, what she thinks at the moment when she thinks about stats, she doesn't think she'll answer this by stats so much as by mathematics. And then she thinks financial maths, because that is where we work with formulas…Any other…thoughts about that?
129	KH	Well I think, if you compare them, stats is a good way to compare. So I mean it can be something as trivial as a pie chart, you know, saying this is the cost of the car and this percentage is maintenance, and another pie chart from another agent…
130	NM	It now sounds…it now sounds, those questions, we are forcing the stats
131	KH	Well ja, absolutely
132	NM	It sounds like that to me.
133	DH	But, even…you might use financial maths to generate your columns of

		numbers, but how are you ever going to compare them? You're gonna do a regression on them, you're gonna plot them, you're gonna see what the mean is, then you're gonna do the mean-median-mode, box-and-whisker plot, something that…
134	NM	Can you be specific exactly there and say this particular part…
135	DH	Why should we have to? Any…
136	Lecturer	Hang on, it is a meaningful question. If we have to think, …(inaudible)…what is stats and what is maths. Hey? And I am sensing here that, and tell me from what, that you are saying, well if I want to know what is the cost of maintenance I am going to calculate it.
137	NM	Mmm (agreement)
138	Lecturer	Any other opinion about this?
139	NM	My problem is, can they at least point out… at a particular point and say this tool here, from stats, this part, this comparing, what exactly are you…how exactly, that comparing you will be doing?
140	Lecturer	OK. So in other words, if you've been… now…is there anybody that wants to answer?
141	RK	No, it's just like that, just like she said. You know like using ah…regression analysis, or…whatever
142	NM	Regression line yes. How exactly?
143	RK	No I just…can't give you an example, can't think of it now… binomial, binomial what, you know, those things they use to compare…data
144	Lecturer	NM is challenging you on a very common sense level
145	NM	Yes
146	Lecturer	Don't tell me your tools, I can give you all the formulas from financial maths. You give me the formulas of stats, I give you the formulas back. On a common sense level.
147	DH	It's asking you to make a comparison. It doesn't ask what the cost of maintenance was, but, how do you make a comparison for it, unless you compare it to something else.
148	NM	Cos all I see. I am seeing numbers here. I am seeing numbers from one garage or wherever, and I'm also seeing numbers, yes, it's clear, I can see it with just financial maths. Now that part about stats, I need an example a specific example, it's just…
149	DH	Can we use from another area?
150	GK	…can I just…
151	Lecturer	No, our challenge here is on the car data.
152	GK	Can I just say what I am thinking. I am thinking that…no…what we're comparing, even if we use financial math to get the figures and so on. In the end you want to see what is this figures tell us, how can we interpret them. And then presently there we are talking about the maintenance of this make this make of car, let's say, what ah Renault, ah at the maybe at the age…after five years…ah…if I took it for maintenance it will cost me this much, after five years, after two years, you get what I'm saying, after each year, or maybe two years you service it. And from there you compare…you compare the prices of the … (lots of weighing gestures)
153	NM	I'm looking at the numbers.
154	GK	And you see what they tell you. Do they tell you that this is the best car to buy, because maintenance wise it's cheaper? Or is it more expensive?
155	NM	All I am saying is I can get the answer form just looking at those

		numbers. Done. What else can I do?
156	GK	You want to buy a better car? One that won't cost you more?
157	NM	Yes. The answers are just there. After doing financial maths, I just get the numbers there. This is more expensive, this is less.
158	Lecturer	I hear GK, you are not convincing her.
159	SDS	I think the problem will be the amount of data you have. If you are looking at just three numbers, like you can…three numbers isn't…a need…enough…for the statistical tools to be used. Because like she said, she can just find the answer in terms of what she's looking for. But if she had a whole bunch, like this whole list here, she can't just look at the numbers, it's quite tedious, whereas if you do a box-and-whisker plot, you can see, like this…is way out for maintenance, I can't afford that, and that's too low, I am looking somewhere between, that's where I'll go and look for…to pay maintenance.
160	NM	I am, it's just that I'm also thinking that the garage would have collected the information for me in a way, but OK, it's fine.
161	Lecturer	Anyone here? Not another example, take up our problem. Her basic argument is, I can see the figures here. I can see this number is smaller than that one, What stats do you want me to do?
162	NM	Exactly. Yes.
163	KH	Except this doesn't tell you much. You've to see a proportion of something. So the cost of the car and what proportion of that cost would you then spend on maintenance or whatever.
164	Lecturer	She [NM] says here's the price of the car, here's the bill from the …garage, I can divide...that's maths, that's not stats. That's what she is challenging you…
165	KH	But that's just one car, as SDS says, we have a whole list of cars.
166	GK	Mm (agreement)
167	DH	What's the easiest way to compare? You can compare with the numbers, but what is the easiest way to compare?
168	NM	You know what guys, you're not giving me the part where you tell me exactly the thing that we will be comparing. You just keep on telling me compare, but what exactly, which part? What is it that we're comparing?
169	DH	Different cars and what the cost of maintenance was.
170	NM	You are just forcing this question into stats, where it doesn't…I fell like that. Because at no instance can you say, e.g. this! At this point we will use this tool, it's either a graph or what…
171	Lecturer	OK, let's try to give her a question, that she has to answer and say what is stats, what is maths.
172	RK	I think, you're not going to get, like 20 cars from with the exact same values from 1994 and 20 cars with the exact same values from 2004 and compare them, you know, maintenance cost. You're not going to get that, because it's not an ideal world. You're going to get a car from 94 with this value and this maintenance cost. And…I don't think if you just…if you add them together and spit out a number…
173	NM	What exactly am I going to do? I want that thing that I'm going to do, from stats.
174	DH	Find the average.
175	NM	Average. OK. It's also maths by the way. (Laughs). It's maths!
176	Lecturer	OK let me see if I can get this to a point. See if this relates? If you feel you

		still need to make a point, do it.
177	Lecturer	If you want to answer a question like, let's stay with the cost of maintenance: what can I expect to budget for the maintenance of my car if want to keep it till its 15 years old? What will you do?
178	NM	Take it in for service after a certain period, I don't know how they do it
179	GG	Six months
180	NM	Is it six months
	Lecturer	OK. I'm changing it. I say, you are the person working at the garage and your clients ask you, what should I expect to budget for maintenance if this car gets older and older. Can you tell them exactly?
	NM	(Smiles) No.
	DH	Unless he's got lots and lots of old data.
	Lecturer	Can he ever tell you exactly?
	DH	He can give you an estimate. An informed estimate. He's got lots of good data and he has graphed it and he uses the graph to extrapolate,
	Lecturer	OK, so…
181	DH	…and summarize all that data into a meaningful form to use.
182	Lecturer	What are we after here? Very important. And I am glad for that challenge, keep putting us to that challenge. If the answers are certain, there's no place for stats. Stats doesn't deal with certainty. Voila, NM your car service will cost exactly this next time, and exactly this the next time, and the next time. Whether you are a person who drives over potholes, and whether you're the person who let your car overheat or whatever, the…cost of the service will be exactly the same each time – then there's no need for statistics, right. But if the service depends on things, that vary, and that is not certain, then we start saying one calculation, the answer to one calculation is not giving me the picture.
183	GK	Mm (agreement)
184	Lecturer	All right. So let's remember this: if the answer is certain, or the question is what is the most or the least, what is the most expensive, the least expensive, then we don't need stats…..(unclear) but when the answer is uncertain, one math answer, or ten maths answers can't give us the picture. Then we need stats.
185	Lecturer	I want to push on. The credibility of data. Is that something we can answer with stats?
186	Class	No.
187	Lecturer	Credibility here is a judgement that you make on trust. ….That's not something you can measure, therefore stats can't give you an answer. Stats can only deal with something you can measure – measurement also mean count.
188	RK	Why don't you clarify the credibility by doing a little study into what you're saying.

Appendix C: Full transcript of Session 3: Grasping the system dynamics of used car prices

Full transcript: Session 3 Grasping the system dynamics of used car prices		
Session 3: Whole class discussion		
Turn	Discursant	Utterance
1	Lecturer	Thinking about 2nd hand cars. What kinds of questions can be answered by statistics and what not. Can you give me some feedback?
2	DH	Where there are measures involved or counts.
3	KH	Trends.
4	Lecturer	We can take the used car business as example, but I'm trying for us to get to know in general the kind of questions statistics can answer.
5	KH	Specific kinds of questions, like, what is the average price for an Opel Corsa, that kind of question
6	Lecturer	That it can answer, but can stats answer what car do you like?
7	Class	UhUh (Disagreement)
8	Lecturer	It can't, right! Can statistics answer a question like: What is a reasonable price for a car to pay?
9	Class	Mumble yes.
10	GK	Yes.
11	Lecturer	How? What will you take as…What will you look at, GK if you want to determine what is a reasonable price for a RunX?
12	GK	I will compare the model…of the RunX that I am looking for…thereafter I'll compare the year… that it is made. Thereafter go check on the price I want, the price that is affordable for me.
13	Lecturer	If I think in terms of statistics then, I see…I imagine GK taking the RunX model she wants, a 140i…but just one. Are you going to take a look at one RunX 140i?
14	GK	Noo, I don't think it's a 140i, I take, no I will compare them, you know, I'll take… their year of make
15	Lecturer	OK, so she's not going to take RunX all of them, you'll compare them in years
16	GK	Mm (agrees with me)
17	GK	And thereafter I'll check out their kilometres as we said, how far they've traveled so far…thereafter check the price of each…and maybe the colour.
18	GK	Mm (agreement)
19	Lecturer	Right?
20	GK	Then I take the one that suit me best.
21	Lecturer	But the question was what is a reasonable price to pay.
22	GK	Mm!

23	Lecturer	So do you say your price will depend on the kilometres? What is reasonable will depend on the kilometres?
24	GK	(Looks around for support) Not just that.
25	KH	Yes it does.
26	Lecturer	You are not willing to pay the same price for a car with 100 000 on the clock and one that has 50 000 on the clock…
27	GK	Mm
28	Lecturer	It will depend in a sense on the km, right?
29	Lecturer	GG, your hand was up?
30	GG	Oh no, I was just going to say that…I will first look at the price, then decide on the kilometres. So you look at…what…at the price range you are willing to pay and then you look at the kilometres and all the other variables…(inaudible) what GK says…
31	Lecturer	OK, so, think carefully here. There are two ways that GK and GG propose to go about it. GK says I know which model I want and then I will go and see what can I get for that and what can I expect to pay. GG says I know which model I want and I have a certain budget, I can afford a certain car, a price, so I will fix the price and see what I can get for that price.
32	GG	Mm
33	Lecturer	So these are different ways to go about it right?...Which means the relationships that you set up between your variables are a little bit different. Your classification. So, if we just think of what kind of data, how you must deal with your data to work in GG's way. See if this makes sense. GG says, I have a budget,…so say your budget is what?
34	GG	A hundred and twenty… five thousand?
35	Lecturer	(At white board) Maximum hundred and twenty five thousand rands. And she wants to know what she can get for that price. And now she goes to the website. What should she choose first [in setting up the search]. Do you only want a RunX?
36	GG	Yes.
37	Lecturer	OK, so we'll stick to a RunX. We can leave it…the models, still open…you might be able to get a bigger RunX for this price that's still good.
38	GG	Yes.
39	Lecturer	Now let's imagine GG goes to the website. You are going to select price range first, right?
40	GG	Mmm (Agreement)
41	Lecturer	And you're going to get [a window which asks] prices between what and what you are willing to look at?
42	GG	Probably between 130 and 120 (gestures a range while she talks). OK so my highest is 125000, so 125000 is my max then my minimum would be …. Probably a 100 000. I wouldn't really go for anything less than that, mm…in case, yeah, that's what I'm willing to pay, so I'll pay up to that amount, probably not less than 100 000.
43	Lecturer	OK. So you look for cars in the 100 000 to, you said 130 000 bracket.
44	GG	OK.
45	Lecturer	Which is interesting for me. You want to pay a maximum of R125, but you look a little higher.
46	GG	Well it all depends on the other variables that fall under it.
47	Lecturer	Uh hu?

48	GG	So that's why, if I find a car, a RunX that is 125 000, and see that there is a newer model, with less mileage on it, that's a little bit more expensive, I might consider buying that one.
49	Lecturer	OK, save a bit and put it with that price 'cos it is now really a good buy for that little bit extra.
50	GG	Ja.
51	Lecturer	OK, good.
52	Lecturer	You have decided you are not going to look only for cars that cost exactly 125 000 rands, which makes sense. She is willing to go a bit lower to see maybe there is a bargain on this side, and a bit higher, maybe there is good value on that side. OK, can statistics tell her to do that?
53	Class	No
54	DH	No, but you need data to do that for yourself.
55	Lecturer	Ja, but that is a good reasoning decision that comes from common sense, from knowing the world a bit. Do you agree? Statistics can't deal with it [such a decision] can't help us to do that.
56	Lecturer	So before we deal with statistics, we must make good decisions about what we questions we want to ask, what ranges we want to put…to the data we are going to look at. That doesn't just jump out of the pages. So a good deal of thinking has to go into that and justify that, before you go looking at the data itself.
57	Lecturer	Right, so GG is now looking for cars. Can you imagine she's on the website, she has set the range for price at 100 to 130 thousand, and that is her first variable. Now she is going to get a *lot* of cars for that price range. Different models, different kilometre readings, different prices, different colours, sold in different places… What's the next important thing for you, GG?
58	MM	Extras
59	Lecturer	Lots of different extras…
60	GG	Probably the kilometres, like how much mileage it has done. After that…
61	Lecturer	Why? What is your reason for now choosing kilometres, uhm…
62	SDS	…above colour?
63	GG	Well, colour is actually my last one. Like my order would be…first the price, then the kilometres, then the model, … and then the colour. So, I think kilometres is to see how much mileage the car has done. To see if it's…If it has been driven too much then…maybe I don't want a car that already has 30 000 km on it, maybe I want a car with only 16 000 km on it.
64	Lecturer	OK let's stop there. She said kilometres is important. Do you all agree? Why? Why do you say kilometres is important?
65	Class	Scattering of yes
66	SM	I think it is not important
67	Lecturer	Not Important? Why?
68	SM	No, it's not
69	Lecturer	Not for you? OK, let's get the "important" argument, no let's get the "not-important" argument. Why do you say it is not important?
70	SM	It depends on how often the car has been taken to service and then, after…how many kilometres. So kilometres for me is not important.
71	Class	Silence.

72	NM	For me, it's just a model. I feel the model is more important than…
73	Lecturer	Do you mean by model the year of make?
74	NM	Ja
75	GG	Oh, I was taking it as the model of the car. Ja.
76	Lecturer	The engine capacity? Let's call that now the engine capacity.
77	Lecturer	Why, NM? Why do you take the year of make as more important than the number of kilometres on the clock?
78	NM	Eish, just a hunch.
79	GG	Maybe … it seems nicer to have a 2007 model than a 2004 model.
80	NM	Just that.
81	Lecturer	So your criteria for a car, I am putting words in your mouth, you actually want a brand new one.
82	NM	(Nods)
83	Lecturer	So the newer you can get it the better.
84	GG	But you can get a nice car for R130 000 if you're willing to pay for it. I am just saying. So you shouldn't actually look for second hand if you're willing to pay R130 000 for a new car.
85	Lecturer	That much…
86	GG	Ja.
87	Lecturer	Do you hear what she is saying? But let's get to why you say kilometres are not important. And then again it is about what do you assume happens…with cars as they drive, and what do you assume happens when it goes for its service, cos you said, kilometres is not important if it goes regularly for a service, you're happy.
88	SM	Yes.
89	Lecturer	Why?
90	Class	Silence.
91	Lecturer	MM?
92	MM	Do you want to say something?
93	Class	Laughter about MM turning the tables on the lecturer.
94	RC	No, well I think, ah, even if you put a car in for a service, they're not gonna put in a brand new engine, so if you have done 100 000 kilometres on this engine, they're not going to replace it, it's still going to have the same wear and tear than a car that has done 100 000 kilometres. Uhm, I mean, they will replace the brake pads in a service, but they are not going to replace important bits that you get wear and tear…(inaudible).
95	SM	Uhm, that depends on the lifespan of the engine. Most of the cars, maybe, you can drive them for more than 500 [-thousand] kilometres, only to find that the engine is still running very strongly.
96	RC	Yes, but…if…I mean,
97	SM	So we should not judge…
98	RC	…if most, if your car has a lifespan of 100 000 kilometres you'd rather buy one of 10 000 kilometres than 90 000 kilometres (inaudible) More life…for you.
99	RK	Just a comment. I, I think uh, from experience, I have seen some cars even that aren't strong, they don't even last as long as some second-hand cars, you know…it depends on the model, the company, I don't know…So, going for a new car doesn't necessarily mean you're doing good buying a good car. It depends on the model of the car.

100	Lecturer	I want you to listen to something. How as we are trying to understand this (the situation) our argument shifts. Right? It's true what you are saying, in your view, but what is a good car then for that price, if we look at different models? So there is so much information we can work with, our brains can't deal with it all at once. So we have to categorize, we have to make groups. We have to say first we look at RunX, and then we are going to look at another car. But we have to do it in pieces, because our brains can't do it. Good.
101	Lecturer	So, let's say…OK, I think, I think you had a good argument there, that if you think the maximum life is measured in kilometres, then rather buy a baby in terms of kilometres.
102	GG	…so that it can grow up with you…
103	Lecturer	So *kilometres* is now the next one. If you look at value for money, then we must also think, what does kilometres have to do with value for money? It all has to do with our understanding of how a car gets older and what it does when a car gets older. In the end it is the moving parts that get wear and tear.
104	Lecturer	Uhm, you have data about the depreciation of cars over years, is kilometres a factor in there?
105	GK	No.
106	Lecturer	Look at that data.
107	SM	It's not a factor.
108	Lecturer	It gives you prices. The new price, and the price a year later and a year later. Do they mention kilometres?
109	Lecturer	So if kilometres is an important one, what have they done with kilometres to compile that table, what do you think?
110	NM	I just want to say… the way I think about the car…it is as if I'm looking at a washing machine, somehow in my head (roll hands like wheels rolling)… so I'm thinking, if the washing machine can still do the job (roll hands energetically) still wash the clothes, even if it's old, but it can still do the job…I am somehow struggling to understand why…it doesn't matter how many washes the washing machine has done, but in future it can still do the job, washing that washing. So I'm also looking at the car in that way. So what matters is it can still move, it can still take me from one place to the next place, as much as the washing machine can still wash those clothes.
111	SDS	I think the important thing is for how long? If you get a baby car, it will last you much longer than time than an old car, where its gonna reach that end at some stage.
112	NM	Mm
113	DH	[Example of someone else's washing machine 16 years old that broke with the washing in it]
114	Lecturer	We can accept machinery needs service, and its parts does get wear and tear do get wear and tear. There comes a point where it is just too expensive to still keep this thing on the road.

Appendix D: Full transcript of Session 4: The meaning of the mean

Full transcript: The meaning of the mean		
Session 4: Discussion of Group A: Discursants KH,RK, GK, SM		
Turn	**Discursant**	**Utterance**
1	KH	So the first question is why do we use the mean as a measure of centre, is that right?
2	RK	Mmm.
3	Lecturer	We all know how to calculate the mean, I am reacting to what I hear... We all know how to calculate the mean. I want the common sense meaning of it, why the heck do we do it?
4	KH	(Referring to textbook) It says here, the mean is the balance point of the distribution, it balances...the mean's the balance point.
5	GK	Page what?
6	KH	We have to get the common sense here.
7	SM	Orraait we can say that it 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.
11	SM	Ja... in...
12	GK	Aren't we supposed to start with the standard deviation first?
13	KH	These three questions here....[ ]
14	RK	... the common sense behind that...
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).
16	RK	Mm
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.
19	KH	So then you know that half the data is above and half below, exactly half and half (looks at SM).
20	RK	I want to say, if you say the average height of kids...let's say one meter two [1.2 meter]...you say that generally...you find kids *around* that (moves hand horizontally at the same height at which he indicated 1 meter 2).
21	KH	Yes.
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.
25	GK	Sorry I missed that one, what did you say?
26	KH	(to RK) Why doesn't the median do the same?
27	SM	Inaudible.

28	RK	I think the median is like... if you have it ordered.
29	GK	Data , mm.
30	RK	You take the middle value.
31	GK	The middle value is the median yes.
32	SM	That is the mean.
33	KH	So if you have very tall kids in the class you're not going to get the mean there...you get the median.
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...
35	GK	Yes, you're given all the names.
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	Mmm.
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.
45	KH	Yes, yes an impression.
46	RK	An impression. I just get the right words.
47	GK	To express it ,mmm.
48	RK:	No but I want to get the general idea...
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.
53	SM	...if you talk...
54	KH	Reads from textbook: They say the right choice depends on the shape of your distribution. If you have a normal shaped distribution, the mean and the standard deviation are the most suitable summaries. If you have a skewed your distribution, there's the median and the quartiles that's the summaries.
55	RK, others	Mm (agreement).
56	Lecturer	I'm after, what's the common sense behind it. What's the common sense of mean.
57	KH	It gives you a kind of general impression of whatever the situation is.
58	Lecturer	Does it?
59		(KH looks at RK, they mumble it does).

60	Lecturer	Suddenly you add everything together, all the prices of the cars, and you *divide* by the number of cars. What does that *mean*, if you do that? What does it imply?
61	RK	Generally the cars cost...(hand movements indicate a horizontal group/cluster)...an amount.
62	Lecturer	Think of the calculation. What does that calculation mean? Go back to grade one. When you teach children to divide, what are they doing?
63	KH	Breaking up into equal parts...
64	RK	Or sharing...
5	Lecturer	Yes. *Sharing.*
66	RK	Dividing (gestures vertical cuts).
67	KH	Equal portions...so if you divide by eight, you divide whatever it is into 8 equal portions.
68	Lecturer	So now here you are saying, take the money, share it equally between all the cars…
69	KH	If you do that ,but you don't do that.
70	GK	Which means the average price you can get for a car, it falls among, arranges around a particular number, isn't it?
71	RK	Unless…it is not an exact value.
72	GK	Mmm, just around a particular value.
73	Lecturer	And that *value* is the one that you get when you say, oh, let's pretend they all cost the same.
74	KH	Yes.
75	Lecturer	What is the common sense behind that? What does it help you to pretend they are all the same? They are not the same!
76	KH	Smiles (you've got me-smile).
		Lecturer leaves the table.
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).
78	KH	It gives the impression of what...I still think that's accurate [our description].
79	GK	Mmm.
80	KH	Ja, it might not be exact, but...
81	GK	It gives an impression of?
82	KH	...whatever the situation is, heights of kids or prices of Toyota RunX's...
83	KH	(Refers to textbook again) Let's say we use the mean for the normal distribution, for that is the important point, for normal shaped distributions you use the mean and the standard deviation.
84	RK	Yes.
85	KH	It might not be... a common sense sort of explanation, but it is important to remember.
86	KH	I wonder what is this divide story (hearing lecturer talking at the

		other table).
87	KH	Let's go and see what they say in the summary (refer to textbook). Summary of the chapter. The mean is the balance...the mean is the summary the sum of the values divided by how many there are...(shrugs).
88	GK	Inaudible....the mean, giving us an indication of the situation, you know what I'm saying?
89	KH	Because, the fact that you have the price of cars, (pointing to RK), if you want to know what is the average price of cars, if you want to know if it is in your price range...
90	GK	Where it lies.
91	KH	That's what the mean gives you.
92	GK	Mmm.
93	KH	It gives you an idea of the value of the cars. It's always just an idea, it's not a... (gestures).
94	RK	(Refers to textbook, page 23)...Raw data, a long list of values...it gives you value to make sense of raw data.
95	KH	Like a summary.
96	GK	(Reads) Raw data a long list of values it's hard to make sense.
97	RK	Yeah.
98	GK	Mm.
99	RK	Mm.
100	GK	Which means ...
101	RK	If you have raw data, if you take the mean, it will begin to give you a sense of...whatever you are looking for.
102	GK	Mm, because it doesn't make sense…
103	RK	When your data is not…
104	GK	…organised or not.
105	RK	It gives you a sense.
106	GK	So in other words are you saying the mean can also help in the raw data...to to to give some sense of what the data is all about??
107	RK	Yes, yes.
108	GK	(to SM) Aré, the mean itself né, because remember raw data is a long list of values which is hard to make sense of.
109	SM	Ja.
110	GK	Mm… and by using the mean it can assist us in understanding I mean giving us an understanding of the raw data and so, because it ahm... (tick tick with pencil on table)... it gives us the average (swirling gesture with hand) like now it gives us the average amount that you can get for the cars in particular, aré, for the RunX.
111	SM	Mmm.
112	GK:	I don't know if it makes sense, but that's what I'm getting. (Sits back in chair).
113	Lecturer	(Returns to Table A) Pretend you didn't know the word average.
114	Lecturer	Does it *help* you to see what is going on in your data set if you think about...you're sharing the price, the total price there? Does it help you to understand?
115	RK	(Nods in agreement).
116	GK	It tells us, if we talking about sharing the money, it tells us, how

		much... how many of us are going to get that ah amount. You know what I'm saying. It's how many (swirling movement with hand, palm up) of us are going to get that amount. Even if there are others...are going to get a little bit more, but most of us will get... that amount.
117	RK	Generally.
118	KH	Does it always, does it tell us that?
119	GK	It does!
120	GK	Look at the median.
121	RK	I think what GK is trying to say, it gives us a sense of what each person is going to get.
122	GK	Mm, that's what...
123	KH	Yes, yes as a group, in the group.
124	GK	Mm...but it doesn't mean that all of us are going to get the same.
125	RK	Not exactly yes, but more or less.
126	GK	...just giving us a picture.
127	KH	If it was an equal distribution within the group (gestures).
128	GK	That's what we get.
129	KH	OK.
130	Lecturer	(Returning to Table A): I noticed when you said that you said *most*of us are going to get that. I want to challenge the word *most*. Does the mean tell you what *most* of you *are* going to get?
131	KH	No.
132	RK	I think, I think *all* of us.
133	KH	Yes, if it was distributed fairly... (gestures) then it would be all of us.
134	GK	Even if it is all of us, you might find that one is out, or even two is out.
135	KH	What if you take the mean though, if you take the mean...and you work that out....and you multiplied it, you get back to your value. It's like with the cars here, if you added up all the prices and you divide it by the cars you get your mean, if you multiply that by the cars you get back to the total price.
136	GK	Mm.
137	KH	We get the total price.
138	GK	That is what I'm saying. It tells us of how much is each going to get in the group. But at the end of the day it doesn't tell us *exactly* that we all gonna get the same. There might be one or two that is out.
139	Lecturer	That's now when you go back to the original distribution.
140	GK	Yea.
141	Lecturer	And you compare to the mean, you say, no we didn't all get the mean.
142	GK	Mmm.
143	Lecturer	So it's a ...[the mean] is something that we do in our heads to get a, seems to me a number that says, what *would* we all get *if* we all got the same. Right?
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 (gesture shaking hand) an approximation.
145	KH	Yes.

146	Lecturer	An approximation of *what.*
147	RK	Ah, like ah, I'll be taking an example. There's an amount of money to share between people, then we say, OK we take the amount and we divide it by the number of persons, then we say that approximately...Oh no, let's go over to the prices of cars. The prices vary, so if you took the mean of the prices you'll say that generally in this shop, if you want a RunX, it's going to cost *about* this amount. But it won't be exactly that amount. If, if somebody who wants to buy a car gives that information, that would give him a sense of how he should prepare.
148	GK	Ja.
149	RK	But then it's not going to be exactly that amount, it's going to be plus or minus.
150	GK	That amount.
151	RK	I don't know if it makes sense.
152	Lecturer	Ask your friend.
153	Lecturer	(to SM) Does it make sense to you?
154	SM	I think it's making a lot of sense...Because my understanding about the mean...when you when we talk of the prices of the cars...so for us ...the average says to us, that is when you add the total price of the cars and divide by the number of cars then the amount that you're going to get that is the average *amount* that you're going to pay. It's not the exact amount it's the average amount that you are going to pay.
155	RK	More or less.
156	SM	Yes, if you talk of the car that is priced at 80000 and then at 90000 the two cars. Then the average amount is 85000.
157	KH	So it's like taking the total amount and share it out equally amongst each individual (hand moves horizontally across at the same height).
158	SM	Yes.
159	GK	So it is an indication of how much we can expect to pay.
160	SM	Yes, yes, it just gives us the average not the *exact* amount...what you're going to pay.
161	RK	So if we have raw data, just prices of cars, it will be confusing to consider each individual price like that (gesture going down a list) then you have to start deciding how am I going to buy one...but then…the mean tells you exactly how much approximately you are going to pay.
162	Lecturer	Carry on with the rest of the questions.
Session 4: Discussion of Group B: Discursants: SDS, NM, GG, MM		
163	MM	So it's the mean and standard deviation, isn't it?
164	Lecturer	You all know how to calculate the mean...why the heck do we do it?
165	GG	Well, it's the average.
166	NM	It's the balance point of the distribution (from textbook).
167	GG	We're supposed to be discussing it.
168	SDS	(To MM) Yes, she said...
169	SDS	The logic behind the mean ... as a measure of centre. (Pages through textbook).

170	NM	I'm thinking of this in terms of maybe working on the mean of learners' marks. Let's say they wrote a test, then we work out their mean mark. What exactly are you trying to say? What information are you trying to deduct from that? Working on the mean, what does that tell you about the performance of those learners?
171	GG	So what are you actually looking for.
172	NM	What are you looking for.
173	MM	You are basically saying, ja, you are basically saying I mean all of those kids, I mean they got sort of like that centre in your mark (hands gesture a small cluster).
174	NM	It's not a centre mark. The mean the way usually I look at it, is
175	SDS	The mean gets affected by all values.
176	NM	Uhhuh.
177	SDS	The median, it doesn't really matter what the amount of the value is, it is the number ... of them. So if you have an outlier (gestures a point far to her right with one hand) that will affect your mean, that over here is quite large so your mean will be quite high (left hand indicates a shift towards the right). If it's still the median it will still be at the same position (left hand moves back to original position) even if your outlier was closer in (right hand indicates a point moving left).
178	NM	But now if I go back to the story about learners' marks for example, what exactly (inaudible.) Is there a specific value you are looking for or what? Cos the way I see it is, if the mean is *low* for that group then I get the sense my group is not performing well.
179	Group	Yes.
180	NM	And if the mean is just maybe...50% (shakes hand horizontally at the same level) still I wouldn't be happy about it (weighing movements with hands). So, that *mean* for those marks, there *is* a certain number that we are looking at as a result. That I need my learners to get this mean and if they are not getting this mean, it means they are performing below it [this mean] on average.
181	NM	If I use it [this mean] on the marks then I would know their mean is low, meaning *most* (gestures swirling movements with both hands as if including all) of them didn't perform well, that's why their mean, maybe, would be *low*. If they performed well, their mean would be higher, that would mean generally (hands swirling to the outside indicating ambivalence, uncertainty) they are performing... better.
182	SDS	You see I don't know if you can say that ... because if…
183	NM	Is, if it's at school.
184	SDS	No, I don't mean that, I understand the point you're making, but I'm saying we can't say *generally* (gestures with both hands scooping together) they are performing...because there could be ten of them that have quite high marks, and then you have...twenty of them that have really really low marks, and it could give you...
185	NM	Even in a class of thirty or fifty, it is an indication that they are not performing, if that's the case.
186	SDS	(Nods) Ten out of thirty?
187	NM	Ten out of thirty, and yet when the mean is 50 you will say they are

		performing well, and 50 again, is it not that.
188	SDS	So should we say that the mean is more...indicative of....
189	NM	The group's… performance.
190	SDS	Performance.
191	SDS	…more indicative of the values of your cases, whereas the median isn't.
192	NM	(Shakes head in disagreement). The median is just telling you about the...
193	SDS	Where half the dots is... here or there (Gestures with hand along a horizontal line).
194	NM	(Nods head in agreement).
195	SDS	So mean is indicative of the
196	NM	performance
197	SDS	Of values of the cases. The actual *values.*
198	NM	Can we not go back to to to stats alone, just try and keep the conver... the minute we use the terms again somehow I'm lost ... somehow I'm mystified.
199	SDS	Well we can't say performance, because that's only for marks. The average for anything.
200	NM	OK I see where you're going.
201	GG	It's got something to do with, like you said about, the mean is like half is below and half above.
202	SDS	(Shakes head in disagreement).
203	GG	Something about balancing.
204	SDS	Median.
205	GG	OK, ja, 'cos the mean is the average.
206	Group	Mmm.
207	NM	After calculating the mean, do you then set your own standard, like you ...calculate the mean for this particular group, then there is the ideal value that we would like to get, compare with your ideal values? (Moving hands up and down)...Are they performing according to this ideal, or are they below this ideal, or... maybe we are running away from the question.
208	MM	I get you very well.
209	NM	What is it that we've been saying.
210	SDS	...indicative of the value of the cases.
211	MM	What do you mean?
212	SDS	It indicates...I don't wanna say average. It indicates the average, you see, the thing for me is that it's looking at the values of your cases, your cases can be marks or.... (to NM) each person is a case, and then marks is the value.
213	NM	Mm.
214	SDS	It's indicative of that *value*, that variable (gestures). And if it's just median, it doesn't really tell you....about the values.
215	Group	(Silence).
216	SDS	Because you could have... (draws on back of note pack) could have OK, one like this. (She draws a dot plot with outlier on the right) Where your mean will be a little higher, because of the outlier, whereas your median will still be ...7....so it will be here somewhere,

		K? But if I have to draw another picture and I still have these, and I brought this outlier inwards (draws a point closer to the rest of the points), the mean will go closer towards the median...but this one here is *very* dependent on the values. That's the value of having the mean, 'cause you're looking at particular values, not only looking at position... of the data plot.
217	Lecturer	So what's the common sense behind it.
218	Group	(Silence).
219	NM	We're struggling to point in general terms. We just have one specific example, now we're trying to put it in general.
220	SDS	That's [per our example] an indication of performance in terms of marks.
221	NM	In terms of marks. We're looking at the marks for kids...
222	Lecturer	So what do you *do* when you calculate the mean?
223	NM	When you calculate the mean...aren't you going to add up all those marks and divide by the number of learners?
224	Lecturer	And what does it mean to do that?
225	NM	To do that, it tells me that if the mean is *low.*
226	Lecturer	No! I don't say interpret the mean, I say interpret your action of adding everything together and dividing it by the number of...
227	GG	Find the average?
228	Lecturer	Mm? Say you didn't know that was "the average"? I'm trying to challenge you to get beyond the things you have "learned". To say, What do I *do* when I do it (calculate the mean). So pretend you are in grade one, so the teacher teaches you to divide, what is she teaching you to *do*? What are you doing when you are dividing?
229	Group	(Silence).
230	SDS	Levelling out.
231	Lecturer	Example?
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 (gesture 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.
233	Group	(Silence).
234	Lecturer	Comment?
235	Lecturer	(To others in group) What do you get from your understanding of what you did in learned in grade one of division?
236	NM	You were sharing equally.
237	MM	You were trying to share things equally.
238	Lecturer	Ja! So is this what you are doing here? We say let's put *all* the money in a pot and we share it equally. We *pretend* that each car cost the same.
239	Group	(Silence).
240	Lecturer	Is that it? Can you give another meaning to that calculation?
241	Group	(Silence).
242	Lecturer	Does it help you if you think about data to think about sharing?
243	Group	(Silence).
244	Lecturer	I'm asking you about common sense. There's no argument like this

		in the textbook. Nobody talks about things like this except teachers – we have to make sense OK so shoot! (leaves table).
245	MM	OK so what is it we are trying to share...are we trying to share this one mark to all of those kids or what?
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.
250	NM	Uh-hu the mean.
251	MM	Uh-hu.
252	NM	So why is it important?
253	SDS	So why *is* it important?
254	GG	The fact that it is sharing amongst...groups.
255	NM	... to see how much they will get...compared to the whole. What is the whole, is the whole not getting...full marks? How far are they from the whole if they were to share...
256	MM	You are talking standard deviation.
257	NM	Ha? No I mean I'm still going back to this interpretation of mine of if they are below average or above average or what...somehow I keep going back to that.
258	MM	If you are calculating the mean, you are trying to say, I mean, with all these kids, uhm, ...eish
259	NM	How much...
260	SDS	We can talk of this in terms of balance.
261	NM	Trying to balance...to balance.
262	GG	Balancing what?
263	GG	(Gestures, checking if lecturer is looking) It's just like...like...there's something in the air!
264	SDS	Balancing the values of the cases, so each case has the same value
265	NM	But why?
266	GG	Ja, why?
267	SDS	Why? So that's.... (gestures, draws funny face).
268	GG	Ja, you see, it's (airy hand movements).
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).
271	NM	Yes, I keep going back to it it's a judgement, judging if they
272	SDS	Judging on assessment criteria.
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, 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.
274	SDS	The...cluster.
275	GG	Cluster?
276	SDS	It's far away from the cluster of values.
277	GG	Ja
278	MM	You see from what I know, let me just try to explain this...Suppose this is a classroom with learners, and that's their average (pointing to his book) for this kids here. I mean the assumption is if I was to take this learner who is basically out of that class and put it in there, so this learner is likely to get this mark...*in* that class.
279	SDS	(Pulls a doubting face)
280	GG	And?
281	SDS	You say *if* a learner comes to that class he is likely to get 34 because that is the class average.
282	MM	Ja, that's the average mark.
283	NM	But there are other factors.
284	SDS	(Laughs) Ja! But yes, in a way, you saying that's a way of using the mean.
285	MM	Ja...
286	GG	(Interrupts) Well that doesn't say very much for the teacher
287	NM	And besides it could be...it could be that this child has better information than these and therefore this child *might* perform better. There are other factors.
288	MM	I dunno what to do.
289	Group	(All laugh).
290	GG	(To SDS) What did you say about the judging thing?
291	MM	Ja, I mean it's all about judging.
292	SDS	(Reads from her book) Balancing values so that each case gets the same value so that you can judge and assess a set of cases.
293	MM	Balancing what?
Session 4: Whole class discussion		
294	Lecturer	This kind of question (what is the common sense behind statistical tools and procedures) you must take up the challenge with. We are teachers. If we were engineers, we could just... press the button on a computer. You need to know more as teachers. You need to think harder than engineers...OK?
295	KH:	Cos we're better than engineers.
296	Lecturer:	We are! (Some laughter).
297	KH:	We just need to be paid.
298	Lecturer:	OK the first challenge was the *mean.* Let us explore what you said. I don't want you to change what you have written now, but you can take notes....
299	Lecturer:	If I were you I'd say I am going to make a concept map...
300	Lecturer:	So let's talk about the mean. I heard in all groups you said, the mean, the calculation of the mean, actually says we're sharing equally. Right? You're doing the same thing you did in Grade 1, you are sharing equally. Now why on earth, if we *know* the prices are not the same, do we want to *pretend* that they are the same?

301	Class	(Silence).
302	Lecturer	If you were a clever child... it should bother you knowing oh no they're not the same, why do we pretend they are the same? That's why I am asking you what's the logic behind it.
303	Class	(Silence).
304	Lecturer	So the thinking behind it is: Well, *if* they were all the same, *then* it would be this price. I heard KH saying you're sort of calculating a norm, a standard for these prices. Now the word *norm* the word *standard* I'm not sure how to interpret it, but I'm taking from you it's an ideal value. If everything was shared perfectly fairly it could've been the ideal price, OK. Luckily it isn't, because then we wouldn't have car markets...
305	RC	Surely it's a value you can use to quickly compare everything else to it?
306	Lecturer	And doesn't it satisfy that for you, now can, now you say OK the mean for you is something you want to compare other things to. So does this way of thinking about the mean do it for you?
307	SDS	I don't like the way you say it's an ideal value. Cos you could say for a test the ideal value would be a hundred percent but...the average is 35. So I think the comparative is much better, 'cos you can compare that average to other averages.
308	Lecturer	To other *averages* now?
309	RC	To other scores.
310	KH	Inaudible.
311	Lecturer	OK, so it's two things and I am trying to keep the conversation on one, and I'll park that one. So an average is useful to compare different data sets, right. So let's just stick again with the meaning of average that you get from primary school and the children in your class will know: add the scores and divide the total by the number of values. It means fair sharing. If you now draw the graph of the mean, the bar that's the mean, you know, everybody will get the same.
312	Class	Mmm (agreement).
313	Lecturer	It's hard to think about.
314	Lecturer	So, let's put the word ideal just in brackets, but it gives you something and now we can compare the other data points to it.
315	DH	I also just think of the mean as, you've got a whole lot of data, so you've got one guy that comes to represent the set, becomes the representative of the set.
316	KH	Gives you an impression.
317	Lecturer	Why do you need to calculate to get that?
318	DH	That's correct, but it becomes the one value that goes with the, think about our class averages that we calculate. We judge a class by that average, it seems that one number represents how well that class did in the test.
319	Lecturer	Mmm, and I want to challenge that.
320	DH	Yes I want too.
321	Lecturer	So let's think. You want to bring in that it is representative.
322	DH	It becomes representative.
323	Lecturer	Let's think about representative. I am having pictures in my mind. I want you to do it as well. (Check other tape for lecturer's gestures). I

		want you to do it as well. Graphs. I am seeing pictures, dot plots. Representative. Then GK said, *most* have that value. *Most* will be that price. I know she has corrected herself, it isn't most if we go and count. But if we say representative, gee you want to see values in the dataset that has that value that the mean has. Right? But *is* the mean always a value, the same as a value in your dataset?
324	Class	No, mm-mm.
325	Lecturer	You can get a *mean* that *no* one of your numbers are (sic).
326	Class	Agrees.
327	Lecturer	So how's that representative?
328	DH	It's a kind of a centre or middle. It's some kind of middle. It gives you a sense of something from the middle of the dataset.
329	KH	It's an impression. It gives you an impression of the price of a RunX as compared to an impression of the price of a Mercedes.
330	Lecturer	OK. So there are three different ways to conceive it, and I want you to write them down and think about them. The one is this fair share option, the other one is signal in noise, and we haven't talked about it. It's what you think you see if you don't see well. A signal in the noise. There are so many prices. Can't talk about it.
331	Class	Yes.
332	Lecturer	But if I mention this *one*, it signals to me OK that's it.
333	Class	(Groups talk with lots of energy).
334	KH	(To RK) It's like you said, all this information is just confusing, so the noise is *data* (gestures).
335	SM	Confusing set of data.
336	GK	That's what they warn us in school about raw data.
337	Class	Yes, mm.
338	KH	So the noise is data and the signal is the mean.
339	GK	To lecturer: And the third one?
340	KH	What is your project about. You analyze textbooks right?
341	GK	(Shakes her head, drops head in palm).
342	Lecturer	Right! And the *third* one is the balance point idea which is in your textbook as well. If you look at the graph with points that is on a measurement scale, you can say where can I balance that.
343	RC	That middle thing.
344	Lecturer	(Nods agreement) But middle again says more...can also be equal on both sides, so it's more the balance idea, the balance that you get. It's three different ways of thinking logically about that measurement that came to be called the mean.
345	Class	(Inaudible).
346	Lecturer	Good.

Reference List

Ahn, W. K., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition, 54*(3), 299 - 352.

Allmond, S., & Makar, K. (2010). Developing primary students' ability to pose questions in statistical investigations. Paper presented at the Eighth International Conference on Teaching Statistics. Retrieved from www. stat. auckland.ac.nz/~iase/publications

Anggoro, F., Gentner, D., & Klibanoff, R. (2005). How to go from nest to home: Children's learning of relational categories. Paper presented at the Twenty-seventh annual meeting of the Cognitive Science Society.

Arnold, P. (2008). What about the P in the PPDAC cycle? An initial look at posing questions for statistical investigation. Paper presented at the Eleventh International Congress of Mathematics Education. Retrieved from http://tsg.icme11.org/tsg/show/15

Arnold, P. (2009). Context and its role in posing investigative questions. Paper presented at the Sixth International Forum for Research on Statistical Reasoning, Thinking and Literacy.

Bakker, A. (2003). The early history of average values and implications for education. *Journal of Statistics Education, 11*(1), 1 - 17.

Bakker, A. (2004a). *Design research in statistics education: On symbolizing and computer tools*. Unpublished doctoral thesis, Utrecht

Bakker, A. (2004b). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal, 3*(2), 64 - 83. Retrieved from www.stat.auckland.ac.nz/serj/

Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning, 13*(1), 5 - 26.

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.

Ball, D. (Ed.). (2000). *Working on the inside: Using one's own practice as a site for studying teaching and learning*. Mahwah, New Jersey: Erlbaum.

Batanero, C. (2004). Statistics education as a field for research and practice. Paper presented at the ICME 10, Copenhagen, Denmark.

Ben-Zvi, D. (2004). Reasoning about data analysis. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 121 - 146). Dordrecht: Kluwer Academic Press.

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.

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.

Biehler, R. (1999). Discussion: Learning to think statistically and to cope with variation. *International Statistical Review, 67*(3), 259 - 262.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning. The SOLO taxonomy*. New York: Academic.

Brickell, G., Ferry, B., & Harper, B. (2002). Developing informal reasoning skills in ill-structured environments. A case study into problem-solving strategies. Paper presented at the Winds of change in the sea of learning: 19th Annual conference of the Australian society for computers in learning in tertiaty education (ASCILITE). Retrieved from http://www.ascilite.org.au/conferences/aucklandoz/proceedings/papers/094.pdf

Brodie, K. (1999). Working with pupils' meanings: Changing practices among teachers enrolled on an in-service course in South Africa. Paper presented at the Twenty third Psychology of Mathematics Education Conference.

Brodie, K., Lelliott, T., & Davis, H. (2002). Developing learner-centred practices through the FDE programme. In Y. Reed (Ed.), *Challenges of teacher development: An investigation of take-up in South Africa*. Pretoria: Van Schaik.

Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education: A joint ICMI/IASE study* (pp. 57 - 69): Dordrecht: Springer.

Chance, B., DelMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295 - 324). Dordrecht: Kluwer Academic Press.

Chin, C., & Kayalvizhi, G. (2002). Posing problems for open Investigations: What questions do pupils ask? *Research in Science and Technological Education, 20*(2), 269 - 287.

Clements, D. H., & Battista, M. T. (1992). Geometry and spatial reasoning. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 420 - 464). New York, NY: MacMillan.

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly, 104*(9), 801 - 823.

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

Cobb, P., & McClain, K. (2004). Principles of instructional design for supporting the development of students' statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 375 - 396). Dordrecht: Kluwer Academic Press.

Confrey, J., & Lachance, A. (2000). Transformative teaching experiments through conjecture-driven research design. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education*. Mahwah, New Jersey: Erlbaum.

Cummins, D. (1995). Naive theories and causal deduction. *Memory and Cognition, 23*(5), 646 - 658.

Da Silva, C. B., & Coutinho, C. (2008). Reasoning about variation of a univariate distribution: A study with secondary mathematics teachers. Paper presented at the ICME/IASE 2008. Retrieved from http://www.ugr.es/~icmi/iase_study/Files/Topic2/T2P7_Silva.pdf

Davis, B. (1997). Listening for differences: An evolving conception of mathematics teaching. *Journal for Research in Mathematics Education, 28*(3), 355 - 376.

DelMas, R., Garfield, J., & Zieffler, A. (2006). *The tyranny of context*. Paper presented at the Sixth International Collaboration for Research on Statistical Reasoning, Thinking, and Literacy.

DelMas, R., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal, 4*(1), 55 - 82.

Department of Basic Education. (2010). *Curriculum and assessment policy statement (CAPS). Mathematics grades 10 - 12. Final draft*. Retrieved from http://www.education.gov.za/Curriculum/CurriculumAssessmentPolicyStatements.

Department of Education. (2003). *National Curriculum Statement Grades 10-12. (General). Mathematics*. Retrieved from http://wced.pgwc.gov.za.

DiSessa, A. (1993). Towards an epistemology of physics. *Cognition and Instruction, 10*, 105 - 226.

Donaldson, M. L. (1986). *Children's explanations: A psycholinguistic study*. Cambridge, UK: CUP.

Erduran, S., & Pilar Jiménez-Aleixandre, M. (Eds.). (2008). *Argumentation in science education*: Springer.

Fielding-Wells, J. (2010). Linking problems,conclusions and evidence: Primary students' early experiences of planning statistical investigations Paper presented at the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia.

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1993). The effects of statistical training on thinking about everyday problems. In R. E. Nisbett (Ed.), *Rules for reasoning*. Hillsdale, New Jersey: Erlbaum.

Franklin, C., & Garfield, J. (2006). The GAISE Project: Developing statistics education guidelines for grades pre-k - 12 and college courses. In G. Burrill & P. Elliott (Eds.), *Thinking and Reasoning With Data and Chance: 68th NCTM Yearbook* (Vol. 68, pp. 345 - 376). Reston, VA: National Council of Teachers of Mathematics.

Gadamer, H. (1993). *Reason in the age of science* (F. G. Lawrence, Trans.). Cambridge, MA: The MIT Press.

Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education, 10*(3). Retrieved from http://www.amstat.org/publications/jse/v10n3/garfield.html

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Dordrecht: Springer.

Gelman, S. A. (1988). The development of induction with natural kind and artefact categories. *Cognitive Psychology, 20*, 65 - 95.

Gentner, D., & Kurtz, K. (2005). Learning and using relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman & P. W. Wolff (Eds.), *Categorization inside and outside the laboratory* (pp. 151 - 175). Washington, DC: APA.

Gierdien, F. (2008). Teacher learning about probabilisitc reasoning in relation to teaching it in an advanced certificate in education (ACE) programme. *South African Journal of Education, 28*, 19 - 38.
Gigerenzer, G. (2004). Mindless statistics. *The Journal for Socio-Economics, 33*, 587 - 606.
Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. New York: OUP.
Gil, E., & Ben-Zvi, D. (2011). Explanations and context in the emergence of students' informal inferential reasoning. *Mathematical Thinking and Learning, 13*(1), 87 - 108.
Glassner, A., Weinstock, M., & Neuman, Y. (2005). Pupils' evaluation and generation of evidence and explanation in argumentation. *British Journal of Educational Psychology, 75*, 105 - 118.
Graham, A. (2006). *Developing thinking in statistics*. London: The Open University in association with Paul Chapman Publishing.
Gravemeijer, K. (2002). Emergent modelling as the basis for an instructional sequence on data analysis. Paper presented at the Sixth International Conference on Teaching Statistics, Cape Town, South Africa.
Groarke, L. (Fall 2012). Informal Logic. *The Stanford Encyclopedia of Philosophy*. Retrieved from http://plato.stanford.edu/archives/fall2012/entries/logic-informal/
Groth, R. E., & Bergner, J. A. (2006). Preservice Elementary Teachers' Conceptual and Procedural Knowledge of Mean, Median, and Mode. *Mathematical Thinking & Learning, 8*(1), 37-63.
Hacking, I. (1975). *The emergence of probability. A philosophical study of early ideas about probability, induction and statistical inference*. Cambridge: CUP.
Hacking, I. (1990). The taming of chance. Cambridge, MA: CUP.
Hall, C. C., Ariss, L., & Todorov, A. (2007). The illusion of knowledge: When more information reduces accuracy and increases confidence. *Organizational Behaviour and Human Decision Processes 103*, 277 - 290.
Hiebert, J., & Carpenter, T. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65 - 97). Reston, VA: NCTM.
Hiebert, J., & Grouws, D. A. (2007). The effect of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371 - 404). Charlotte, NC.: INformation Age Publishing.
Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1 - 28). Hillsdale, New Jersey: Erlbaum.
Jacobbe, T. (2008). Elementary school teachers' understanding of the mean and median. Joint ICME 18/IASE study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. In C. Batanero, G. Burrill, C. Reading & A. Rossman (Eds.), *Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*.

Jepson, C., Nisbett, R. E., & Krantz, D. H. (1993). Inductive reasoning: Competence or skill? In R. E. Nisbett (Ed.), *Rules for reasoning* (pp. 70 - 83). Hillsdale, New Jersey: Erlbaum.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases*. Cambridge, UK: CUP.

Keogh, B., & Naylor, S. (1999). Concept cartoons, teaching and learning in science: An evaluation. *International Journal of Science Education, 21*(4), 431 - 446.

Kim, D., Ferrini-Mundy, J., & Sfard, A. (2012). How does language impact the learning of mathematics? Comparison of English and Korean speaking university students' discourses on infinity. *International Journal of Educational Research, 51* – 52(0), 86 - 108. doi: 10.1016/j.ijer.2012.01.004

Kintsch, W. (1994). Text comprehension, memory and learning. *American Psychologist, 49*(4), 294 - 303. doi: 10.1037/0003-066X.49.4.294

Klauer, K. J., & Phye, G. D. (2008). Inductive reasoning: A training approach. *Review of Educational Research, 78*(1), 85 - 123.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction, 6*, 59 - 98.

Konold, C., & Higgins, T. L. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 193 - 215). Reston, VA: National Council of Teachers of Mathematics.

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.

Kuhn, D. (1991). *The skills of argument*. Cambridge: CUP.

Lakoff, F., & Núñez, R. E. (2000). *Where mathematics comes from. How the embodied mind brings mathematics into being*. New York: Basic Books.

Lampen, E. (2009). The zeroth problem and statistical reasoning. Paper presented at the The 33rd Conference of the International Group for the Psychology of Mathematics Education.

Lampen, E. (2010). Structuring contexts for statistical treatment: Initializing statistical reasoning. Paper presented at the Eighth International Conference on Teaching Statistics.

Lampen, E. (2011). Structuring everyday contexts for statistical treatment: The struggle for shared focus. African Journal for Research in MST Education, 15(1), 51 - 66.

Langrall, C., Mooney, E., & Williams, N. (2005). *Students' use of context knowledge in interpreting data.* Paper presented at the Building connections: Research, theory, and practice. 28th Annual Conference of the Mathematics Education Research Group of Australasia Sydney.

Langrall, C., Nisbet, S., Mooney, E., & Jansem, S. (2011). The role of context expertise when comparing data. *Mathematical Thinking and Learning, 13*(1), 47 - 67.

Lecoutre, M. P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics, 23*, 557 - 568.

Lehrer, R., & Kim, M. (2009). Structuring variability by negotiating its measure. *Mathematics Education Research Journal 21*(2), 116 - 133.

Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition & Instruction*, *14*(1), 69.
Lehrer, R., & Schauble, L. (2002). Distribution: A resource for understanding error and natural variation. Paper presented at the Sixth International Conference on Teaching Statistics. Retrieved from www. stat. auckland.ac.nz/~iase/publications
Lehrer, R., Kim, M., & Schauble, L. (2007). Supporting the development of conceptions of statistics by engaging students in measuring and modelling variability. *International Journal of Computers for Mathematical Learning, 12,* 195 - 216.
Lesser, L. M., & Blake, S. (2007). Mathematical power: exploring critical pedagogy in mathematics and statistics. *Journal for Critical Education Policy Studies, 5*(1).
Lester, F. K. (2005). On the theoretical, conceptual and philosophic foundations for research in mathematics education. *ZDM, 37*(6), 457 - 467.
Lipton, P. (2004). *Inference to the best explanation*. London, UK and New York, NY: Routledge.
Luria, A. R. (1976). *Cognitive development. Its cultural and social foundations* (M. Lopez-Morillas & L. Solotaroff, Trans.). Cambridge, Massachusetts: Harvard University Press.
Madden, S. (2011). Statistically, technologically, and contextually provocative tasks: Supporting teachers’ informal inferential reasoning. *Mathematical Thinking and Learning, 13*(1), 109 - 131.
Makar, K. (2004). *Developing statistical inquiry: Prospective secondary mathematics and science teachers’ investigations of equity and fairness through analysis of accountability data*. Unpublished doctoral dissertation, University of Texas at Austin, Texas.
Makar, K., & Confrey, J. (2005). "Variation-talk": Articulating meaning in statistics. *Statistics Education Research Journal, 4*(1), 27 - 54.
Makar, K., & McPhee, D. (2009). Young children’s exploration of average in an inquiry classroom. Paper presented at the 32nd Annual Conference of the Mathematics Education Research Group of Australasia.
Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning, 13*(1), 152 - 173.
Makina, A. (2009). *An investigation of the role of visualization in data handling in Grade 9 within a problem-centred context*. Unpublished Master’s thesis, University of South Africa.
Mallows, C. (1998). The zeroth problem. *The American Statistician, 52*(1), 1 - 9.
Masnick, A. M., Klahr, D., & Morris, B. J. (2007). Separating signal from noise: Children’s understanding of error and variability in experimental outcomes. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 3 - 26). New York, London: Erlbaum.
Mathews, D., & Clark, J. M. (2007). Successful Students’ Conceptions of Mean, Standard Deviation, and The Central Limit Theorem. Paper presented at the Research in Undergraduate Mathematics Education Community. Retrieved from http://www1hollins.edu/faculty/clarkjm/stats1.pdf
Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review, 62*(3), 279 - 300.

Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability and knowledge levels. *Cognition and Instruction, 14*(2), 139 - 178.
Mercier, H., & Sperber, D. (2008). Intuitive and reflective inferences. In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond*. London: OUP.
Mokros, J., & Russell, S. J. (1995). Children's concept of average and representativeness. *Journal for Research in Mathematics Education, 26*(1), 20 - 39.
Moore, D. S. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 950 - 137). Washington, DC: National Academy Press
Moore, D. S. (1998). Statistics among the liberal arts. *Journal of the American Statistical Association, 93*(444), 1253 - 1259.
Moore, D. S. (2004). *The basic practice of statistics* (3rd ed.). New York: W.H. Freeman.
Moritz, J. (2004). Reasoning about covariation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 227 - 256). Dordrecht: Kluwer Academic Press.
Nisbett, R. E. (1993). *Rules for reasoning*. Hillsdale, New Jersey: Erlbaum.
Nisbett, R. E., & Borgida, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology, 32*, 932 - 943.
Nisbett, R. E., Krantz, D. H., & Kunda, C. J. Z. (1993). The use of statistical heuristics in everyday inductive reasoning. In R. E. Nisbett (Ed.), *Rules for reasoning* (pp. 15 - 54). Hillsdale, New Jersey: Erlbaum.
North, D., & Scheiber, J. (2008). Introducing statistics at school level in South Africa: The crucial role played by the national statistics office in training in-service teachers. In C. Batanero, G. Burrill, C. Reading & A. Rossman (Eds.), *Joint ICME/IASE Study: Teaching statistics in school mathematics: Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE round table conference*. Monterrey, Mexico: International Statistics Institute.
Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97*, 185 - 200.
O'Toole, T. (2006). Building powerful understanding by connecting informal and formal knowledge. Paper presented at the MERGA 29. Retrieved from www.merga.net.au/documents/RP432006.pdf
Peräkylä, A., & Ruusuvuori, J. (2011). Analyzing talk and text. *The Sage handbook of qualitative research* (4 ed., pp. 529 - 544). Los Angeles, London: Sage.
Pfannkuch, M. (2008). Training teachers to develop statistical thinking. Paper presented at the Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference.
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. doi: 10.1080/10986065.2011.538302
Pfannkuch, M., & Horring, J. (2005). Developing statistical thinking in a secondary school: A collaborative curriculum development. Paper presented at the Curricular development in statistics education: International Association for Statistical Education 2004 Round Table, Voorburg, The Netherlands.

Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi & J. Garfield (Eds.), *Statistical literacy, reasoning, and thinking: Goals, definitions and challenges* (pp. 17 - 46). Dordrecht: Kluwer Academic Press.
Phillip, R. A. (2007). Mathematics teachers' beliefs and affect. In F. K. Lester (Ed.), *Second handbook on mathematics teaching and learning* (pp. 257 - 315). Charlotte,NC: Information Age Publishing.
Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children*. New York: Norton.
Plomp, T. (2006). Educational design research: a research approach to address complex problems in educational practice. Paper presented at the The Fifth International Forum on Educational Technology.
Porkess, R. (2004). *Collins dictionary of statistics*. London: HarperCollins.
Quattrone, G. A., & Jones, E. E. (1980). The perception of variabiltiy within in-groups and out-groups: Implications for the law of large numbers. *Journal of Personality and Social Psychology, 38,* 141 - 152.
Rao, C. R. (1999). *Statistics and truth. Putting chance at work* (2 ed.). Singapore: World Scientific Publishing.
Reading, C., & Shaughnessy, J. M. (2004). Reasoning about variation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 201 - 226). Dordrecht: Kluwer Academic Press.
Rein, J. R., Goldwater, M. B., & Markman, A. B. (2010). What is typical about the typicality effect in category-based induction? *Memory and Cognition, 38*(3), 377 - 388.
Rost, G. C. (2011). *Object categories provide semantic representation for 3-year-olds' word learning*. Unpublished Dissertation, Graduate College of the University of Iowa.
Sacks, H. (1984). 'On doing "being ordinary"'. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 513 – 529). Cambridge: CUP.
Sanchez, E., & Blancarte, A. (2008). Statistical thinking as a fundamental topic in training the teachers. Paper presented at the Joint ICMI/IASE Study: Statistics education in school mathematics: Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference Monterrey, Mexico.
Sánchez, E., Borim da Silva, C., & Coutinho, C. (2011). Teachers' understanding of variation. In C. Batanero, G. Burrill & C. Reading (Eds.), *Teaching statistics in school mathematics - Challenges for teaching and teacher education: A Joint ICMI/IASE Study* (pp. 211 - 221). Dordrecht: Springer Science and Business Media.
Schulz, L. E., Bonawitz, E. B., & Standing, H. R. (2008). Word, thought, and deed: The role of object categories in children's inductive inferences and exploratory play. *Developmental Psychology, 44*(5), 1266 - 1276.
Schwartz, D. L., & Goldman, S. R. (1996). Why people are not like marbles in an urn: an effect of context on statistical reasoning. *Applied Cognitive Psychology, 10,* S99 - S112.
Sfard, A. (2000). Steering (dis)course between methaphors and rigor: Using focal analysis to investigate an emergence of mathematical objects. *Journal for Research in Mathematics Education, 31*(3), 296 - 327.

Sfard, A. (2008). *Thinking as communication*. New York: CUP.
Sfard, A. (2012). Introduction: Developing mathematical discourse : Some insights from communicational research. *International Journal of Educational Research, 51–52*, 1 - 9. doi: 10.1016/j.ijer.2011.12.013
Shaughnessy, 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.
Shin, N., Jonassen, D. H., & McGee, S. (2003). Predictors of well-structured and ill-structured problem solving in an astronomy simulation. *Journal of Research in Science Teaching, 40*(1), 6 - 33.
Silverman, D. (2006). *Interpreting qualitative data* (3rd ed.). London: Sage.
Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency llusions and other fallacies. *Organizational Behavior and Human Decision Processes, 91*, 296 - 309.
Stake, R. E. (2003). Case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Strategies of qualitative inquiry* (pp. 134 - 164). Thousand Oaks, London: SAGE Publications.
Strauss, S., & Bichler, E. (1988). The development of children's concepts of the arithmetic average. *Journal for Research in Mathematics Education, 19*(1), 64 - 80.
Tyack, D., & Cuban, L. (1995). *Tinkering toward Utopia: A century of public school reform*. Cambridge: Harvard University Press.
Toulmin, S. E. (1958). *The uses of argument* (Updated edition). Cambridge: CUP.
Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105 - 110.
Usiskin, Z., & Dossey, J. (2004). *Mathematics education in the United States 2004: A capsule summary fact book*. Reston, VA: National Council of Teachers of Mathematics.
Van der Merwe, L., & Wilkinson, A. (2011). Mapping the field of statistics education research in search of scholarship. *International Journal for the Scholarship of Teaching and Learning, 5*(1).
Vygotsky, L. (1986). *Thought and language* (A. Kozulin, Trans.). Cambridge, Massachusetts: MIT Press.
Walton, D. N. (1990). What is reasoning? What is an argument? *The Journal of Philosophy, 87*(8), 339 - 419.
Watier, N. N., Lamontagne, C., & Chartier, S. (2011). What does the mean mean? *Journal of Statistics Education, 19*(2).
Watkins, A. E., Schaeffer, R. l., & Cobb, G. W. (2001). *Statistics in action*. Emeryville, Calif.: Key Curriculum Press.
Watson, J. (1998). The role of statistical literacy in decisions about risk: Where to start. *For the Learning of Mathematics, 18*(3), 25 - 27.
Watson, J. (2006). Statistical literacy at school. Growth and goals. In A. Schoenfeld (Ed.), *Studies in mathematical thinking and learning*. London: Erlbaum.
Watson, J. M., & Moritz, J. B. (2000). The longitudinal development of the understanding of average. *Mathematical Thinking and Learning, 2*(1), 11- 50.
Watson, J., & Kelly, B. (2003). The vocabulary of statistical literacy. Paper presented at the Joint Conferences of New Zealand Association for Research in Education and the Australian Association for Research in Education: Educational Research, Risks & Dilemmas, Auckland.

Watson, J., & Moritz, J. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics, 37*, 145 - 168.
Watson, J., & Moritz, J. (2000). The longitudinal development of the understanding of average. *Mathematical Thinking and Learning, 2*(1), 11 - 50.
Watson, J., Callingham, R., & Kelly, B. (2007). Students' appreciation of expectation and variation as a foundation for statistical understanding. Mathematical Thinking and Learning, 9(2), 83 - 130.
Watson, J., Kelly, B., Callingham, R., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology, 34*(1).
Wessels, H. (2006). *Types and levels of data arrangement and representation in statistics as modeled by Grade 4 - 7 learners*. Unpublished Doctoral dissertation, University of South Africa.
Wessels, H. (2007). Statistics in the South African School Curriculum: Content, Assessment and Teacher Training. Paper presented at the ICMI/IASE 2008.
Wessels, H. (2011). Statistics in the South African school curriculum. In C. Batanero, G. Burrill & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education. A joint ICMI / IASE study* (pp. 21 - 25). Dordrecht: Springer.
Wessels, H., & Nieuwoudt, H. (2011). Teachers' professional development needs in data handling and probability. *Pythagoras, 32*(1), 1 - 9.
Wild, C. (2006). On cooperation and competition. Paper presented at the Seventh International Conference on the Teaching of Statistics.
Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical inquiry. *International Statistical Review, 67*(3), 223 - 248.
Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 35 - 62): New York Academic Press.
Zwaan, R. A., & Kaschak, M. P. (2009). Language comprehension as a means of 're-situating' oneself. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 368 - 381). Cambridge: CUP.