LEARNING CORRELATION AND REGRESSION WITHIN AUTHENTIC CONTEXTS

Dierdorp, Adri

Learning correlation and regression within authentic contexts / A. Dierdorp – Utrecht: Freudenthal Institute for Science and Mathematics Education, Faculty of Science, Utrecht University / FIsme Scientific Library (formerly published as CD-β Scientific Library), no 79, 2013.

Dissertation Utrecht University. With references. Met een samenvatting in het Nederlands.

ISBN: 978-90-70786-19-9
Key words: Correlation and regression / Authentic professional practices / Informal inferential reasoning / Variability / Sampling / Coherence

Cover design: Vormgeving Faculteit Bètawetenschappen
Printed by: Drukkerij Wilco, Amersfoort

LEARNING CORRELATION AND REGRESSION WITHIN AUTHENTIC CONTEXTS

HET LEREN VAN CORRELATIE EN REGRESSIE AAN DE HAND VAN AUTHENTIEKE CONTEXTEN

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van
de rector magnificus, prof. dr. G.J. van der Zwaan, ingevolge het besluit van
het college voor promoties in het openbaar te verdedigen
op 4 november 2013 des middags te 12.45 uur

door
Adrianus Dierdorp

geboren op 6 april 1958
te Velsen, Nederland

Promotoren:	Prof.dr. J.A. van Maanen
	Prof.dr. H.M.C. Eijkelhof
Co-promotor:	Dr. A. Bakker

This dissertation was supported by funding from Platform Beta Techniek, DUDOC Program.

Table of contents

For my son and parents

Fragment Bridge over troubled water (Simon and Garfunkel)

Chapter 1
INTRODUCTION

1 Introduction

To draw inferences about real-world situations, correlation and linear regression models are applied to describe a linear relation between two variables with a formula. To construct such a model, the observed data points are represented in a scatterplot – one variable on the horizontal axis and the other on the vertical axis. After the scatterplot is drawn, the data are examined to find a pattern that can often be modelled with a line (Figure 1). The line with the best fit is called “regression line.” The formula for the regression line can be used to summarize the current situation or make predictions. The strength of this relation can be indicated by a measure called “correlation.”

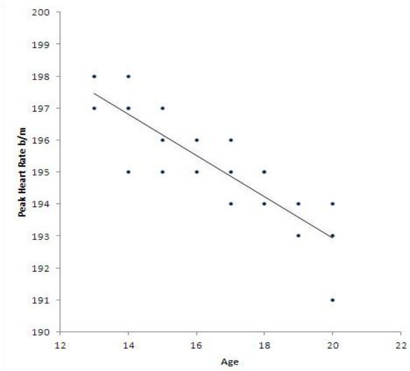

Figure 1. Regression line to approximate a relationship between age and peak heart rate.

The relevance of correlation and regression models in the real world implies the importance of supporting secondary school students’ learning and understanding of these statistical techniques. The relevance of regression models within statistics as a whole is evident from the history of statistics. Many statisticians have contributed to the development of modelling by correlation and regression. Since Pearson (1857-1936), this type of modelling has been a part of statistics (Stigler, 1986). The origin of regression can be found in the natural sciences, in real-world problems in astronomy and geodesy such as the distribution of measurement errors for heavenly bodies moving around the sun (Stigler, 1986). Nowadays, many academic statistics courses acknowledge the relevance of correlation and regression and include these techniques in their curricula. In secondary school this relevance is less trivial. For many upper-secondary school students in the Netherlands correlation and regression are an optional topic within mathematics, often taught in only a few lessons. In secondary education, mathematics textbooks often take a formal methodological approach with few or no real-world examples (Estepa & Sánchez-Cobo, 1998). Despite the techniques often being introduced through a context, the focus is on learning the techniques in a formal way. After such lessons, the students are supposed to be able to construct a formula for a regression line and calculate a measure for correlation, but it may be asked whether they understand what they are doing. The offered data set is often carefully chosen and the search for a signal amongst the “noise” (variability) has not been the focus of teaching (Konold & Pollatsek, 2002). Afterwards, students probably will still not be able to draw real-world inferences when they meet real-world problems with messy data, especially when they are working with their own sampling and measurements. These formal approaches ignore the fact that variability is omnipresent in real-world data and can affect people’s lives (Wild & Pfannkuch, 1999). When learning is tightly focused on formal learning objectives, it can become meaningless for students (Ainley, Pratt, & Hansen, 2006).

Our study is intended to find a more meaningful way of teaching correlation and regression. Rather than starting with explaining the formal techniques, what Freudenthal (1973) called anti-didactic inversion, we start with authentic applications and work towards understanding correlation and regression and their applicability. Applicability is one reason to use authentic professional practices as a basis for the design of teaching materials. Teaching that is based on authentic professional practices can provide students with a sense of relevance for real world situations and a better understanding of the real world (Berlin & White, 1992). It can motivate students to learn correlation and regression and to apply their knowledge in other scientific and real-life contexts (Bray, 1969).

Another reason to base design and teaching on authentic professional practices is that students may get the opportunity to experience coherence between mathematics and the natural sciences. Professionals in mathematics-related fields draw upon interdisciplinary knowledge when they have to solve their real-world problems (English, 2009). This is especially true when professionals have to deal with statistical inferences. Traditionally, in the Netherlands, but also in other countries, school subjects such as mathematics, biology, geography and physics are taught separately in secondary education. Despite the creation of new school subjects that combine subjects (e.g., Nature, Life and Technology), teaching subjects separately remains the core of secondary education. Separate teaching of subjects often results in the acknowledged problem that students are not able to use concepts they learn in one subject in other subjects. Bransford, Brown and Cocking (2000, p. 62) wrote: “Transfer is also affected by the context of original learning; people can learn in one context, yet fail to transfer to other contexts.” This lack of transfer points to the fact that many students do not see the coherence between different disciplines. When teaching is inspired by authentic professional practices and goes beyond mathematization by considering interdisciplinary elements, students can experience the teaching as more meaningful (Prins, 2010).

2 Purpose and main research question of the project

The purpose of this research project is to investigate how to support upper-secondary school students’ learning and understanding of statistical correlation and regression models in such a way that they can apply the techniques in new situations and experience coherence between mathematics and the natural sciences. This purpose would be achieved were we to design a valid and effective teaching and learning strategy.

By “valid” we mean that the strategy must be in line with prevailing epistemological ideas of the school subjects involved. The strategy is considered effective if students can solve a real-world problem by correctly using correlation and regression models, if they understand the concepts and process of modelling and if they are able to combine mathematical and statistical techniques with concepts from the natural sciences to solve the real-world problem. The fundamental assumption underlying our research project is that basing the design of the teaching and learning strategy on authentic professional practices can help us achieve our purpose. In line with this purpose our main research question is:

> *What are characteristics of a valid and effective teaching and learning strategy to teach students about correlation and regression in such a way that they experience coherence between mathematics and the natural sciences?*

3 Authentic professional practices

We define an authentic professional practice as the practice of a professional who works according to a characteristic procedure leading to an outcome (cf. Bulte et al., 2005). To engage students in meaningful educational tasks it is necessary to adapt the authentic professional practices for educational use. Often a simplification is needed because the real-world situation may be too complex for secondary school students (Westbroek, 2005).

Innovation commissions in the Netherlands promote context-based education. Some of these commissions for science education even promote inspiring teaching through authentic practices (Apotheker, Bulte, de Kleijn, Van Koten, Meinema, & Seller, 2007; Boersma et al., 2007). Authentic practices often use mathematical tools to model data. Mathematics has a long tradition of using models and modelling and some researchers in relation to modelling stress the importance of linking back mathematical solutions to the contextual problem (e.g., Galbraith & Stillmann, 2006). In our opinion this linking back is an important element of learning when this learning is inspired by authentic professional practices. When drawing inferences for a real-world problem, the solution must make sense in the real world.

Research into mathematical modelling shows that several educational choices are possible. Zbiek and Conner (2006, p. 89) start their report with a possible motivation for engaging students in modelling:

> To prepare students to work professionally with mathematical modelling, to motivate students to study mathematics by showing them the real-world applicability of mathematical ideas and to provide students with opportunities to integrate mathematics with other areas of the curriculum.

This is a fair reason to engage students in modelling. However, in their conclusions, Zbiek and Conner focus on mathematical concepts and procedures. For Galbraith and Stillmann (2006, p. 143) the modelling processes are the most important and “other” mathematical concepts and procedures are an “additional benefit.”

In science education there are some successful examples of teaching and learning strategies inspired by authentic practices. For example, Prins

(2010) used an authentic professional practice of water treatment to teach students chemical contents. He elaborated the claim of Bennett and Holman (2002) that it must be possible to involve students in models and modelling and evolving their understanding when the approach is linked to recognizable real-world problems. This water treatment practice is of great importance worldwide and thus recognizable for students. One of the steps in his modelling approach was regression. From a chemistry-didactical perspective, Prins focused on modelling chemical processes with regression, but we had our focus on statistics as the bridge between mathematics and professional practices in which scientific knowledge is used to solve authentic problems with correlation and regression.

There is little research on basing teaching and learning strategies for mathematics on authentic professional practices. In this study we see supporting students to learn about correlation and regression inspired by authentic professional practices as a chance to learn about how students draw inferences at school level beyond correlated real-world data. Little is known about this, even though many real-world problems deal with relationships between two or more quantitative variables and less is written about teaching correlation and regression to upper-secondary school students.

Summarizing, we investigated whether a teaching and learning strategy inspired by authentic professional practices can be a useful approach to teach students about correlation and regression in such a way that they experience coherence between mathematics and the natural sciences. This strategy includes a specially designed instructional unit and its rationale which for example describes how the unit can be used. As we explain in section 7 of this chapter, we conducted design research (Barab & Squire, 2004; Van den Akker, Gravemeijer, McKenney, & Nieveen, 2006) to answer our research question.

4 Informal inferential reasoning at the basis of four studies

Over the last decades there has been a shift in statistics education research from statistical procedures to conceptual understanding (Makar & Rubin, 2009) and informal inferential reasoning (Makar & Ben-Zvi, 2011; Pratt & Ainley, 2008). In everyday life, informal reasoning can be characterized as a process in which a person builds a model of an authentic problem (Perkins, Farady, & Bushey, 1991); it is used in non-deductive situations, such as decision making (Voss, Perkins, & Segal, 1991). In education, informal knowledge may help students to combine knowledge based on real-world experience with their pre-knowledge and should be considered in designing curricula (Gravemeijer & Doorman, 1999; Smith, diSessa, & Roschelle, 1993). In the context of statistics education a definition for informal inferential reasoning is: “the way in which students use their informal statistical knowledge to make arguments to support inferences about unknown populations based on observed samples” (Zieffler, Garfield, delMas, & Reading, 2008, p. 44). In line with the shift to informal inferential reasoning in current statistics education research, we wanted to investigate how to support students’ informal inferential reasoning when learning about correlation and regression as used by professionals. To support this reasoning it is important that students develop several statistical key concepts, for example variability and sampling (Bakker, 2004).

In our research project, modelling by means of correlation and regression was central, but we focused different studies on different aspects. In Chapter 2 we address informal inferential reasoning in the broad sense. In the next two chapters we focus on key concepts that underpin informal inferential reasoning. In Chapter 3 we present our investigation of how students reason about *variability* when engaged with measurement tasks using correlation and regression models. An advantage of involving students in measurement activities is that it invites them to make connections between the real world and the world of data. Also, measurement activities seem suitable for teaching about variability in interdisciplinary contexts. In Chapter 4 we investigate how students can develop a rich understanding of sampling by shuttling between context and statistics. This shuttling between – in our case scientific – contexts and statistics is important beyond the development of statistical concepts, such as sampling. Hence we zoom out in Chapter 5 to study more broadly to what extent students experience the coherence between mathematics and the natural sciences, with statistical concepts and modelling techniques as the bridging tools. In relation to this coherence and informal inferential reasoning Ben-Zvi and Garfield (2010, p. 359) wrote:

> ... statistics can be viewed as a type of bridge that connects mathematics and science, in that it provides the mathematical foundations for analysing data gathered in the real world (science). It seems natural to introduce and use statistics as part of both mathematics instruction (e.g., ideas of chance, measures of centre

and variability) as well as in science (e.g., variability in characteristics of plants and animals).

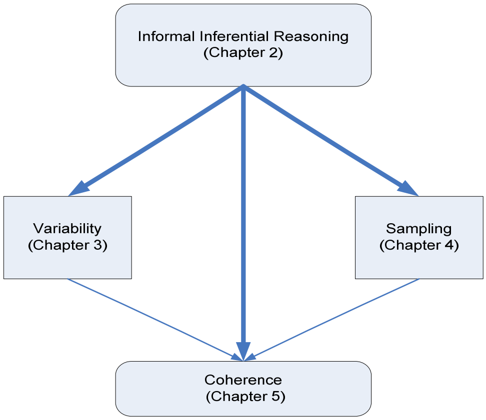

Figure 2. In this study the focus on Informal Inferential Reasoning (Chapter 2) led to studies of two key concepts underlying this reasoning, variability (Chapter 3) and sampling (Chapter 4). The first three studies, in particular the study in Chapter 2, were at the basis of our study of coherence (Chapter 5).

5 Students' informal inferential reasoning

Informal inferential reasoning (Pfannkuch, 2006a) is about drawing generalizations from data samples with respect to populations (Ben-Zvi, Gil, & Apel, 2007) or processes (Bakker, Kent, Derry, Noss, & Hoyles, 2008). In drawing informal inferences, the focus is on reasoning and conceptual understanding and less on statistical procedures (Gil & Ben-Zvi, 2011). An explicit link between statistical inferences and tendency is made by Makar and Rubin (2009, p. 85). They see "acknowledgement of a mechanism or tendency that extends beyond the data at hand" as an important idea to draw inferences. We focus on tendencies that can be modelled by means of correlation or regression. There is also evidence that if students are able to draw informal inferences, they have easier access to formal methods (Zieffler, Garfield, delMas, & Reading, 2008). For making informal statistical inferences, Makar and Rubin (2009, p. 85) considered four elements critical:

- Notions of uncertainty and variability articulated through language that breaks from the mathematical convention of claims of certainty;
- Reliance on the concept of aggregate (as opposed to individual points) through the use of generalizations about the group;
- Acknowledgement of a mechanism or tendency that extends beyond the data at hand;
- Evidence for reasoning based on purposeful use of data.

These elements are needed for accessibility to informal inferential reasoning with data. Real contexts play a significant role in students' informal inferential reasoning. Pfannkuch (2011) stressed this significance by arguing that learning about real-world situations (data in context) and the knowledge students bring to tasks and their physical and social learning environment (learning-experience-contexts) need to be taken into account when students have to develop this kind of reasoning. Pfannkuch (p. 44) reasoned about "a constant interaction between the contextual and statistical domains." Such interaction inspired us to investigate how to support students' informal inferential reasoning (Chapter 2), their shuttling between contextual and statistical domains (Chapter 4) and their experiences of coherence between scientific contexts and the mathematical underpinning of correlation and regression (Chapter 5).

5.1 Students' reasoning about variability through measurement activities

Variability is a big idea related to informal inferential reasoning (Rubin, Hammerman, & Konold, 2006). Reasoning about variability is important because it is omnipresent and has a practical effect on people's lives (Wild & Pfannkuch, 1999). Since this omnipresent variability is one of the key ideas of statistics, Wild and Pfannkuch (p. 226) see the consideration of variability as a fundamental type of statistical thinking. They distinguish the components: "noticing and acknowledging," "measuring and modelling for the purposes of prediction, explanation and control," "explaining and dealing with," and "investigative strategies." By investigative strategy they do not mean the investigative cycle (Problem, Planning, Data Collection, Analysis, Conclusions), but for example looking at the patterns of variability in many different ways. Finding a trend (regression line) or discussing a deviation from the trend is one such example. In our study we assumed that the components of variability mentioned by Wild and Pfannkuch are important for secondary school students too and that measuring and modelling are suitable activities to stimulate reasoning about variability. However, there are many types of variability; the type referred to most often in this study is variability around a model (e.g. Figure 3). This variability is partly due to real (natural) variability and partly due to measurement error.

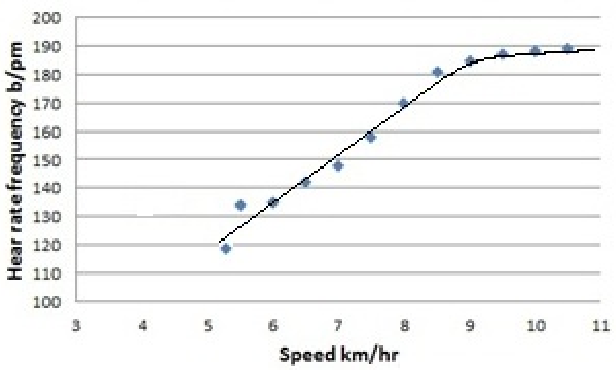

Figure 3. Variability around a model for heart rate of frequency of a student per speed of a threadmill.

Tasks about variability can be rewarding and meaningful for students. As an example, Makar and Rubin (2009, p. 83) mention: "one area of focus has been on reasoning about variation and distributions within the context of making meaning of the data." When students get the opportunity to analyse their own measurements, they have to take into account variability in order to draw a meaningful inference. However, Smith, Van den Heuvel-Panhuizen and Teppo (2011) argue that the conceptual principles that underlie measurement procedures are poorly learned by students. These researchers (2011, p. 617) wrote:

> Weak learning of measurement—particularly of the conceptual principles that underlie measurement procedures—undermines students' ability to learn and understand more advanced mathematical and scientific content and hence their access to important kinds of skilled work—both professional and not.

We suggest a learning focus on measurement, because measurement activities may be suitable to give students a sense of coherence between subjects. A reason for this is that measurement is at the interface of phenomena and data. The coordination of the two is important in authentic professional practices. In this study we investigate whether students demonstrate understanding of all four components of reasoning about variability when they engage in measurement activities inspired by authentic professional practices (Chapter 3). This would underpin our idea that measurement activities can assist in developing a rich notion of variability.

5.2 Students' conceptual development of sampling for shuttling between context and statistics

In this study we focus on another statistical key concept, sampling. In Chapter 3 the connection (shuttling) between the real world and data was made through measurement. In the study reported in Chapter 4, we designed sampling tasks based on authentic professional practices that asked for an approach in which students shuttle between the context and the statistics. According to Wild and Pfannkuch (1999, p. 28) statisticians experience this "shuttling between the contextual and statistical spheres" as a continuous process. The shuttling requires a deep understanding of several concepts, some of which are statistical (cf. Bakker & Derry, 2011). From an inferentialist perspective, students thus need to broaden the scope of using and understanding statistical concepts beyond the statistical domain. It is for this reason that in this study we focus on sampling and its underlying concepts such as sample size, randomness, distribution, informal confidence interval and the relationship between sample and population (based on Pfannkuch. 2008). In short: shuttling is the main learning goal and conceptual development underlying sampling is the means to this end.

Sampling is a neglected area in secondary school statistics education, although recently more has been published on students' understanding of sampling (e.g., Garfield, delMas, & Chance, 1999; Konold & Kazak, 2008; Saldanha & Thompson, 2003; Wild, Pfannkuch, Regan, & Horton, 2011). There is very little research on upper-secondary school students. Because sampling is a key aspect related to variability and informal inferential reasoning, we stress in our study the importance of supporting students to reason about sampling.

5.3 Coherence between mathematics and the natural sciences

Chapters 2 and 3 focus on students' connections and shuttling between scientific phenomena and data with a focus on variability and sampling. Our last study investigated the broader theme of coherence, statistics as a bridge between mathematics and the natural sciences. In this context, coherence at the attained curriculum level can be defined as students' ability to apply knowledge from both mathematics and the natural sciences when solving a real-world problem. The need for this study stems from the criterion that our teaching and learning strategy must be effective in the sense that students can solve real-world problems, understand the concepts and process of modelling and are able to combine mathematical and statistical techniques with concepts of the natural sciences to solve these problems. Edelson and Reiser (2006, p. 335) claim that "engaging learners in authentic practices can assist them in understanding the structure of knowledge, or the epistemology, of the domain under study". In line with these researchers we suggest that it can help to make the strategy effective if the students understand the concepts of mathematics and the natural sciences, but also see the coherence between them.

As part of our teaching and learning strategy we had to design an instructional unit, because while many existing educational materials do present a problem by introducing a real-world context to show the students when certain mathematical concepts can be used, they do not link back to formulate a real-world solution. This implies that students often see the different school subjects as isolated subjects (Claxton, 1991). As suggested before, students are often offered very clean noiseless data sets which results in less understanding of variability and sampling.

In sum, our teaching and learning strategy is meant to teach correlation and regression in such a way that students can experience coherence between mathematics and the natural sciences. Authentic professional practices are chosen to stimulate students' ability to use correlation and regression in new contextual situations. Informal inferential reasoning is a key part of students' statistical reasoning in this strategy. Closely related to informal inferential reasoning are variability and sampling, which are given full attention in the strategy.

6 Specific research questions

To test whether the strategy is effective it is important to know whether students are supported to make predictions by using correlation and regression models when they analyse correlated data. Therefore the question in our first study is:

RQ1: *How does a teaching and learning strategy based on authentic practices support students in making statistical inferences about authentic problems with the help of correlation and linear regression?*

As a result of the cyclic design process, indications were found that when students had to make inferences, they needed to reason about variability. Therefore the instructional unit was adapted and investigated in study 2:

RQ2: *How do secondary students consider variability within measurement activities based on authentic professional practices?*

To study the support of students' conceptual understanding of another statistical key concept, sampling, we investigated in study 3:

RQ3: *What is the potential of tasks based on authentic professional practices to support students in developing concepts that underlie sampling in such a way that they can shuttle between contextual and statistical spheres?*

To know whether the strategy of teaching statistical modelling inspired by authentic professional practices offers students the possibility to experience coherence between mathematics and the natural sciences, we addressed the following question in the last study:

RQ4: *To what extent can professional practices serve as meaningful contexts to show students connections between mathematics, statistics, science and professional practices?*

7 Educational context

All student participants were upper-secondary school students (Grade 11 and 12, aged 16-18 years) at three similar Dutch schools and all within the pre-university track (VWO). These grades were chosen because the strategy was conducted to contribute to a recently introduced school subject, "Nature, Life and Technology." Since 2007, NLT is an optional subject for students of science streams in addition to the usual science and mathematics subjects. About 40% of Dutch schools offer this subject to their students. The main aims of NLT (Eijkelhof & Krüger, 2009, p. 2) are to:

1. offer pupils a modern view on science and technology, which includes insight into current developments in a wide range of topics, the cooperation between various types of experts and the role of mathematics in science;
2. challenge pupils to study developments in science and technology in depth;
3. offer options in choosing topics according to the interest of pupils, the expertise of teachers and opportunities in the region of the school;
4. assist pupils in their orientation on a career in science and technology;
5. contribute to continuous innovation in science and mathematics education.

Besides the author, two other teacher-volunteers were involved in this study. The background of the first teacher (also the researcher) is mathematics and biology, the second teacher is a mathematician and the last teacher's background is physics and mathematics. The author is the principal researcher, the co-promotor contributed to the data collection, offered editorial support and acted as an advisor, the promotors contributed editorial support, advice and general supervision.

8 Outline of the PhD project

8.1 Design research

The design research we conducted consisted of three phases. In phase 1, we designed a teaching and learning strategy inspired by a literature study and meetings which were held with students, teachers, academic experts from educational institutes, universities and innovation commissions and professional experts.

First, there was the challenge of finding suitable authentic professional practices. As the decision was taken to base the strategy on educationalized authentic professional practices (AuPPs) we formulated the following set of selection criteria based on the aims of NLT, the criteria for the selection of authentic modelling (Prins, Bulte, van Driel, & Pilot, 2008) and advice from respective professionals:

1. Professionals in the AuPP use correlation and regression models.
2. The modelling activities seem suitable to adapt for students in grades 11 and 12.
3. The AuPP offers a chance for students to perform a measurement experiment.
4. A sufficiently rich and authentic data set from this AuPP was at our disposal.
5. The AuPP should give students the opportunity to identify themselves with the professional.
6. The AuPP should induce students to appreciate the educational activities.
7. The AuPP should give students the opportunity to make their own inferences based on the data.
8. The authentic activities from the AuPP should give students the opportunity to experience coherence between mathematics and natural sciences.

It was not easy to find authentic professional practices that met all these criteria. Some practices were too difficult for educational use (criterion 2). Often more than two variables were involved or the regression model was not linear. Fifteen AuPPs were investigated. Table 1 indicates for each AuPP which criteria were met (+) or not (-).

Input on criteria led to the choice of three AuPPs that met most criteria: “sport physiologist,” “researcher water treatment for monitoring height of dykes,” and “calibrator.” From the three chosen AuPPs the sport physiologist’s practice was the only one that met criterion three. For this AuPP a small pilot with two students was conducted. This resulted in a first draft of the educational teaching and learning strategy to be tested in the second phase of the design research. The AuPPs of researchers monitoring the height of dykes to prevent flooding and professionals calibrating measuring instruments were chosen because they met criterion 4.

Phase 2 of the design research contained an iterative approach of designing the strategy, evaluation and revision. Our revisions were mainly concerned with elaborating variability to help students to think more about variability in statistical data. Especially the first cycles needed more attention to variability. The strategy was tested six times over a period of three years (2008 – 2011). We call each of these six experiments a "macrocycle." (Figure 4).

Table 1
Potentially AuPPs and the eight criteria.

AuPP	Variables	criterion 1	2	3	4	5	6	7	8
Bookbinder	Back vs. Thickness	+	+	+	-	+	-	+	+
Crash tester	Power vs. Deviation	+	-	+	-	+	+	+	+
Bungee jump	Time vs. Weight	+	-	+	-	+	+	+	+
Laboratory technician	Power vs. Elongation line (fishing)	+	+	+	-	+	-	+	+
Sport physiologist	Heart rate vs. Effort (power / time)	+	+	+	+	+	+	+	+
Laboratory technician	Sand content vs. Solids	+	-	+	-	-	-	+	+
Sport physiologist	Step size vs. Velocity	+	+	+	-	+	+	+	+
Laboratory technician	Light intensity vs. Amperage	+	-	+	-	+	+	+	+
Laboratory technician	Concentration vs. Speed of sound	+	+	+	-	+	-	+	+
Laboratory technician	Forward voltage vs. Temperature	+	+	+	-	-	-	+	+
Laboratory technician	Light transmission vs. Concentration	+	+	+	-	-	-	+	+
Sport physiologist	Weight vs. Contraction of muscle	+	+	-	-	+	+	-	-
Laboratory technician	Sound vs. Distance	+	+	+	-	+	+	-	+
Calibrator	E.g. Temperature vs. Temperature	+	+	-	+	+	+	+	+
Researcher Water treatment	Time vs. Height deviation	+	+	-	+	+	+	+	+

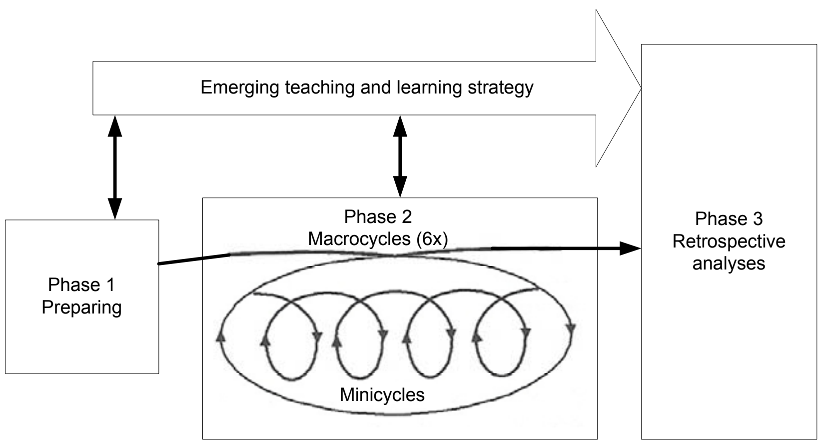

Figure 4. Design research (based on Gravemeijer & Cobb, 2007).

Each macrocycle contained around nineteen lessons (minicycle). The analyses of the lessons often influenced the next lessons. During four macrocycles all lessons were recorded. In two cycles it was only possible to record part of the lessons.

Phase 3 includes the retrospective analyses of all the research findings. Our research questions required the use of a mixed method approach. The use of qualitative as well as quantitative approaches helped to get a holistic view of the research outcomes. The quantitative approach examines whether the interventions of the design research are effective and the qualitative approaches examines how and why they work.

8.2 The design of the instructional unit

The special designed instructional unit includes three chapters, each with several tasks and subtasks (See Appendix B). Each chapter was inspired by one AuPP. The intention of the first chapter was to introduce conceptual ideas in an intuitive way and stimulate students' motivation to learn. The intention of the second chapter was to provide students with a more formal approach to the concepts and chapter three to give the students the possibility to apply what they had learned.

We based chapter one on the practice of sport instructors and physiologists who identify the best training program for clients. Apart from introducing students to the need for statistics in such an authentic practice, this chapter was also meant to familiarize them with collecting data, scatterplots, variability, sampling and correlation in an informal sense.

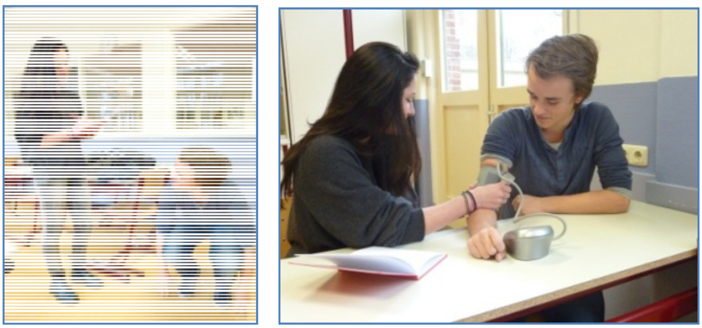

Figure 5. Students from the fifth macrocycle measuring physical condition using a sphygmomanometer (right figure).

In chapter one, students had to visually estimate the linear part of the model, but it had been planned that students would later see the need for a more reliable method to find the formula for a regression line and to measure correlation. This was intended to motivate them to learn about regression lines and correlation in relation to scientific applications. We further assumed that if the teacher inspired students to participate in classroom discussions, they would be able to reason in depth about variability.

Two key tasks were measuring condition (Figure 5) and measuring the peak heart rate and the threshold heart rate frequency (Figure 6).

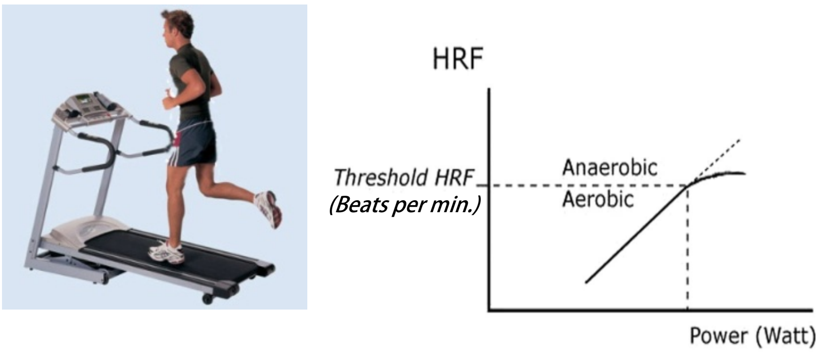

Figure 6. Figure from the instructional unit to explain to students how they could perform the experiment of running a treadmill to determine their peak heart rate and their threshold heart rate frequency.

We based chapter two on the practice of monitoring the height of dykes. A dyke is a natural or artificial slope or wall to regulate water levels, preventing flooding. A major problem is that the height of dykes decreases over time. Dykes should therefore be monitored to predict when action must be taken to prevent flooding. This authentic problem was introduced by showing the students a video.

The students were given authentic, satellite generated data (Dentz et al., 2006) for the detection and monitoring of surface deformation. Students had to model the data set to predict the day to heighten the dyke (to simplify the problem seasonal cyclic issues are ignored), which shows the deformation of the height at a particular location on a certain dyke over eight years (Figure 7).

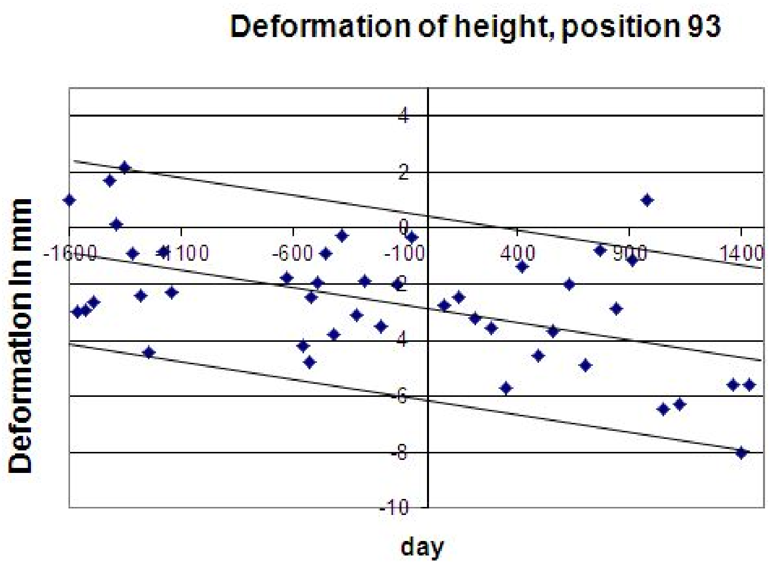

Figure 7. Picture made by a student using Excel to predict the last day to decide about heightening a dyke. The student used the outer lines to indicate the margin around the regression line (in the middle).

The key educational task in chapter two is to draw an inference about when the height of a dyke should be increased to prevent flooding. The students needed to have an aggregate view of a data set, presented in scatterplots, in order to see trends. To support students' understanding of correlation and regression, this part was designed as follows: First, students would speculate about possible strategies for producing a regression line. It was also expected that once they saw the meaning of using regression lines when making predictions in the dyke monitoring context they would learn more about the mathematical background of regression. Next, using a spreadsheet, the students learned how to calculate the formula of a regression line and how to calculate correlation coefficients. At the end of the second chapter the students practiced their newly-acquired skills in a new situation: to find and judge the regression line for two measuring instruments after they made a scatterplot for two similar variables (temperature versus temperature) instead of two different variables (heart rate versus power or deformation versus day).

We based chapter three on the authentic professional practice of calibrating measurement devices with the help of regression and correlation. An authentic key task for the students was to analyse the data from two thermometers and making decisions about which thermometer was most suitable for a particular purpose. In this case, correlation, which was elaborated in the first two chapters, was used as a measure for how well the measurements from a thermometer matched a standard. What was new in this chapter was that there was no time dimension when using regression lines and there was no future value to predict. The students were expected to experience a non-trivial transfer to a new situation of comparing two variables of the same kind.

9 Structure of the thesis

The PhD thesis comprises a set of articles (Chapters 2, 3, 4 and 5), each addressing a different perspective of this research to answer the main research question (Chapter 1). Table 2 gives an overview of the titles with the corresponding research questions of each study.

Table 2
Outline of the research

Chapter	Title	Main research question
1	Introduction	What are characteristics of a valid and effective teaching and learning strategy to teach students about correlation and regression in such a way that they experience coherence between mathematics and the natural sciences?
		Research question
2	Authentic practices as contexts for learning to draw inferences beyond correlated data.	How does a teaching and learning strategy based on authentic practices support students in making statistical inferences about authentic problems with the help of correlation and linear regression?
3	Secondary students' considerations of variability in measurement activities based on authentic practices	How do secondary students consider variability within measurement activities based on authentic professional practices?
4	Supporting students' conceptual development for shuttling between context and statistics: The case of sampling	What is the potential of tasks based on authentic professional practices to support students in developing concepts that underlie sampling in such a way that they can shuttle between contextual and statistical spheres?
5	Meaningful statistics in professional practices as a bridge between mathematics and science: An evaluation of design research.	To what extent can professional practices serve as meaningful contexts to show students connections between mathematics, statistics, science and professional practices?
6	Conclusion and discussion	

References

Ainley, J., Pratt, D., & Hansen, A. (2006). Connecting engagement and focus in pedagogic task design. *British Educational Research Journal, 32*(1), 23–38.

Apotheker, J., Bulte, A., De Kleijn, E., Van Koten, G., Meinema, H., & Seller, F. (2010). *Scheikunde in de dynamiek van de toekomst over de ontwikkeling van scheikunde in de school van de 21e eeuw* [Chemistry in the dynamic of the future; on the development of chemistry in the 21st century school]. Enschede the Netherlands: SLO.

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools.* Utrecht, the Netherlands: CD-Bèta Press.

Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning, 13,* 5–26.

Bakker, A., Kent, P., Derry, J., Noss, R., & Hoyles, C. (2008). Statistical inference at work: The case of statistical process control. *Statistics Education Research Journal, 7*(2), 130–145.

Barab, S. A., & Squire, K. D. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences, 13*, 1–14.

Bennett, J., & Holmann, J. (2002). Context-based approaches to the teaching of chemistry: What are they and what are their effects? In J. K. Gilbert, O. De Jong, R. Justi, D. Treagust, & J. Van Driel (Eds.), *Chemical education. Towards research based practice* (Vol. 17, pp. 165–184). Dordrecht, the Netherlands: Kluwer.

Ben-Zvi, D., & Garfield, (2010). Introducing the emerging discipline of statistics education. *School Science and Mathematics, 108*, 355–361.

Ben-Zvi, D., Gil, E., & Apel, N. (2007, August). *What is hidden beyond the data? Helping young students to reason and argue about some wider universe.* Paper presented at the Fifth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-5), University of Warwick, UK.

Berlin, D. F., & White, A. L. (1992). Report from the NSF/SSMA Wingspread Conference: A network for Integrated Science and Mathematics Teaching and Learning. *School Science and Mathematics, 92*, 340–342.

Boersma, K. Th., Van Graft, M., Harteveld, A., de Hullu, E., De Knecht van Eekelen, A., … Mazereeuw, M., (2007*). Leerlijn biologie van 4 tot 18 jaar. Uitwerking van de concept-context benadering tot doelstellingen voor het biologieonderwijs* [Biology curriculum from 4 to 18 years. Elaboration of the concept-context approach to targets for biology education]. Utrecht, the Netherlands: NIBI.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Bulte, A. M. W., Klaassen, K., Westbroek, H. B., Stolk , M. J., Prins, G. T., Genseberger, R., De Jong, O., & Pilot, A. (2005). Modules for a new chemistry curriculum, research on a meaningful relation between contexts and concepts. In P. Nentwig & D. Waddington (Eds.), *Context based learning of science* (pp. 273–299). Münster, Germany: Waxmann.

Bray, E. C. (1969). MINNEMAST. An elementary math-science program. *School and Mathematics, 69*, 541–547.

Claxton, G. (1991). *Educating the enquiring mind: The challenge for school science.* New York: Harvester Wheatsheaf.

Dentz, F., Halderen, L. van, Possel, B., Esfahany, S. S., Slobbe, C., & Wortel, T. (2006). *Poseidon. On the potential of satellite radar interferometry for monitoring dikes of the Netherlands*. Delft, the Netherlands: Delft University of Technology.

Edelson, D. C., & Reiser, B. J. (2006). Making authentic practices accessible to learners: Design challenges and strategies. In R. Keith Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences.* (pp. 335–354). New York: Cambridge University Press.

English, L. D. (2009). Promoting interdisciplinarity through mathematical modeling. *ZDM - The International Journal on Mathematics Education, 41*, 161–18.

Estepa, A., & Sánchez-Cobo, F. T. (1998). Correlation and regression in secondary school text books. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee & W.-K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics*, (vol. 2, pp. 671–676). Voorburg, the Netherlands: International Statistical Institute.

Eijkelhof, H. M. C., & Krüger, J. (2009, August). *Improving the quality of innovative science teaching materials.* Paper presented at the European Science Education Research Association conference (ESERA), Istanbul, Turkey.

Galbraith, P., & Stillmann, G. (2006). A framework for identifying student blockages during transitions in the modeling. *Zentralblatt für Didaktik der Mathematik, 38*, 143–162.

Garfield, J., delMas, R., & Chance, B. (1999). *Developing statistical reasoning about sampling distributions.* Presented at the First International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL), Kibbutz Be'eri, Israel.

Gil, E., & Ben-Zvi, D. (2011). Explanations and context in the emergence of students' informal inferential reasoning. *Mathematical Thinking and Learning, 13*(1-2), 87–102.

Gravemeijer, K., & Cobb, P. (2007). Design research from a learning design perspective. In J. J. H. Van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds), *Educational design research.* London: Routledge, Taylor & Francis.

Gravemeijer, K., & Doorman, K. (1999). Context problems in realistic mathematics education: A calculus course example. *Educational Studies in Mathematics, 39*(1/3), 111–129.

Freudenthal, H. (1973). *Mathematics as an educational task.* Dordrecht, the Netherlands: Reidel.

Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education, 2*(1), 1–37.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*(4), 259–289.

Makar, K., & Ben-Zvi, D. (2011). The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning, 13*(1&2), 1–4.

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

Perkins, D. N., Farady, M., & Bushey, B. (1991). Everyday reasoning and the roots of intelligence. In J. F. Voss, D. N. Perkins, and J. Segal (Eds.), *Informal reasoning and education* (pp. 83–105). Hillsdale, NJ: Lawrence Erlbaum.

Pfannkuch, M. (2006a). Informal inferential reasoning. Working cooperatively in statistics education In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics.* Voorburg, the Netherlands: International Statistical Institute.

Pfannkuch, M. (2008, July). *Building sampling concepts for statistical inference: A case study.* Paper presented at the International Congress on Mathematical Education, Monterrey, Mexico.

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.

Pratt, D., & Ainley, J. (Eds.). (2008). Informal inferential reasoning. Special Issue of the *Statistics Education Research Journal, 7*(2), 3–4.

Prins, G. T. (2010). *Teaching and learning of modeling in chemistry education*. Utrecht, the Netherlands: CD-Bèta Press.

Prins, G. T., Bulte, M. W., Van Driel, J. H., & Pilot, A. (2008). Selection of authentic modelling practices as contexts for chemistry education. *International Journal of Science Education, 30*(14), 1867–1890.

Rubin, A., Hammerman, J. K. L., & Konold, C. (2006). Exploring informal inference with interactive visualization software. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics* [CD-ROM]. Voorburg: the Netherlands: International Statistical Institute.

Saldanha, L., & Thompson, P. (2003). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51*, 257–270.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences, 3*(2), 115–163.

Smith III, J. P., Van den Heuvel-Panhuizen, M., & Teppo, A. R. (Eds.). (2011). Learning, teaching, and using measurement [Special issue]. *ZDM - The International Journal on Mathematics Education,43(5).*

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, Mass.: Belknap Press of Harvard University Press.

Van den Akker, J. J. H., Gravemeijer, K., McKenney, S., & Nieveen, N. (2006). *Educational design research.* London: Routledge, Taylor & Francis.

Voss, J. F., Perkins, D. N., & Segal, J. W. (1991). Preface. In J. F. Voss, D. N. Perkins, & J. Segal (Eds.), *Informal reasoning and education* (pp. vii–xvii). Hillsdale, NJ: Lawrence Erlbaum Associates.

Westbroek, H. B. (2005). *Characteristics of meaningful chemistry education - The case of water quality.* Utrecht: CD Bèta Press.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*, 223–265.

Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. J. (2011). Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society, 174*, 247–295.

Zbiek, R., & Conner, A. (2006). Beyond motivation: Exploring mathematical modeling as a context or deepening students' understanding of curricular mathematics. *Educational Studies in Mathematics, 63*, 89–112.

Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistical Education Research Journal, 7*(2), 40–58.

Chapter 2

AUTHENTIC PRACTICES AS CONTEXTS FOR LEARNING TO DRAW INFERENCES BEYOND CORRELATED DATA

Dierdorp, A., Bakker. A., Eijkelhof, H. M. C., & Van Maanen, J. A. (2011). Authentic practices as contexts for learning to draw inferences beyond correlated data. *Mathematical Thinking and Learning, 13*, 132-151.

Abstract

To support 11th-grade students' informal inferential reasoning, a teaching and learning strategy was designed based on authentic practices in which professionals use correlation or linear regression. These practices included identifying suitable physical training programmes, dyke monitoring, and the calibration of measurement instruments. The question addressed in this study is: How does a teaching and learning strategy based on authentic practices support students in making statistical inferences about authentic problems with the help of correlation and linear regression? To respond to this question we used video-recordings of lessons, audio-taped interviews, classroom field notes, and student work from a teaching experiment with 12 Dutch students (aged 16–17 years). The analysis provided insights into how the teaching and learning strategies based on authentic practices supported them to draw inferences about authentic problems using correlated data. The evidence illustrates how an understanding of the authentic problem being solved, collecting their own data to become acquainted with the situation, and learning to coordinate individual and aggregate views on data sets seemed to support these students in learning to draw inferences that make sense in the context.

Keywords: *statistical reasoning, inferential reasoning, authentic professional practice*

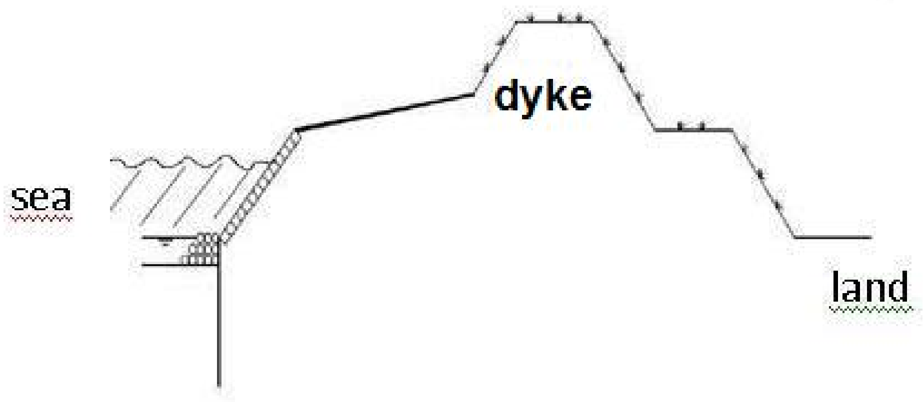

Figure 1. Schematic picture of a dyke

1 Introduction

This paper analyses how a teaching and learning strategy, based on authentic practices using correlation and regression, supported the development of informal inferential reasoning by 11thgrade students (16–17 year olds). Informal inferential reasoning (Pfannkuch, 2006) can be characterized as the reasoning processes required to make informal statistical inferences, which Makar and Rubin (2009) defined as probabilistic generalizations beyond the data that are supported by data-based evidence. The importance of looking "beyond the data" (Curcio, 1987, p. 384) to some wider universe is widely acknowledged to be at the heart of statistics. The GAISE report (Franklin et al., 2005), for example, argued that by the end of secondary school students should have learned to look beyond the data, hinting at generalizations from a sample to a process or a population.

Statistical inference, typically connoting formal statistical techniques, has long been considered a topic too difficult for most secondary school students. To give students a sense of what it means to draw a statistical inference without formal techniques such as hypothesis testing, researchers over the past few years have worked on ways to study and develop informal statistical inferences at primary and middle school levels (Ben-Zvi, 2006; Paparistodemou & Meletiou-Mavrotheris, 2007) or in the workplace (Bakker, Kent, Noss, & Hoyles, 2009), but rarely at upper-secondary school level, the focus of this paper. Despite the growing interest in informal inferential reasoning, research in this area is still in its infancy. Most studies concern the comparison of two data sets (e.g., boys vs. girls; grade 6 vs. grade 7), but little is known about how students draw inferences at school level beyond correlated data, even though many real-world problems deal with relationships between two or more quantitative variables. Learning to draw conclusions through correlation and regression is important because it allows students to study trends and to learn to provide statistical evidence for predictions beyond correlated data.

One of the problems with the teaching of correlation and regression to 11th-graders is that there is a lack of educational materials that take advantage of informal inferential reasoning in contexts that are meaningful to students. Existing materials are often quite formal and hardly engaging (Vuijk, 2001), or otherwise too informal for grade 11 (Roodhart, Kindt, Pligge, & Simon, 1998).Wild, Pfannkuch, Regan, and Horton (2009, p. 2) noted more generally:

> "Mathematics teachers desperately need interesting learning activities with obvious real-world relevance and statistics is capable of satisfying that need."

Statistics, mathematics, and science education have increasingly aimed to be "context-based" to show students the relevance of what they learn (Lijnse & Klaassen, 2004; Schoenfeld, 1985; Shaughnessy, Garfield, & Greer, 1996). One trend in science education that we find particularly promising for statistics education is that of using authentic practices as a source of inspiration for designing educational materials. In our study, we considered authentic professional practices in which particular functional scientific knowledge is needed, for example the practice of predicting dyke heights, for which regression lines are used. The underlying conjecture we made is that students will then more easily see the point of developing such knowledge than they do in traditional approaches in which theory is taught prior to application. For example, inspired by real weather forecasting, Lee and Songer (2003) designed weather forecasting tasks in which students had to collect local data, compare weather data from geographically different regions, interpret real-time weather maps, and make forecasts. They found that this approach helped students to develop rich understanding about scientific knowledge.

In line with this idea we have designed a teaching and learning strategy (explained in more detail later) on the basis of authentic practices in which professionals use correlation or linear regression to deal with scientific problems that arise in their work. Our approach is in line with the aims of a new school subject called Nature, Life, and Technology for grades 11 and 12 (16–18 year olds), which the Dutch government introduced to foster better connections between mathematics and the natural sciences and to make these subjects more attractive to students. The research reported in this paper deals with students taking this subject and was designed to investigate the conjecture that when learning activity contexts are chosen from authentic practices that are recognizable and relevant from a student perspective, students can be supported to learn statistical reasoning about real-world phenomena. This paper therefore focuses on the following research question:

How does a teaching and learning strategy based on authentic practices support students in making statistical inferences about authentic problems with the help of correlation and linear regression?

With informal inferential reasoning beyond correlated data as a learning goal, we first address the conceptual understanding which, according to the literature, might support this. Next we summarize the literature on how authentic practices can inspire the design of educational materials that foster such conceptual understanding. Then we elaborate on our teaching and learning strategy by describing our educational materials and the accompanying hypothetical learning trajectory. Subsequently we describe the setting of this research and the data sources, and clarify the analytical method followed. The research question and findings are addressed in the last two sections.

2 Theoretical background

2.1 Learning goal: informal inferential reasoning supported by linear regression models

Makar and Rubin (2009, p. 85) argued that the following ideas are critical for students in order to be able to draw statistical inferences:

1. notion of *uncertainty* and *variability* articulated through language that broke from mathematical convention of claims of certainty,
2. reliance on the concepts of *aggregate* (as opposed to individual points) through the use of generalizations about the group,
3. acknowledgement of a *mechanism* or *tendency* that extends beyond the data at hand,
4. *evidence* for reasoning based on purposeful use of data.

The development of these ideas requires considerable support. For example, students' difficulties in understanding the sources of variability in data have been widely reported (Moritz, 2004; Shaughnessy & Noll, 2006; Wild & Pfannkuch, 1999). It is also well known from the statistics education literature that aggregate thinking and identifying tendencies are not easily developed by school students (Bakker & Gravemeijer, 2004; Konold & Higgins, 2002). More generally, it is known that people find it hard to use data as evidence for conclusions (Andriessen, 2006; Kuhn Berland & Reiser, 2009). Although some researchers report valuable student intuitions about correlation (Cobb, McClain, & Gravemeijer, 2003; Moritz, 2004), educational research further shows students' difficulties in judging correlation and regression, both on an intuitive level and after teaching. The literature reports that seeing trends in scatter plots is far from trivial (Batanero, Green, & Serrano, 1998; Ben-Zvi, 2000), even for professionals who know the context very well (Hoyles, Bakker, Kent, & Noss, 2007; Noss, Pozzi, & Hoyles, 1999). A focus on the aforementioned list of ideas guided the design of our teaching and learning strategy in addition to the design heuristics elaborated in the next section.

2.2 Design heuristics: using authentic practices as inspiration for educational materials

A key concern for educational researchers is tomake scientific concepts and activities meaningful to students. Over the past decades various approaches, such as context-based science and mathematics education, have been proposed and tested (Bennett, Lubben, & Hogarth, 2007). Contexts, however, can be used in very different ways (Gilbert, 2006). One extreme is to use contexts as areas of application without framing such settings in advance. This approach has often been criticized. Freudenthal (1973), for example, referred to it as an "anti-didactic inversion" and suggested confronting students with problem situations first so as to promote "guided reinvention," discovering particular mathematical knowledge under the guidance of the teacher and the educational materials. Another extreme is to ask students to participate in authentic out-of-school practices such as environmentalist groups (Roth & Lee, 2006). This is, however, not possible in mass schooling situations.

Our own use of context is that of using authentic practices in which particular statistical or scientific knowledge is used as a rich source of inspiration for designing educational materials. Such practices have clear motives and the scientific knowledge used is functional (Prins, Bulte, Van Driel, & Pilot, 2009). If it is possible to adapt such a practice to an educational context that students still recognize as relevant, one might expect students to see the need to learn and use particular scientific knowledge. For many science educators this has been a reason to seek inspiration for the design of educational materials for science in authentic practices (Edelson, 1998; Lee & Songer, 2003; Prins et al., 2009; Roth & Lee, 2006;Westbroek, 2005;Westra, 2008).

Westbroek (2005) and Westra (2008) combined this approach of basing educational materials on authentic practices with a problem-posing approach (Lowrie, 2002; Lijnse & Klaassen, 2004; Kortland, 2001). This latter approach emphasizes that students should always know what the motives for their learning activities are. This requires designers to investigate how elements from authentic practices could be "educationalized" to elements of a learning environment in which students experience the concepts learned and the learning itself as relevant (Westbroek, Klaassen, Bulte, & Pilot, 2010). Relevant elements of an authentic practice are *motives* to perform certain *actions* and *procedures,* and *knowledge* is used as a tool to achieve particular *goals*.

Westbroek (2005), for example, based her educational materials on the professional practice of water quality testing to engage students in the question of how drinking and swimming water are tested in practice. In line with the problem-posing approach this helped students to predict the next step in testing water quality and develop the chemical knowledge required in this step. The students worked with measurement tools similar to, but simpler than, those used in actual practice. In this way, Westbroek designed an educational version of an authentic practice that gave students a clear goal and made the chemical knowledge to be learned meaningful to them. Prins and colleagues (2009) proceeded along these lines and used authentic motives for modeling drinking water treatment. They concluded that an authentic practice can offer inspiration for designing an environment that meaningfully involves students in modeling processes. Since the emphasis in educational materials designed by Prins and associates is on chemistry, the statistical knowledge of regression required to model water treatment processes was at the service of chemical modeling. In contrast, we chose regression and correlation as central means to support informal inferential reasoning.

3 Teaching and learning strategy

As mentioned in the introduction, there were no materials available that matched our ideas of a teaching and learning strategy on how to support the intended learning. We therefore decided to design materials ourselves by means of educational design research (Barab & Squire, 2004; Van den Akker, Gravemeijer, McKenney, & Nieveen, 2006). We use the term "teaching and learning strategy" to include both the concrete educational materials (teaching unit with three chapters, Excel spreadsheet, and TI-Nspire software) and a hypothetical learning trajectory (HLT). Simon (1995) defined HLT in the context of a lesson as a description of students' prior knowledge, learning goals, and hypotheses about how the learning processes were supported by the educational materials. Bakker (2004) used the HLT notion for longer-term learning processes (about 15 lessons). In such longer learning trajectories, hypotheses can be global and local. The global ones address the overarching ideas, which in our case are variability, aggregate, tendency, and evidence (see previous section); the local ones are tied to each specific task. Given that we have already discussed the learning goals of the teaching and learning strategy, we now summarize only students' prior knowledge and the ideas behind the educational materials.

3.1 Students' prior knowledge

Dutch students typically learn less about statistics than students in countries that are well represented in the statistics education literature (e.g., New Zealand Ministry of Education, 2007; Council for theMathematical Sciences, 2004; National Council of Teaching Mathematics, 2000). By the end of grade 10, Dutch students have learned little more than some descriptive statistics with a focus on graphical representations and measures of center and spread. Scatter plots were new to our students in grade 11. In this paper we first report on "learning to make statistical inferences about an authentic problem" and elaborate on four conceptual ideas that still had to be developed:

1. supporting students to coordinate individual and aggregate views on data sets,
2. recognizing a trend that extended beyond the data,
3. making sense of variability,
4. using data as evidence when drawing an inference.

The students were not used to aggregate features of data sets such as trends, or in this case regression lines, and they were not used to thinking about variability. In the educational materials specially designed as part of the teaching and learning strategy we created opportunities for students to reason informally about underlying trends, to extend their understanding of these trends in order to make better inferences. Our students also had little experience with spreadsheets.

3.2 Educational materials

We briefly describe the educational materials and accompanying HLT. The educational materials consisted of three chapters, each based on an authentic practice in which correlation and/or regression was used to address a problem.

3.2.1 Chapter 1: Identifying a Sports Program to Improve Physical Condition

The first chapter was based on the practice of sport instructors and physiotherapists who identify the best training program for clients (mostly athletes). Apart from introducing students to the need for statistics in such an authentic practice, this chapter also served to acquaint students with collecting data themselves and to familiarize them with scatter plots, variability, and correlation in an informal sense. To illustrate the HLT of this chapter we first describe a central task: just like athletes, the students had to collect data of their heart rate frequency (HRF) when increasing the intensity of their training. HRF increases proportionally with the invested power. If the efforts exceed a certain point, threshold HRF, the linear proportionality will disappear and the HRF will approach the peak heart rate (Gellish et al., 2007). The threshold HRF is where required energy does not only rely on aerobic sources anymore but starts to draw on anaerobic sources. When drawing on anaerobic sources the body produces lactic acid, which causes muscle pain and consequent “burn” during intensive training. It is important for athletes to keep their use of energy in the aerobic area, because if someone frequently trains in the anaerobic area, his or her threshold HRF will decrease, and muscle problems will occur at lower HRFs. The upper bound where the HRF still acts as a linear function of the invested power, called the point of deflection (Conconi, Ferrari, Ziglio, Droghetti, & Codeca, 1987), is a good approximation for the threshold HRF (Figure 2), from which the optimal HRF for training programs can be determined.

Determining the threshold HRF can be seen as an informal statistical inference because it is measured only indirectly through a point of deflection that is derived from a set of data that inevitably involve measurement error. Therefore it is a generalization beyond the data which should be phrased in probabilistic terms (i.e., non-deterministic or uncertain language) (Makar & Rubin, 2009).

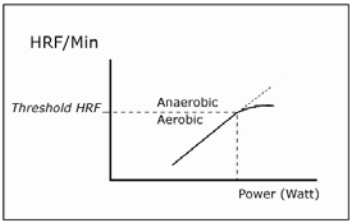

Figure 2. Theoretical model of heart rate frequency as a function of power during a cycling test (reconstructed from Conconi et al., 1987).

Like for many other statistical inferences, especially about scientific phenomena, a graphical representation of the data is crucial. The global conjecture from the HLT for Chapter 1 was that students would recognize the linear trend in the data up to a certain value and find the threshold HRF from a scatter plot. In Chapter 1 visually estimating the location of the linear model suffices, but it was hoped that students would later see the need for more sophisticated and reliable methods to measure correlation and model data sets. This would motivate them to learn more about regression lines and correlation in relation to scientific applications. It was further assumed that if the teacher stimulated students to reason in classroom discussions, that they would get a feel for natural variability as well as measurement error through their experience of collecting data. This would presumably give them a sense of which mechanisms cause variability while perceiving tendencies in the data.

3.2.2 Chapter 2: Monitoring the height of dykes

The second chapter is based on the practice of monitoring the height of dykes. A dyke is an artificial longitudinal land elevation with the aim of preventing flooding. Dyke monitoring is a vital issue for the Netherlands because the majority of the Dutch population lives below sea level. A major problem is that dykes' heights decrease over time, for instance through subsidence (the downward displacement of land relative to sea level). At some points (called "critical values" by the Ministry of Transport and Water Management) they are so low that high sea water levels are a danger. Dykes should therefore be monitored so action can be taken to prevent flooding. Moreover, sudden height changes can indicate weaknesses in the dykes. We introduced this authentic problem by showing the students a video. To find a way to deal with this problem we gave students authentic data (Dentz et al., 2006) generated by satellites for the detection and monitoring of surface deformation. The data set could be modelled with a linear function (we chose to ignore the seasonal cyclic issues to simplify the problem), which shows the deformation of the height of a certain dyke at a particular position over several years.

The key educational task in Chapter 2 is to make an inference about when the height of a dyke should be increased, just as professional research organizations for dyke monitoring do. These predictions require students to have an aggregate view of a data set, presented in scatter plots, in order to see trends. To support students' understanding, the chapter was designed as follows. First, students speculate about possible criteria for producing a regression line. Once they see the power of regression lines when making predictions in the dyke monitoring context they learn more about the mathematical background of regression. Next, using a spreadsheet the students learn how to make scatter plots, how to calculate the formula of a regression line, and how to calculate correlation coefficients. At the end of the chapter the students practice the skills they have developed in a new situation.

3.2.3 Chapter 3: Calibration of thermometers

The third chapter assesses students' aptitude for making inferences with the help of regression and correlation. The students were given a final assignment about the calibration of thermometers to make decisions about which thermometer was most suitable for a particular purpose. In this case, correlation, which was elaborated in the first two chapters, can be used as a measure for how well the measurements of a thermometer fit those of a standard. What makes this problem context different from the dyke monitoring is that there is no time dimension when using regression lines and there is no future value to predict. This implies a non-trivial transfer to a different situation of comparing two variables of the same kind instead of one being time. Whereas correlation coefficients were not used to make predictions in the dyke monitoring context, here they help students to make a decision about which thermometer is the best.

4 Method

To gain insight into how the teaching and learning strategy supported students' inferential reasoning, we analyzed the data collected in a teaching experiment with

12 students, from December 2008 to March 2009. This teaching experiment was part of a design-based research with three design cycles (Barab & Squire, 2004; Van den Akker, Gravemeijer, McKenney, & Nieveen, 2006), but given the paper's main question we do not report on the cyclical design process. All students (5 boys, 7 girls) were from the grade 11 pre-university track, consisting of academically successful students who had chosen to study the aforementioned new school subject Nature, Life, and Technology. Courses within this subject typically involve project-based learning and put less emphasis on direct instruction and the textbook than typical mathematics and science chapters. Students participated in classroom discussion but also worked in pairs and small groups. Our strategy required students to reason more than they were used to.

Data collection in the teaching experiment included transcripts from video-recorded lessons and audio-taped interviews, field notes, and a final assignment. The teaching and learning strategy was enacted by the first author in 19 lessons (a total of 16 hours) each of which consisted of several tasks.

In order to give an impression of how students' learning was supported by the educational materials and how well this learning was predicted in the HLT we compared the actual learning trajectory (ALT) with the conjectures from the HLT for each task (while we recognize that it is not possible to detect students' "actual" learning, we chose to use this more conventional phrase to describe the observed learning that was inferred from the data we collected.) To find evidence and counter-evidence for the conjectures formulated in the HLT, the transcripts, field notes, interviews, and final assignment were analyzed with the help of a data analysis matrix.

Table 1 shows 1 of the 94 rows (one for each task or subtask, columns 1–2) in the data analyses matrix used to compare our predictions (column 3) with the actual observations (columns 4–6). This subtask will be described in the results section. The fourth column gives representative examples of transcripts from classroom interaction, followed by a conclusion or clarifying comments based on field notes, and transcripts. The result column (Res) summarizes how well the HLT and ALT match by means of "–", "±", or "+." The − sign was used when the observations suggested that the conjectures were confirmed for a maximum of one-third of the students. The + sign was used when observations suggested that the conjectures formulated in the third column were confirmed for at least two-thirds of the students; for the intermediate cases we used the ± sign. The choice of the three categories (−, ±, +) was motivated by the fact that two categories would be too coarse for an evaluation and more than three categories would suggest more precision than can be justified. In cases where the observations did not include all students, the clarification column also includes the number of observations (e.g., “based on eight observations”). In five cases we were not able to assign one of the three categories and left the matrix cells empty because the observations did not provide us with enough information. The first author made the first coding decisions; the others judged the texts of the columns “Conjecture,” “Transcript excerpt,” and “Clarification” on clarity and consistency and did 10 random checks of the “Result” column. This led to small changes of the content of several matrix cells.

In this paper, we focus on those tasks that involved statistical inferences to look for a general trend in the actual learning of drawing contextually realistic inferences. On the basis of the observations of the aforementioned analyses we illustrate students’ struggles and successes in order to gain insight into how the teaching and learning strategy supported students to make statistical inferences beyond correlated data from authentic practices.

Table 1
Row 5f from data analysis matrix

No	***Task***	***Conjecture***	***Transcript Excerpt***	***Clarification***	***Res***
5f	The graph of Gellish’s formula is an approximation of reality. The Measured values did not lie exactly on the line of the formula. What might the graph in reality have looked like? Make a sketch of at least 30 points	Students take variability into account	T: Do you think all Gellish’s data points are just on this line [made by Leon]? Leon: No, they are scattered around the line.	4 of 8 observed students used a ruler to draw the line. Only 2 made a cloud of data points	-

5 Result

5.1 Learning to make statistical inferences about an authentic problem

To address the research question, we compared the findings of the ALT with the conjectures in the HLT and the data analysis matrix explained in the method section. For each inference-related task Table 2 shows to what extent our conjectures were supported by the data. The conjectures in the HLT were often supported by the ALT and more so toward the end than in the beginning. Because many of these conjectures were of the form “students will learn/understand that …” we concluded that the overall pattern was that students tended to make better inferences about authentic problems as they gained experience, even though the tasks became more difficult toward the end. We also conclude that the teaching and learning strategy supported the students’ learning process reasonably well. To give a sense of the quality of inferences we next give a few examples in chronological order. Tasks illustrated below in subsequent sections, including 5f, which we predicted badly, are indicated by bold type numerals in Table 2. All tasks in this table involve informal inferential reasoning (IIR).

Table 2
ALT results compared with HLT conjectures for the tasks involving IIR

Task:	5d	**5f**	6a	**6c**	7	8	9c	9e	10b	11c	15
+				x	x			x	x	x	x
±	x		x								
-		x				x	x				

Task:	17	23b	23c	**24aI**	**24aII**	**24c**	**25d**	**34a**	34c	42
+	x		x	x		x	x	x		x
±		x							x	
-					x					

Note: Tasks 5d until 17 concern the authentic practice of identifying a sports program to improve physical condition; the others concern monitoring the height of dykes.

5.2 Supporting students to coordinate individual and aggregate views on data sets

In mathematics lessons the students had always used mathematical functions rather than correlated data. Hence, we first wanted to ensure they understood the origin of linear models such as Gellish's formula: PHR = 207 – 0.7A, where PHR is peak heart rate and A is age. After a brief introduction to scatter plots we therefore asked students to sketch what Gellish's original data points might have looked like before he came up with this formula. We expected that students would sketch similar "clouds" but now surrounding a line.

Despite our attempt to focus the students' view on the variability of the individual data points behind the aggregate of Gellish's formula, half of the observed students used a ruler to draw a line instead of plotting a cloud of data points. Two other students drew points that were almost collinear. Only two students took natural variability into account and scattered the points around the line representing Gellish's formula. Table 1 summarizes the comparison of the HLT and ALT for this task.

When the teacher subsequently asked the students who used the ruler if the data points that Gellish found were all exactly on the line, they realized the points must have been scattered around the line. In this task, help with understanding variability was apparently not sufficient in the first HLT, but the episode also indicates that what students learned here was something new to them. Apparently it was necessary to offer students more support than we had anticipated to think through what real data summarized by a linear trend might look like. Even with these academically bright students a brief introduction appeared necessary to help them link the aggregate view of the line with the data-point view underlying Gellish's formula. In future research we will investigate if this can be done by adjusting the materials and making use of explicit classroom discussion.

Another interpretation is that their conflict between their experience with functions previously (no variability) and the variability they were confronted with in the tasks supported them in developing new, albeit still emerging, understandings of variability in authentic contexts (Makar, Bakker, & Ben-Zvi, this issue). In other words, the fact that they initially saw this as perfectly linear allowed them to confront this assumption, and this was facilitated by the authentic nature of the problem context.

We wondered why students had struggled with subtask 5f and realized that they had had too little experience with collecting data and with this type of correlated data. Moreover, from a student perspective there might not have been a good reason to sketch such data—in contrast to the aims of a problem-posing approach. To support students making realistic inferences we realized it was important to bring students into a position, guided by the design of the learning and teaching strategy, to formulate the need for extending their knowledge. In a next subtask, 6c, the teacher more carefully prepared a motive for students' activities. Preferably students themselves should be aware of the authentic problem and formulate this problem to be further investigated. We suggest that if students have a better view of the authentic problem, it will be easier for them to have an aggregate view on the data. For example, the teacher tried to assist students to understand the point of the next task.

In this case, after reviewing the information on aerobic and anaerobic use of energy in a classroom discussion, the teacher hoped students would "reinvent" the idea of doing a physical test in order to have a more complete view on the problem. One student, Jolan, used the phrase "stress test" to describe a test similar to the one further on in the educational materials. Her description of the "stress test" suggests that she was not thinking of individual measures but had ideas of what the model would look like, facilitating an aggregate view of the data when she would draw a graph of her data collected in the next task.

In the next lesson the students went to the gym to do a test similar to the one Jolan described. In the running test the students gathered their own data in order to make an informal inference about the threshold point. During the physical test, Leon kept a record of Jolan's heart rate values. The teacher wanted to support Leon making an inference and tried to understand if Leon had an aggregate view on the data.

T: Do you already see a trend?

Leon: Well, it does not go up very quickly. Now it is [increasing by] 4. Before that 3, and before that 4. Before those 13; and before that 21.

Leon, like others, seemed to focus on the individual deviations of the data points instead of the aggregate. He noticed that the values were reaching the limit and did not "go up very quickly," but he did not demonstrate a more aggregate view of the data. Some students showed evidence of an aggregate view already from their graph. Most students drew a graph corresponding to the model in Figure 1, but some drew a graph like that in Figure 2.

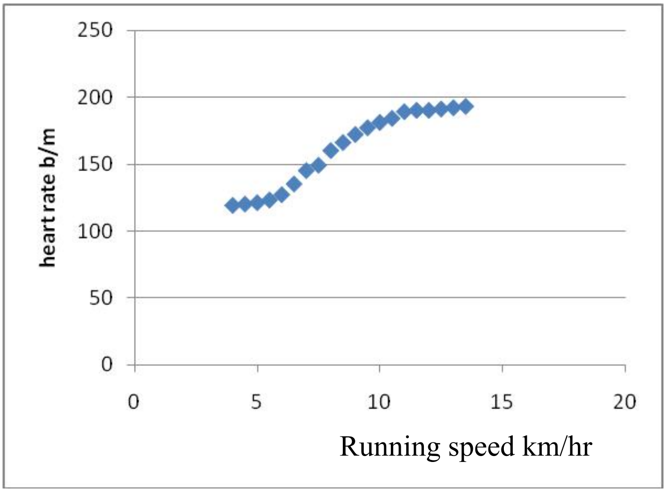

Figure 3. Graph made by Maria of her data collected in the gym (Task 6c).

Initially, it was not clear if Maria had demonstrated an aggregate view of the data necessary to identify the threshold value (which might not be one of the data points); however, when the teacher said that her graph (Figure 3) looked like Figure 1, Maria did not fully agree:

Maria: You can clearly see a difference between walking and running. When you suddenly start to run, your heart rate accelerates.
T: [addressed to the whole class] Maria has another remark to make!
Maria: That there is suddenly a steep rise. I think that's the point of deflection from walking to running.

Maria's comments suggest that she indeed seemed to adopt the aggregate view required by linking the data to the contextual problem. She was not focused on the individual data points but recognized a trend and modelled her data. She noticed that her graph differed from the graph in Figure 1, owing to another deviation from a linear trend (Figure 3). Immersion in the contextual situation most likely helped her find a plausible explanation for this (cf. Gil & Ben-Zvi, this issue; Roth & Lee, 2006) and thus to look beyond the data.

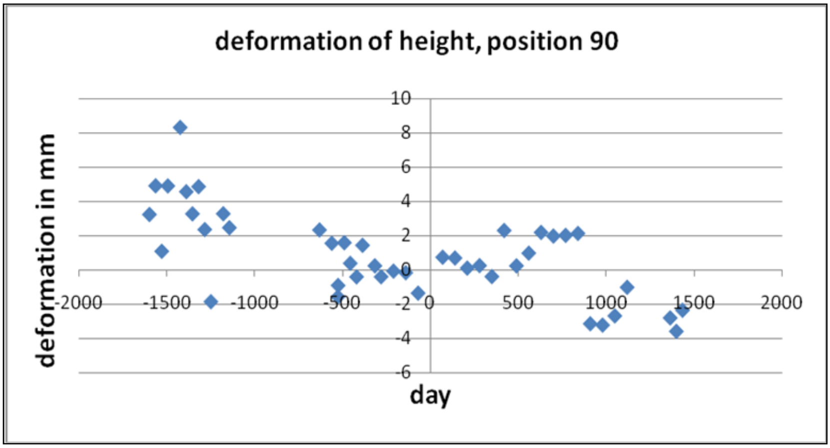

Figure 4. Data used in Task 24: deformation of height from a dyke at position 90 for 3034 days (more than eight years). Day = 0 is taken as reference point for comparison with other positions.

5.3 Recognizing a trend that extended beyond the data

For a different type of reasoning about trends in order to make informal inferences we jump to Chapter 2 in which students learned to produce scatter plots using a spreadsheet (Excel). Subtask 24a showed a scatter plot (Figure 4) and asked, “What happens with the height of the dyke at this position during these years and what intervention may be needed?” In the HLT we conjectured that the students by thinking through the context would recognize a downward trend and suggest a realistic solution, taking into account a safety margin (e.g., sea levels might be higher than normal, there might be a storm or high tide, and ground movements cause temporary changes to the height of the dykes). Our analysis suggested that the students recognized the trend beyond the data and made an inference about heightening the dyke, but they struggled to articulate a solution that drew on the context of the authentic problem (see the + and − in Table 3). During this task the focus was on recognizing the trend. The students struggled to find a realistic solution in which variability is taken into account.

Table 3

Row 24a from the Data Analysis Matrix

No	Task	Conjecture	Transcript excerpt	Clarification	Res
24a I	What happens with the height of the dyke at this position during these years?	Students recognise a downward trend.	Ciska: The dyke becomes less high. T: Yes, especially at the end of the scatter plot you recognised a clear trend.What can you do? Ciska: Heighten the dyke.	Most students recognised the trend.	+
24aII	What intervention may be needed?	The students can find a realistic solution.	Sanno:Produce a regression line. T: And when you have the regression line, what then? Sanno:Then you can make a prediction when to heighten the dyke.	Most students had problems finding a realistic solution. They used only the regression value.	-

5.4 Making sense of variability

In line with the HLT, the teacher asked the students to give reasons for the variability in the deformation of the height of the dyke (subtask 24c), and thus encouraged them to treat the data as "numbers with a context" (Moore, 1992, p. 15). This was considered important to assist students to recognize the variation around a roughly linear trend, something that they found difficult during the aforementioned Task 5. We allowed students to ignore the cyclic effect caused by ground movements because this would require too much time to be spent on geographical information.

T: You saw [in the video] that after the flooding disaster a standard (critical value) was formulated for the height of dykes, and that nowadays still 30% of the dykes. . .

Marco: Are not high enough.

T: Yes, they didn't meet the standard.

Maria: I understand that. Sea level rises also. They can heighten the dykes, but after 50 years these are too low again, because the sea level rises.

T: Yes, you have to intervene in time.

Maria: Yes, but I understand that they can't do everything right.
T: How can you anticipate?
Maria: Heighten at once. Heighten more than needed at the moment.

This example illustrates how the teacher supported these students' awareness of the contextual problem with reference to the introductory video the students had watched. We conjectured students would see that variability affects the timing of heightening the dyke, and Maria indeed seemed to do so (when probed).

Several students (e.g., Susan) did not immediately see the importance of a margin given the variability of both the dyke height and the sea level (subtask 24c), but some did (e.g.,Maria using probabilistic [uncertain] language like "about," "could," and "surprise"). Our idea was to focus their attention on the need for a measure of this variability.

T: Why could it be important to have a measure, a norm, for variability?
Maria: Then you can see how much they deviate. If you think, because of the regression line, that there will be flooding in about 5 years, you do not necessarily have to do anything. But because there is variation, it could be that there is a too low position after three years.
T: Yes, indeed.
Jolan: So, you have to take a lower height into account.
Susan: But why is it important?
Maria: To prevent an unpleasant surprise.

Table 4
Results of the final assignment

Students	*Draw Regression Line*	*Calculate Regression Line*	*Calculate Correlation*	*Conclusion*	*Link the Context*
Marco & Danny	+	+	+	+	+
Birte & Leon	+	+	+	+	+
Sanno & Ciska	+	-	+	+	-
Maria & Jolan	+	-	+	+	-
Theo & Susan	+	+	+	+	+
Hanne & Wiebe	-	-	+	+	+

Note: A plus sign refers to an appropriate answer, a minus sign to an inappropriate or missing answer. The final assignment is described in the section on the educational materials.

This last remark from Maria satisfied Susan (analysis of Susan's final assignment suggested that she came to understand the importance of variability, see Table 4). This last excerpt suggests that the use of an authentic practice helped students to think about variability in concrete terms such as varying dyke heights, surprises, and chance of flooding. The analysis provided further evidence that they gained insight into correlated data in relation to the authentic problems in using the context of the problem to explain their reasoning. We also think that the authentic practices chosen were appropriate to foster students' awareness of the role of variability. The contextual problems from these practices appeared to help students understand the relevance of variability and seemed to encourage students to make realistic inferences. The teacher-led classroom discussions appeared to further assist students by stressing the role of variability in the problem contexts.

5.5 Using data as evidence when drawing an inference

Understanding variability brought the students closer to addressing a key aim of Chapter 2, which asked them to use the data set as evidence to predict the day when a dyke should be heightened. The next transcript (from subtask 25d) illustrates how these students understood the role of regression lines within this context.

T:	But how do you know what height is needed?
Maria:	Looking at the data of the rising sea level, and watch the subsidence.
Sanno:	Construct a regression line!
T:	What about this regression line?
Sanno:	Then you can predict the day to take action.

Sanno probably meant that the formula of the regression line would give him the opportunity to predict the (precise) day before which action has to be taken, possibly forgetting about the variability of dyke height.

The context about monitoring the heights of dykes is complex. Drawing a valid inference — in this case a prediction of when to heighten a dyke — not only requires coordinating knowledge about regression lines and the context of the data but also knowledge of risk (cf. Eijkelhof, 1996; Pratt et al., in press). In the unlikely event that dykes are relatively low and water levels high, such a situation can have a disastrous effect; hence there is a clear need to stay well above the given critical value. In theory, it is possible that students interpreted the given critical value as already taking such safety margins into account. However, we have no data suggesting they did at this stage.

In the next example it appeared that students had no problems with formal calculations (subtask 34a) but struggled to find a solution for the authentic problem, because they again did not take the variability into account.

Susan: You have to see when it [the regression value] is –25.
Jolan: You can construct a formula.
Susan: Yes, we have a and b [regression coefficients].
Jolan: Then you can use your calculator.
T: Yes, but now you want to know when the deformation is −25.
Sanno: You substitute H=−25 [using the formula for the regression line]. I get d = 5116.9 [day].
Susan: That's the moment the flooding starts.

From this wording ("the moment the flooding starts") we infer that Susan probably did not realize that the regression line only predicts when the dyke height might reach a particular value. Other students did not take the safety margin into account either. After some interrogation by the teacher Danny realized this, but he did not know how to deal with this extra complication.

T: What did you calculate?
Marco: The a and b, from the regression line [as in $y = ax + b$], I suppose.
Danny: From the trend line: $y=-0.00533x + 2.273104$.
T: Yes, only we have now an H [height] and a d [day].
Danny: Is that also the critical value?
T: No.
Danny: How can you calculate that one?

Similarly Susan confirmed that she did "not know how to calculate the safety margin." During the final interviews students expressed that they had learned how to calculate the formula for a regression line, but did not mention the role of variability.

R: What did you learn in these lessons?
Danny: I learned how to use a spreadsheet. In the beginning I didn't know how to use it, now I do.

R: So, you didn't use a spreadsheet often?
Danny: Yes, but now more. I understand it better now, and I have learned how to make a graph and a regression line.

The last excerpts and results from the final assignment (Table 4) suggest that students were capable of calculating the formula for the regression line and making inferences based from these lines, but considering and calculating a safety margin remained difficult for them.

The students' final assignment suggested that they were able to make realistic inferences about which thermometer was best to use for sterilization in a hospital pharmacy. Although regression lines and formulas were not strictly necessary to answer this question they were used by five of the six student pairs. All students correctly computed the correlations between thermometer data and a golden standard (see Table 4) and judged the thermometers to be accurate enough based on high correlations (around 0.999). Four pairs of students made critical remarks about their conclusion, showing they had kept an eye on the context. For example, Birte and Leon wrote:

> Whether this is sufficient depends on the function of the thermometer. For measuring the body temperature it is allowed to be less accurate than for measuring the air. In this case it is a thermometer for sterilisation in a hospital pharmacy. This does not have to be quite accurate.

This quote illustrates how students drew informal statistical inferences about authentic problem situations with the help of correlation. We supported them by teaching them in the first two chapters how to produce the formula for a regression line and how to calculate the correlation. We also allowed a lot of time for students to work out the meaning of this all in relation to the authentic problems. As stated before, few students had problems using the learned techniques in the new situation of the calibration of thermometers. We noticed, however, that most students struggled with the fact that the practices of the first two chapters were dealing with a time variable, whereas the thermometers needed an approach where they had to compare two variables that were the same.

6 Discussion

This paper addresses the question of how a teaching and learning strategy supported 11th-graders to make statistical inferences beyond correlated data about authentic problems from professional practices. We summarize our response to the research question, then reflect on the role of context, followed by a discussion of limitations and future research, and finally draw some conclusions regarding the study.

6.1 Learning to make statistical inferences about authentic problems

The results section provides evidence that students' inferential reasoning was reasonably well supported by the educational materials and shows that their observed learning was reasonably well predicted in the HLT. From the analysis it appears that the teaching and learning strategy focusing on core ideas such as coordinating individual and aggregate views on data sets, recognizing trends, making sense of variability, and using data as evidence, helped students to make more and better informal statistical inferences over the course of the teaching experiment. We also provided examples in which students struggled with an aggregate view on a data set.

Learning about correlation and regression seemed meaningful to students because they recognized its role in authentic situations. Recognizing a trend that extended beyond the data seemed to be accessible for most of the students, but making a realistic prediction for the dyke height context made the students struggle, because they often forgot to take the variability into account. We also illustrated through the excerpts that the use of authentic practices also provided the students with multiple challenges. For example, one of the points we intend to improve in a next design cycle is how to address the problem of risk involved in dyke monitoring. With regard to risk (cf. Pratt et al., in press), we had underestimated the difficulty of taking it into account in the dyke context. Giving students a critical value of a dyke height leaves open many issues such as: What risk of flooding do we consider acceptable? What safety margin is already included in a critical value defined by the Ministry of Transport? What variability in sea water levels can we expect?

Another point for consideration is that students had become used to the time dimension in the contexts of Chapters 1 and 2, and to using regression rather than correlation, which hindered transfer to situations such as calibration, in which correlation is not time-dependent but involves two variables of the same type. The results of the final assignment suggest, however, that the authentic practices of the first two chapters helped students to use correlation and linear regression to understand that practitioners in many fields use these techniques in their daily professional work and to realise they can use these techniques in many situations themselves.

6.2 Reflection on context

We chose authentic practices as the contexts for our educational materials, and our study suggests that this had both advantages and challenges. A major advantage was that students generally saw the practical value of correlation and regression and were generally able to draw realistic inferences. The need to account for variability was more critical in the more complex context of Chapter 2, allowing students to build on the foundational skills of data collection and linear modelling from Chapter 1. Apart from challenges in finding suitable practices that appeal to students and in which the desired techniques are used, a principal challenge for the designer is to appropriately modify the level of complexity of authentic practices for student use. For educational purposes the authentic problems and tools in our study needed to be simplified so students could reasonably work with them. For example, dyke monitoring is much more complex than predicting a linear downward trend. There may be sudden subsidence or weaknesses in the structure of the dykes, and there are seasonal effects in ground movements that we chose to ignore in the educational materials. Additionally, we had to choose the knowledge that students needed about causes of variability carefully because many more factors can influence dyke height and measurement error than we could accommodate. Keeping students focused on the learning goals thus requires careful design but also places high demands on the teacher in terms of their own knowledge of the context.

6.3 Limitations and future research

A teaching experiment on correlation and regression with a small group of students in the highest tier of general education can only suggest “proof of principle” that it is possible to base the teaching and learning strategy on authentic professional practices. In addition, the designer of the educational materials was also the teacher in this case. Additional research with more students, other topics, and other teachers is required before we can draw more general conclusions about whether such teaching and learning strategies based on authentic practices is effective at a larger scale. In this paper we have not focused on teaching because we first wanted a better understanding of how the strategies would be enacted in principle. It is for this reason that we used the term “hypothetical learning trajectory”; in future research we intend to use the term “hypothetical *teaching* and learning trajectory.” When working with other teachers, it is necessary to articulate more explicitly what teaching approaches they could adopt to benefit from the potential of the teaching and learning strategy. In the light of our experience in Dutch classrooms we assume that the common focus on self-reliant learning (Bos, 1996) is at odds with our own approach, which involved peer interaction and classroom discussions on “how?” and “why?” questions about contextual problems.

7 Conclusions

Given the results of this study we conclude that the teaching and learning strategy generally supported students in learning to draw informal statistical inferences beyond correlated data. Students were generally able to link much of their formal statistical knowledge and scientific concepts to contextual problems when making informal statistical inferences. Although the study involved only 12 students and favorable conditions, this study suggests strong potential for basing statistics instruction on authentic practices. In these authentic contexts, statistical concepts and techniques are used for purposes that students can relate to and can therefore potentially support students’ learning, provided the teaching and the educational materials are of sufficient quality. As common in design-based research, support cannot be attributed to one feature of the designed teaching and learning strategy but is most likely the result of several features that are attuned to each other. We mention a few of these possible features as well as some deliberations behind them.

The tasks, inspired by authentic problems, seemed realistic enough so that students experienced authenticity and felt engaged. The tasks were of increasing difficulty and provided students with many opportunities to reason about scatter plots, variability, regression lines, and correlation in relation to contextual problems. The tasks in which students had to collect and model their own data fuelled the perceived need to find a solution for the contextual problem and to study the educational materials. This was also enforced by following the problem-posing approach in which students had to formulate the need for extending their knowledge of mathematics and the natural sciences. We suggest that this support assisted students to make statistical inferences about an authentic problem, but also helped students grasp ideas underlying statistical inference such as coordinating individual and aggregate views on data sets, recognizing trends that extended beyond the data, making sense of variability, and using data as evidence when drawing inferences.

References

Andriessen, J. (2006). Arguing to learn. In K. Sawyer (Ed.), *Handbook of the learning sciences* (pp. 443–460). Cambridge, UK: Cambridge University Press.

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Utrecht, the Netherlands: CD-Bèta Press.

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

Bakker, A., Kent, P., Noss, R., & Hoyles, C. (2009). Alternative representations of statistical measures in computer tools to promote communication between employees in automotive manufacturing. *Technology Innovations in Statistics Education, 3*(2). Retrieved from: http://www.escholarship.org/uc/item/53b9122r

Barab, S. A., & Squire, K. D. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences, 13*, 1–14.

Batanero, C., Green, D. R., & Serrano, L. R. (1998). Randomness, its meanings and educational implications. *International Journal of Mathematical Education in Science and Technology, 29*(1), 113–123.

Bennett, J., Lubben, F., & Hogarth, S. (2007). Bringing science to life: A synthesis of the research evidence on the effects of context-based and STS approaches to science teaching. *Science Education, 91*, 347–370.

Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning, 2*, 127–155.

Ben-Zvi, D. (2006). Scaffolding students' informal inference and argumentation. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics* (CD-ROM), Salvador, Bahia, Brazil, 2–7 July, 2006. Voorburg, The Netherlands: International Statistical Institute.

Bos, M. (1996), Zelfwerkzaamheid? Zelfstandig leren! [Self-activation? Self-reliant learning!]. *Euclides*, *72*(1)*,* 15–18.

Cobb, P., McClain, K, & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction, 21*, 1–78.

Conconi, F., Ferrari, M., Ziglio, G., Droghetti, P., & Codeca, L. (1987). Determination of the anaerobic threshold by a noninvasive field test in runners. *Journal of Applied Physiology, 52*, 869–873.

Council for the Mathematical Sciences. (2004). *An international review of UK research in mathematics*. London, UK: Author.

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*, 382–393.

Dentz, F., Halderen, L., Possel, B., Esfahany, S. S., Slobbe, C., & Wortel, T. (2006). *Poseidon. On the potential of satellite radar interferometry for monitoring dikes of the Netherlands*. Delft, the Netherlands: Delft University of Technology.

Edelson, D. C. (1998). Realizing authentic science learning through the adaption of scientific practice. In B. J. Fraser & K. G. Tobin (Eds.), *International handbook of science education* (pp. 317–331). Dordrecht, the Netherlands: Kluwer Academic.

Eijkelhof, H. M. C. (1996). Radiation risk and science educaton. *Radiation Protection Dosimetry, 68*, 273–278.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry,M., & Schaeffer, R. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) college report*. Alexandria VA: American Statistical Association. Freudenthal, H. (1973). *Mathematics as an educational task*. Dordrecht, the Netherlands: Reidel.

Gellish, R. L., Goslin, B. R., Olson, R. E., McDonald, A., Russi, G. D., & Moudgil, G. V. (2007). Longitudinal modeling of the relationship between age and maximal heart rate. *The American College of Sports Medicine, 39*(5), 822–829.

Gilbert, J. (2006). On the nature of "context" in chemical education. *International Journal of Science Education, 28*, 957–976.

Hoyles, C., Bakker, A., Kent, P., & Noss, R. (2007). Attributing meanings to representations of data: The case of statistical process control. *Mathematical Thinking & Learning*, *9*, 331–360.

Konold, C., & Higgins, T. (2002). Highlights of related research. In S. J. Russell, D. Schifter, & V. Bastable (Eds.), *Developing mathematical ideas: Working with data* (pp. 165–201). Parsippany, NJ: Dale Seymour.

Kortland, J. (2001). *A problem posing approach to teaching decision making about the waste issue*. Utrecht, the Netherlands: CD-Bèta Press.

Kuhn Berland, L., & Reiser, B. J. (2009). Making sense of argumentation and explanation. *Science Education, 93*(1), 26–55.

Lee, H. S., & Songer, N. B. (2003). Making authentic science accessible to students. *International Journal of Science Education, 25*, 923–948.

Lijnse, P. L., & Klaassen, K. (2004). Didactical structures as an outcome of research on teaching-learning sequences? *International Journal of Science Education, 26*, 534–554.

Lowrie, T. (2002). Designing a framework for problem posing: Young children generating open-ended tasks. *Contemporary Issues in Early Childhood, 3*, 354–364.

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

Moore, D. S. (1992). Teaching statistics as a respectable subject. In F. S. Gordon (Ed.), *Statistics for the twenty-first century* (pp. 14–25). Washington, DC: Mathematical Association of America.

Moritz, J. B. (2004). Reasoning about co-variation. In J. Garfield & D. Ben-Zvi (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 227–256). Dordrecht, the Netherlands: Kluwer.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

New Zealand Ministry of Education. (2007). *Effective pedagogy in mathematics Pangarau: Best evidence synthesis*. Lead by G. Anthony and M. Walshaw. Wellington, New Zealand: New Zealand Ministry of Education, Crown.

Noss, R., Pozzi, S., & Hoyles, C. (1999). Touching epistemologies:Meanings of average and variation in nursing practice. *Educational Studies in Mathematics, 40*, 25–51.

Paparistodemou, E., & Meletiou-Mavrotheris, M. (2007, August). *Enhancing reasoning about statistical inference in 8-year-old students*. Paper presented at the Fifth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-5), University of Warwick, UK.

Pfannkuch, M. (2006). Informal inferential reasoning. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics* (CD-ROM), Salvador, Bahia, Brazil, 2–7 July, 2006. Voorburg, the Netherlands: International Statistical Institute.

Pratt, D., Ainley, J., Kent, P., Levinson, R., Yogui, C., & Kapadia, R. (in press). Informal inferential reasoning as a contextualised risk-based enterprise. *Mathematical Thinking and Learning.*

Prins, G. T., Bulte, A. M.W., Van Driel, J. H., & Pilot, A. (2009). Students' involvement in authentic modelling practices as contexts in chemistry education. *Research in Science Education, 39*, 681–700.

Roodhart, A., Kindt, M., Pligge, A., & Simon, A. N. (1998). *Mathematics in context: Get the most out of it*. Chicago: Encyclopaedia Britannica Educational Corporation.

Roth, W. M., & Lee, Y. (2006). Contradictions in theorizing and implementing communities in education. *Educational Research Review, 1*(1), 27–40.

Roth, W. M., & McGinn, M. K. (1997). Deinstitutionalizing school science: Implications of a strong view of situated cognition. *Research in Science Education*, *27*, 497–513.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando, FL: Academic Press.

Shaughnessy, J. M., Garfield, J. B., & Greer, B. (1996). Data handling. In A. Bishop (Ed), *International handbook of mathematics education* (pp. 205–238). Dordrecht, The Netherlands: Kluwer Academic.

Shaughnessy, J. M., & Noll, J. (2006, November). *School mathematics students' reasoning about variability in scatterplots*. Paper presented at the twenty-eighth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education. Universidad Pedagógica Nacional, Mérida, México.

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

Van den Akker, J. J. H., Gravemeijer, K., McKenney, S., & Nieveen, N. (2006). *Educational design research*. London, UK: Routledge, Taylor & Francis.

Vuijk, R. (2001). *Getal & Ruimte: Zebra correlatie en regressie*.[Numbers & Space: Zebra correlation and regression]. Groningen, the Netherlands: EPN.

Westbroek, H. B. (2005). *Characteristics of meaningful chemistry education*. Utrecht, The Netherlands: CD-Bèta Press. Westbroek, H., Klaassen, K., Bulte, A., & Pilot, A. (2010). Providing students with a sense of purpose by adapting a professional practice. *International Journal of Science Education, 32*, 603–627.

Westra, R. (2008). *Learning and teaching ecosystem behavior in secondary education*. Utrecht, the Netherlands: CD-Bèta Press.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*, 223–265.

Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. (2009, June). *Precursor statistical inferences*. Paper presented at the United States Conference on Teaching Statistics, Ohio State University, Columbus, Ohio.

Chapter 3

SECONDARY STUDENTS' CONSIDERATIONS OF VARIABILITY IN MEASUREMENT ACTIVITIES BASED ON AUTHENTIC PRACTICES

Abstract

This study analyses considerations of variability by students in Grade 12 (aged 17-18) when they engage in measurement activities that we designed on the basis of authentic professional practices in which correlation and regression are used. Analysis of students' reasoning during these activities in one classroom (N = 13) suggests that students considered variability in several ways: noticing and acknowledging variability, measuring and modelling variability, explaining variability and using investigative strategies. We conclude that the measurement tasks based on authentic professional practices helped students to reason about relevant aspects of variability. Finally, curricular and research implications are discussed.

Keywords: *Variability; Measurement; Modelling; Authentic Professional Practice; Correlation and Regression*

1. Introduction

Variability is everywhere. Variability is the phenomenon that something is apt or liable to vary or change (Reading & Shaughnessy, 2004). Wild and Pfannkuch (1999, p. 235) stress that “variability affects all aspects of life and everything we observe. No two manufactured items are identical, no two organisms are identical or react in identical ways.” Repeated measurements also vary. Such variability leads to a state of not having enough information, certainty and knowledge to accurately analyse and describe a situation. For example, a sports physiologist who measures a person’s heart rate and uses a formula to describe this person’s physical condition faces various ways to consider variability. If she identifies an unusual heart rate, she should check if this is due to true physiological aspects, choice of measurement device, or measurement error.

Variability “is the reason why people have had to develop sophisticated statistical methods to filter out any messages in data [a signal] from the surrounding noise” (Wild & Pfannkuch, 1999, p. 236). A signal in data does not just refer to “true” values approximated with measure of central tendency, but can also describe stability in variability measured with a range, interquartile range or standard deviation. A signal – conceptualized here as stability in variability – can also be the shape of a distribution (Bakker, 2004) or a trend (e.g., Fitzallen, 2012). Variability is a thus a broader concept than spread or variation (Shaughnessy, 2007) and statistics therefore is not merely the science of variability (MacGillivray, 2004), but also the science of identifying and modelling stability or signals in the noise (Konold & Pollatsek, 2002).

In many statistics curricula there is a focus on identifying and measuring centres of data sets (Sorto, 2006) rather than on reasoning about variability. However, variability is an area of difficulty for students. As Reading and Shaughnessy (2004, p. 203) pointed out: “Students’ current lack of understanding of the nature of variability in data may be partly due to the lack of emphasis of variability in our traditional school mathematics curriculum and textbooks.” In a review of Dutch secondary school mathematics textbooks with statistics chapters, we found that none of them had a reference to variability or variation. Hjalmarson, Moore and delMas (2011) stress that the lack of tasks that require students to measure variability may impede their understanding of variability. They found a few examples of tasks in engineering textbooks that provoked students to measure variability, but these tasks were disconnected from a real-world context.

In our research we have designed tasks to engage students in considering variability in rich ways. As we will explain later, we assumed that measurement tasks based on authentic professional practices would assist in this aim. The goal of this article is to gain knowledge about *how students in Grade 12, aged 17-18, consider variability in these measurement tasks based on an authentic professional practice*.

2 Theoretical background

2.1 Variability

Variability is a multifaceted concept, like other statistical key concepts such as distribution (Bakker, 2004) and sampling (Pfannkuch, 2008). Hence Reid and Reading (2008) argue that the development of students' understanding of variability should be evaluated by thorough assessment of multiple facets of variability. This implies that it is important to engage students with rich activities that develop ways to *consider variability* when they engage in measurement activities. In their analysis of statisticians' thinking, Wild and Pfannkuch (1999) distinguished four ways to consider variability:

1. Noticing and acknowledging variability
2. Modelling or measuring variability for the purpose of predicting, explaining and controlling
3. Explaining and dealing with variability
4. Using investigative strategies to handle variability

Reading and Shaughnessy (2004) added "describing" and "representing variability" for educational use. We adapted the framework of Wild and Pfannkuch (1999) and incorporated Reading and Shaughnessy's (2004) additions (describing and representing) which we describe in Section 3.4.

To understand variability, one needs to consider both statistical and contextual worlds in relation to each other (Ben-Zvi & Aridor, in press). However, education typically focuses on only one of these components. Statistics education (as part of mathematics education) tends to focus on *techniques* that often neglect authentic application or cross-curricular content. Conversely, content areas such as science education may focus on scientific content, but pay little attention to statistical techniques (Estepa & Sanchez-Cobo, 1998; Reading & Shaughnessy, 2004). The outcome of this discrepancy is that students are inexperienced in applying statistical concepts to contextually rich content or authentic tasks in science. Makar and Confrey (2007) argue that students who engage in statistical inquiry with a compelling purpose, such as modelling experiences with authentic data, gain a deeper understanding of data analysis, the context itself and modelling. However, one challenge is to design learning activities that connect statistical techniques in authentic contexts that are rich in scientific content. This can help students acknowledge, notice, quantify and control statistical variability and seek meaningful explanations for variability in applications of science. In the next section we argue that measurement activities can establish such scaffolding.

2.2 Measurement

Measurement is the assignment of numbers with units to objects or events (Pedhazur & Pedhazur Schmelkin, 1991) and can be described as ordering our surrounding world through numbers to better control that world (Adams & Harrell, 2003; Buys & de Moor, 2005). It has been gaining greater importance in society, so to participate successfully in modern society, it is important that students learn to measure various phenomena in their environment and learn how to analyse the resulting measurements (Gooya, Khosroshahi, & Teppo, 2011; Lehrer & Kim, 2009).

An advantage of involving students in measurement activities is that it invites them to make connections between the real world and the world of data and thus learn to see that measurement cannot be absolutely accurate (Rabinovich, 2005). Measurement activities are suggested to be suitable for teaching about variability in interdisciplinary contexts (Enderson, 2003). These activities, with a discussion of the measurement process and the resulting data, can increase students' understanding of the nature and importance of measurement (Moore, 1990).

To understand measurement data and models, students need considerable contextual background, including knowledge of the phenomenon measured and the measurement procedure. Graphical representation of the data as part of the modelling process can help students to develop their understanding of variability (delMas & Liu, 2005). Such representations allow students to see shapes or trends in data, help them to make predictions and draw attention to variability. Research on graphing by Roth and Bowen (2003) suggests that students should be involved in measurement processes in order to interpret resulting data and models. They argue that even scientists require much contextual background to interpret graphs and if not familiar with the data generation process, they find it difficult to read graphs in their own discipline. In fact, Roth and Bowen recommend that experience with research and participation in graphing practices was more important for correct graph interpretation than exposure to increasingly complex graphs. We therefore chose to involve students in measurement activities that stress the importance of contextual background when graphing and modelling real data. Inspired by research in science education (Prins, 2010; Westbroek, Klaassen, Bulte, & Pilot, 2010), we chose to base the design of activities on authentic professional practices.

2.3 Authentic professional practices

In this article we define an authentic professional practice (AuPP) as a patterned purposeful activity of professionals working on a problem that is exemplary for their profession. We identify two desired elements of authentic practices to attend to in the design of activities, in particular, the work of a sports physiologist. First, in line with Prins, Bulte, van Driel, & Pilot (2008), we note that professionals need to have specific content knowledge that underpins their practice; for example a sports physiologist must have certain knowledge of biology. Second, the professional works according to statistical procedures accepted in the profession; for example, the sports physiologist may use regression techniques during the modelling of data obtained by a fitness test.

In science education, learning activities based on AuPPs can offer students meaningful contextual references to link abstract concepts (Lee & Butler-Songer, 2003). The activities based on AuPPs inevitably have to be to simplified or modified to make them useful in an educational setting. Dierdorp, Bakker, Eijkelhof, van Maanen (2011) based their design of statistics activities on an authentic practice of monitoring the height of dykes in the Netherlands, in which students used their contextual knowledge to make sense of variability in real data. However, it appears that variability needs more attention in design and teaching, which led to the study reported here.

2.4 Research question

Our article reports on research in which students in grade 12 (17-18 years old) reasoned about variability when engaged in a (simplified) practice of a sports physiologist. For example, to analyse aerobic and anaerobic respiration, students could measure heart rates under increasing power; then they could apply regression techniques to model variability in data in order to determine the ideal heart rate (threshold point) at which the working of muscles turns from aerobic to anaerobic metabolism. The idea underlying our research was that measurement activities based on suitable authentic practices could support students in considering variability in different ways. To evaluate this idea, we asked the following research question:

> *How do secondary students consider variability within measurement activities based on authentic professional practices?*

3 Method

3.1 Research setting

The data presented here stem from a four-year PhD research project investigating how students can learn about the statistical key concepts of correlation and regression in multidisciplinary contexts, experiencing the links between mathematics and the natural sciences. The overall study was based on design research, which involved an iterative design process (Barab & Squire, 2004; Van den Akker, Gravemeijer, McKenney, & Nieveen, 2006) consisting of six research cycles. Each cycle included the design of a hypothetical learning trajectory (Simon, 1995), a teaching experiment of about twenty lessons to implement and assess the instructional unit, analysis of classroom data and revision of the learning trajectory. This paper reports on the analysis of the *fifth research cycle*, which focused on *students' considerations of variability*.

As we were unable to find existing measurement activities in secondary school statistics based on authentic measurements by professionals, we designed an instructional unit ourselves. There was little education research to draw on because most of the research on measurement has been carried out in primary education, focusing on spatial measurement (Lehrer, 2003). Most research concerns relatively straightforward measurement of parameters such as length and volume with simple technologies such as rulers and measuring jugs (Smith III, Van den Heuvel-Panhuizen, & Teppo, 2011). What comes closest to what we envisioned is the work by Lehrer, Kim and Schauble (2007) in primary science education, which incorporates modelling and data analysis (see also English, 2009), topics related closely to variability and measurement.

In the design of the learning trajectory for this fifth teaching experiment research cycle, we searched for suitable AuPPs that contained measurement activities in which professionals use correlation and regression models that could be adapted for students in grades 11 and 12. We also wanted students to appreciate the AuPPs and identify with the professionals in ways that coherently embrace mathematics and the natural sciences. It was particularly important for us that there would be at least one AuPP-based activity in which students could perform a measurement experiment. These considerations led us to the practices of sports physiologists identifying the best training procedure for their clients. Based on typical practices of this profession, we designed two measurement activities over six lessons (Section 3.3).

3.2 Participants

Thirteen students, seven boys and six girls, from an affluent school took part in this study. They were in the beginning of Grade 12 of the pre-university track, which is attended by the top 15% of academically high achieving Dutch students. The first author (T) taught these students at his own school in a small city. Our designed instructional unit was entitled, *Statistics as a bridge between mathematics and the natural sciences* and was part of their school subject *Nature, Life and Technology* (Eijkelhof & Krüger, 2009). The students participated in classroom discussions and worked in pairs and small groups. They were expected to reason more than they were used to in other school subjects.

3.3 The measurement learning activities

In this section we describe the two measurement activities, which spanned three lessons each (50 minutes per lesson). These measurement activities aimed to involve students in *reasoning about variability* in informal ways in relation to correlation and regression to prepare for the learning of formal correlation and regression modelling techniques in subsequent lessons. In the first measurement activity the students had to perform heart rate measurements and use a given model to quantify *physical condition*. In the second measurement activity they were required to model their own measurements of heart rates under increasing power.

3.3.1 Measurement Activity 1 (MA1): Measuring Condition

MA1, which consists of six tasks with several subtasks, concerns the measurement of physical condition. Professional sports physiologists regularly use measurements and modelling techniques in their advice about the best training for their clients; in particular, accurate measurements and suitable statistical techniques are needed when they want to determine the physical condition of a person and assess their potential and risks. We assumed that students could engage easily with this context because many of them are concerned about their physical condition and do some sport themselves. They presumably have some prior knowledge of the AuPP and see the point of it. This would help them see the point of what they learn (cf. Lijnse & Klaassen, 2004).

We developed an analysis framework (Section 3.4) based on Wild and Pfannkuch (1999), with Reading and Shaughnessy's (2004) additions (describing and representing) incorporated. We added "describing" to noticing and acknowledging (consideration 1) and we consider representing to be an investigative strategy (consideration 4). Our framework therefore contains: 1) noticing, acknowledging and describing variability, 2) modelling or measuring variability, 3) explaining variability and 4) using investigative strategies to handle variability.

The design aim of MA1, on the measurement of physical condition, was to prompt reasoning about the ways in which students consider variability. We expected to achieve this goal by allowing students to perform their own measurements and compare these with an existing model. We assumed that suitable AuPPs constitute rich contexts that are meaningful for students, which would make it easier for them to consider variability (Cobb & Moore, 1997) and be motivated to learn (Dierdorp, Bakker, Eijkelhof, & van Maanen, 2013). We conjectured that a) the authentic data would show enough "noise" to urge students to notice and acknowledge variability when they interpreted the data (Konold & Pollatsek, 2002), b) they would understand that the relation between power (measured in Watts) and heart rate could be modelled with a trend line, c) they would explain the noise by sources of variability and d) they would use investigative strategies such as representing the data with graphs to seek ways to interpret the variability.

Task 2a

Introduction

In this task you will measure the heart rate frequency the same way as professional sports physiologists do. The task concerns three measurements. First a measurement at rest. Secondly, a measurement after knee bends. Third, a measurement at rest again. Work in small groups. Every student will take the test and perform the measuring at least once. In the following text you can read how to perform the measurements.

Heart Rate Measurement

The person who takes the physical test needs to sit quietly for about one minute before starting the measurement process. Measure the heart rate of each person at rest. Measuring the heart rate is determining the pulsation frequency. We call this resting heart rate **H1**. Always measure the heart rate with your middle finger (possibly joined by the index finger). The artery is on the side of your thumb.

Have someone else determine H1, so the person who gets tested doesn't need to keep an eye on the time.

Next, the testee does 30 deep knee bends in about 45 seconds. The back remains straight and the feet must keep contact with the ground. Each time your fingertips should touch the ground.

Measure the heart rate (number of pulsations) directly afterwards, for 15 seconds and convert it to rates per minute (**H2**). Sit quietly again. One minute later measure your heart rate again for 15 seconds and convert it to one minute (**H3**).

From these measured values an indication of your physical condition can be calculated using the Index Ruffier-Dickson formula. This index is frequently abbreviated as IRD and is defined as:

$$IRD = \frac{H2 - 70 + 2 \cdot (H3 - H1)}{10}$$

Translation from IRD to a qualitative indication of physical condition:

IRD	**Physical Condition**
Below or equal to 0	excellent
between 0 and 3	very good
between 3 and 6	good
between 6 and 8	moderate
above 8	bad

Figure 1. Task 2a from Measurement Activity MA1 (translated from Dutch).

Heart rate increases with increased power (physical effort), but this happens less rapidly with people who are in good physical condition than with people who are less physically fit. In addition, people who train regularly recover more rapidly after physical effort (heart rate becomes normal again). Researchers have designed suitable tests to quantify physical condition by measuring heart rates. The Ruffier-Dickson test (Paulet, Gratas, Dassonville, & Rochcongar, 1981), for example, uses heart rate frequencies at three relevant moments in a physical exercise to determine physical condition. In MA1, partly presented in Figure 1, students were asked to use and discuss this test. We expected them to demonstrate several ways to consider variability. After the completion of the heart rate measurements, students were asked to calculate their Index Ruffier-Dickson (IRD) and to reason about variability.

3.3.2 Measurement Activity 2 (MA2): Identifying a Suitable Sports Program

MA2, which consists of four tasks with several subtasks, concentrated on presenting and analysing data collected by students performing the Conconi Test (Conconi, Ferrari, Ziglio, Droghetti, & Codeca, 1987), which measures the threshold heart rate frequency (HRF) at which the muscles switch from aerobic to anaerobic combustion. Despite recent studies that have shown the Conconi test has limited levels of accuracy, we decided that it is a good option to offer students because it is still used and is suitable for students to reason about variability in relation to correlation and regression.

The design aim for MA2 was for students to demonstrate all four ways to consider variability: 1) noticing, acknowledging and describing variability, 2) modelling or measuring variability, 3) explaining variability and 4) using investigative strategies to handle variability. As argued in Section 2.2, it would be important for them to collect and model data themselves in order to enrich their ability to interpret the resulting graphs (Roth & Bowen, 2003).

MA2 is based on a practice of sports physiologists who identify the best training program for clients. Just like athletes, students had to measure their heart rate frequency (HRF) while increasing the intensity of their effort (power used at the treadmill). It is important for athletes to stay within the aerobic area, otherwise the muscles produce lactic acid. Training in the aerobic area under the threshold HRF from aerobic to anaerobic will prevent a decrease of this threshold and prevent muscle problems. According to Gellish, Goslin, Olson, McDonald and Moudgil (2007), when people increase their efforts during their training session, HRF increases proportionally with the power. If the power exceeds a certain point, the linear proportionality will disappear and the HRF will approach the peak heart rate (Figure 2). The upper bound where the HRF still behaves as a linear function of the power is called the point of deflection (Gellish et al., 2007). This is a good approximation for the threshold HRF which sports physiologists use to predict to plan the best training program. Measuring the threshold HRF takes place indirectly by analysing a graphical model of the data.

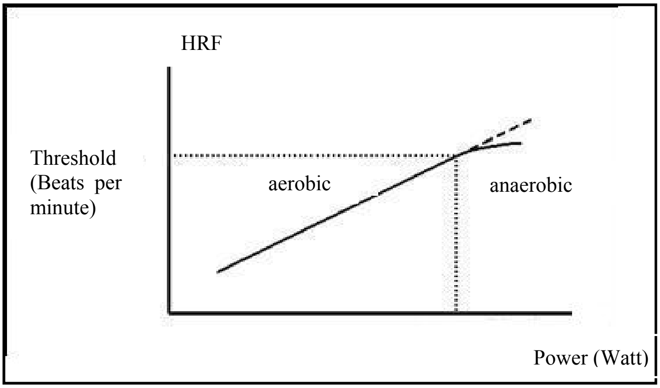

Figure 2. Model to find the threshold (based on Conconi et al., 1987).

During these measurement activities, students have to collect and reason about variability, make tables or graphical models of the data by scatter plots, which have not been addressed at school before. In contrast to MA1, the students had to model the measurement data generated in MA2 themselves using software. For this modelling, it was important that students reason about variability by considering the particular way that their group performed the test (method), the measurement devices they used to collect the data (there were several different devices and some measured by hand) and measurement errors. We also expected that students would use investigative strategies for finding a trend in their data.

3.4 Data collection and analysis

To assess students' reasoning about variability, their work in class was observed and recorded (audio and video) by the first author. Video-recorded lessons (six lessons of sixty minutes each) were transcribed. Data included students' written work, transcripts from video recorded lessons and field notes. One girl was not included in the analyses because of her absence during some lessons.

Using the adapted version of Wild and Pfannkuch's (1999) framework (extended by Reading and Shaughnessy, 2004), we analysed students' spoken interaction and studied their reasoning in depth, using the four ways to consider variability. To identify which considerations were at stake in the interactions between students or between student(s) and their teacher, we developed an analysis framework (Table 1).

Table 1
Analysis framework to identify which ways to consider variability.

Code	Consideration of variability	Example
NAD	*Noticing, acknowledging and describing variability*	The student implicitly refers to variability or explicitly describes variability.
Mod	*Modelling or measuring variability*	The student discusses variability in relation to the model (e.g., between measurements of heart rate and level of power or between age and maximum heart rate).
Exp	*Explaining variability*	The student tries to explain variability in data by indicating that people are different, or that the circumstances are not equal.
Inv	*Using investigative strategies to handle variability*	The student discusses what is necessary to describe the variability (investigative strategy) or how to handle variability. For example, she represents such variability in a graph or table to arrive at a model or identify a trend, or call conditions on which the strategies can be used.

We divided the transcripts of the classroom interaction into utterances (our unit of analysis) in which the researcher recognized one of the four ways to consider variability. This process yielded 82 analysis units (utterances).

This categorization (Table 1) is neither a hierarchy nor a list of exclusive categories. The first category in our analysis framework (NAD) is conditional because acknowledging variability is prerequisite for the other three ways of consideration. Because we wanted unique codes for each utterance, we used code NAD only if no other code applied. The utterances were first categorized by the first author and an independent researcher who was not involved in this study, but an expert in mathematics education and psychology who made coding decisions without access to the decisions of the first author. Kappa was .66, which Cohen (1960) considers substantial.

4 Results

In each measurement activity (MA1 and MA2), we investigated the spoken utterances of students to identify considerations they demonstrated in their reasoning when they engaged in measurement activities based on AuPPs. Table 2 demonstrates that the measurement activities based on authentic practices stimulated students to reason about all four ways to consider variability. Table 2 contains some zeros because in cases where the teacher explicitly asked about one of the four ways to consider variability the reactions of the students were not included in the analysis, because our focus was to investigate whether students themselves would reason about the several ways to consider variability. In the first set of tasks (MA1) 37 (4 + 5 + 16 + 12) utterances were coded and in the second set of tasks (MA2) 45 (7 + 5 + 18 + 15). Table 2 shows that nearly all students noticed variability. Furthermore, the table indicates that most students attempted to explain or work with variability, but few tried to explicitly measure or model (quantify) it.

Table 2
Numbers of spoken student utterances that demonstrate reasoning about a way to consider variability.

	MA1				MA2			
Student	NAD	Mod	Exp	Inv	NAD	Mod	Exp	Inv
Tom	0	1	1	4	0	0	0	1
Alan	0	0	3	1	5	2	4	5
Bert	0	2	3	2	0	0	0	0
Jorr	2	1	2	1	0	1	4	5
Rose	2	0	2	0	1	0	2	1
Maxima	0	0	1	0	0	0	0	2
Elsa	0	0	1	0	0	1	4	0
John	0	0	2	0	0	0	0	0
Simona	0	0	1	1	0	0	0	0
Abel	0	1	0	2	1	1	2	1
Eleonore	0	0	0	0	0	0	1	0
Kai	0	0	0	1	0	0	1	0
Total	4	5	16	12	7	5	18	15
#students	2	4	9	7	3	3	6	6

To give a qualitative illustration of students taking into account the four ways to consider variability we briefly report on their reasoning during MA1. To set the stage, we first sketch how the first measurement activity was introduced. We wanted to involve students in the measurement activities and to become aware of the presence of variability around the regression line. To achieve this, the teacher introduced the following task at the beginning of MA1: "Consider how a sports physiologist could support a client in improving her condition and why it can be useful to measure her heart rate for that." Most students wrote down that the HRF depends on the degree of power. None of the students mentioned anything about variability in their written text. In the next sections we provide examples of how MA1 helped stimulate students to reason about each of the four ways to consider variability.

4.1 Noticing, acknowledging and describing variability

In this subsection we illustrate that when students did not notice variability, very little support from the teacher was needed to stimulate them to reason about variability. Part of MA1 asked students to orally explain the difference between the measured IRD and the condition of their class mates. Although, in their written work, students mentioned several ways to consider variability, they did not mention aspects of variability orally. The teacher responded to this in the next lesson by again leading the discussion about variability. In the next excerpt, Jorr and Tom acknowledged variability in an informal way:

T: Do you expect the same values [IRD] when you run the experiment three times?
Tom: If in between the experiments you recover completely because you have time enough to take a rest, then the values [IRD] will be the same.
[a few minutes later:]
T: Some [students] say if you rest enough, you will find the same [IRD].
Jorr: Well, not quite exactly the same.
T: So, you expect something close to it?
Jorr: Yes, a small deviation. You have to take the mean.
Tom: That is true.

The observations showed that all students agreed and the teacher was satisfied that the students were aware of some variability in their measurements. The analysis indicated that students knew that measurement values need not be exactly the same, but sometimes did not seem to see the need to express this variability. By evoking a cognitive conflict (Watson, 2007), the teacher had an important role in stimulating them to express more precisely the difference between what they said and what they observed. First, Tom expected the same values, later he agreed with Jorr that he expected a small deviation. Therefore, we coded these statements as Tom and Jorr acknowledging variability and trying to describe it informally.

4.2 Modelling or measuring variability

We conjectured that students would understand that a model (in this case heart rate versus power) is a representation of a relationship and that there would be some variability around any model. We give an example in which Tom considered variability when he described the relation between heart rate and a person's physiology:

Tom: Sports physiologists have devices to measure a person's physiology. When you can consider the heart rate as an indicator, it should not be too high or too low. But it is only a partial indicator to measure the effort to do something.
T: So, you say: it is an indicator. For what?
Tom: It is an indicator for the degree of effort you must perform to do something.

Tom reasoned how to measure physiology and mentioned "partial indicator" to refer to the model. He seemed to realize that the heart rate gives indications about a person's physical condition and that fluctuations of heart rates are to be expected. The teacher recognized the importance of Tom's remark because he wanted to make sure that students were aware that a characteristic such as heart rate could be used by a sports physiologist as a (simplified) measure to assess a client's physical condition.

MA1 confronted the students with variability around the (IRD) model. They had to measure heart rates and find a value by the IRD model to predict the physical condition of a "client." We saw some students struggle with the relation between their data and the model. For example, when the teacher showed the scatter plot of the students' measurements and asked if the students recognize a relation between the resting heart rate (H1; Figure 1) and the heart rate after the knee bending (H2), Bert responded with: "There are very few points." In explaining the variability he seemed to realize that a small sample size would make it more difficult to formulate the relation and that a bigger sample size would be preferable when constructing a meaningful model. However, other students could recognize a trend:

Abel: Actually, I can see a line. But only a little.
Kai: No.
Abel: Yes, I do. I see no line, but I recognize a rising regression line. You have to make this data set real.
T: But this is real data [the students collected it themselves].
Abel: But I cannot see the shape of the [regression] line.
T: So, you cannot say something about the relation, but can you say something in common?
Abel: They [H1 and H2] are linearly proportional.
T: But, if you know the formula for the right regression line, what can you do with it?
Abel: Make a prediction.
Jorr: If your condition is above or below the mean.

Abel likely meant that there was a lot of variability in the data, but that he could recognize a pattern.

4.3 Explaining variability

For the third way to consider variability which we expected to see in students' reasoning, we conjectured that students would try to explain variability by indicating that people are different or that the circumstances were not equal. An example is given by Rose who found a higher IRD than she expected: "For me it [IRD value] is not right. I exercise four times a week and swim and cycle every day." When the teacher asked her if she could explain this result, she said that she had heart problems and that her heart might not work in a way that is necessary to measure condition by heart rates.

We think that modelling their own measurements evoked students' reasoning about variability. We observed that all students considered whether their measurement results fit the model and tried to explain any deviations from the model. Although all students considered themselves fit, only three of them found a result which corresponded with their perception of their actual condition. The fact that they were dissatisfied with the outcome implies that the students linked the measurement to the authentic context. Allen, one of the disappointed students, even felt a little offended by the result and used the context to explain his disappointment: "I have got 6.8 and am supposed not to have a good condition [according to the IRD]. This surprises me because I train 5 times a week." Other students made similar comments, so the teacher asked what could be explanations for the disappointing results. The students mentioned that the measurements were inaccurate.

The authentic character of the measurement activities urged some students to seek an explanation for why data varied without being too concerned about actually describing the variability, because they had some contextual knowledge (Reading & Shaughnessy, 2004). The students mentioned several ways to explain the variability. Jorr mentioned the "accuracy of the measurement." In terms of Wild and Pfannkuch (1999), we could say that he meant that there are "accidental" mistakes in the data collection process. Students regularly referred to the difference between people to explain the variability (Schwartz, Goldman, Vye, Barron, & Cognition and Technology Group at Vanderbilt, 1998). For example, Bert referred to a famous athlete:

> Kai: Maybe stress increases your heart rate and relaxation by sports decreases your heart rate.
>
> Bert: If Louis Armstrong digs his front yard his heartbeat is the same as when an ordinary human is asleep.

Bert probably meant cyclist Lance Armstrong, but his message was clear: People are not the same, or the circumstances are not the same. Students often mentioned: "A human is not a machine."

We suggest that indicating that people are different, or that the circumstances are not equal are students' implicit ways to explain and thus deal with variability. Another indication that students deal with variability is found when the teacher asked the students the relevance for a sports physiologist to know the data of the whole group. Six of the students mentioned something like Kai: "Then he knows if the client is below or above the mean", which was related to variability; only two of them mentioned variability explicitly.

4.4 Using investigative strategies to handle variability

We used this fourth code when the students discussed what is necessary to describe the variability (investigative strategy) or how to do this. For example, when the students tried to explain why clients' heart rates did not fit the formula exactly, the authentic character of the task seemed to encourage the students to be more aware of variability and to link their data to the scientific context:

Bert: Maxima, the formula implicated that you have a bad physical condition. Do you smoke?
Maxima: No, I am just not very sporty.
Jorr: Whether someone smokes cannot be found in the formula, but it [smoking] can have an impact on the value [IRD].

Bert was apparently searching for a contextual explanation why Maxima had an extremely poor number to indicate her condition. We think that Bert's remark suggests that the authentic character of the measurement activity made him aware of real variability and see the model as indicative for physical condition. We see this as an example to investigate a need to know more information in order to describe or explain the variability. However, one could also interpret Bert's statement with code EXP (explaining that people are all different). We now discuss an example when students represented variability by a scatter plot and noting what was needed to find the regression line. When the teacher asked them what they meant by a rising regression line, students' mathematical knowledge helped them to reason about the variability of their measurement results:

Bert: We have very few points to draw a regression line.
T: That is true.
Abel: I can recognize a [regression] line.
T: How?

Bert mentioned that there were only a few data points to produce a regression line. He seemed to be aware that more data points makes it easier or perhaps more reliable to find the regression line. When the teacher asked how Abel recognized a regression line, Kai answered first:

Kai: The regression line is the mean of all data points.
Abel: The mean of all data points through the data points.
Kai: With the same number of points above the line as under the line.
Abel: It is necessary to have the same number of points above the line so that the overall result of deviations on the upper side is as large as the overall result at the bottom.

We think that Abel tried to formulate a version of the "sum of residuals" when mentioning overall results of deviations. In this context, the sum of residuals is the summation of all absolute deviations of the heart rate observations from the regression line. Measuring the data themselves and representing them by a graph seemed to encourage them to consider the deviations of their measurements from the IRD model even though they had not learned this idea yet. These formal techniques would be learned later in the course (after MA1 and MA2).

There are several differences between MA1 and MA2. In MA1 students were given a model (the IRD formula); in MA2 they had to model their measurements and they were informed that the point of deflection in a scatter plot indicates the threshold heart rate. IRD is a simple indicator of physical condition, based on data from many people. MA2 was more explicitly linked to the biochemical process of metabolic acidosis. This meant that in MA1, students had to reason about variability with regard to their individual values in relation to an aggregate data set, whereas they could remain focused on an individual's data in MA2 when doing their running test (Conconi et al., 1987). In this running test they gathered data by measuring the heart rates with increasing speed of the treadmill.

In the following, we focus on students' reasoning about patterns of variability (investigative strategy) when they were involved in MA2. Most students were able to find their own threshold point, but some students did not recognize a trend in their data just as when they struggled with comparing the Ruffier-Dickson model with their own variable data. Given their limited experience with statistics and the literature on students' difficulties of coordinating local and global perspectives on data sets (Ben-Zvi & Arcavi, 2001), this should not be too surprising. What might have helped here is that students seemed to have an idea of what the underlying scientific phenomenon was they were measuring—as in the approach of Lehrer, Kim and Schauble (2007) in which students had a sense of the true value they were approximating. The teacher mentioned that the part of their graphs before the threshold point were not totally linear. Alan responded and mentioned the variability implicitly:

Alan: It is not the fact that our heart rate is not linear, but the line is based on something we want to be linear. In my head it is correct. It is not that the heart rate is linear, but because we constructed a linear line as a kind of "guideline" [Mod].

T: So you say that based on this theory, these models, there must be a linear relation because we think that it is linear?

Alan: Yes, we invented a linear relation with values which are not linear.

T: You say: there is no linear relation?

Alan: No, it is not completely linear. It is almost linear.

In fact, Alan claims that the linearity of the model is correct, but that the actual data do not fit the model completely. He stressed the presence of variability and was not the only student who did this. Later, when they constructed a representation of their collected data in a scatter plot they noticed variability and named it “a margin.”

Jorr: You can see this as a margin.
T: What does this margin say to you?
Jorr: The possible deviation for people who score poorly and those who score better [Exp].

Jorr used the phrase “margin” and Elsa agreed with him. She suggested not sticking rigidly to a formula when advising a client, but to deal with the variability and use a margin. In the next excerpt the students try to explain the variability:

Elsa: It has an aspect of randomness. Like a thermostat. You got a standard, but the value can be below or above. There is a margin of error.
T: Error?
Elsa: A range of errors. When the value can be found between two limits and between these limits it is random.
T: Why errors?
Elsa: I interpret them as errors. Like a standard. It can be above or below. It fluctuates between them. The processes in your body are never the same.
Kai: Your body is not a machine [Exp].
T: Your body is not a machine? What do you mean with that?
Kai: I mean, your body is not always working as described by the model.
Alan: Suppose that during training you see a beautiful woman, then your heart rate will become higher.
Elsa: The treadmill is rather long. So, you can make big steps or small steps. Or walk a little faster and slow down a little [Exp].
T: Thus, we have also variability caused by methodological aspects?
Students: Yes.

Again the students explained the real variability (e.g., your body is not a machine, you can make big steps or small steps). In this phase the teacher tried to let the students reason about explanations for variability. The body not being a machine was mentioned multiple times; the students agreed that even if the test were done in a laboratory, the result would still vary.

5 Discussion

As summarised in this article, variability is a key concept in statistics that typically does not receive the attention it deserves in statistics education. As part of a larger research project we have designed measurement tasks that were based on authentic professional practices in which correlation and regression are used. The advantage of measurement is that it is at the interface between context and statistics, where students can get a feel for where variability comes from (e.g., variability in the phenomenon studied versus measurement error). The assumed advantage of basing measurement tasks on an authentic professional practice is that students may then better see the need of learning about statistical modelling techniques such as correlation and regression and thus be motivated to learn about them. Moreover, it is known that students often demonstrate computational habits without realistic considerations when they solve word problems (Cooper & Harries, 2002). More recent studies indicate that more authentic tasks can help to counteract such habits (Verschaffel, Greer, Van Dooren, & Mukhopadhyay, 2009).

To test whether our measurement tasks did help students to consider variability in rich ways, this article addressed the following research question: *How do secondary students consider variability within measurement activities based on an authentic professional practice?* We used an analysis framework based on ways in which statisticians consider variability (Table 1): noticing and acknowledging variability, measuring and modelling variability, explaining variability and using investigative strategies. With this framework we analysed how students considered variability and we found that students demonstrated all these ways of considering variability (though of course in less advanced ways than would statisticians). This suggests that the measurement tasks based on the authentic professional practice of a sports physiologist have the potential to stimulate students to consider variability in all these ways, provided that the teacher helped them to deepen their reasoning about variability.

In more detail, analysis of the transcripts and observations showed that students noticed and acknowledged variability. They experienced that the data they found did not exactly fit the model and they tried to find explanations for this. Furthermore, they were concerned with a margin around the model to predict a client's condition and explained the deviation of the value of condition according to the value obtained by the model they found. To control the variability in data the students suggested doing more measurements, using the same device for every measurement. To identify a suitable sports program the students used investigative strategies by modelling the data to find the threshold point. Finally, we conclude that the two activities supported students to reason about the four ways to consider variability as described in our coding scheme.

Because of the deviations from their own data and the data predicted by the model, students were asking themselves whether the heart rate was the key characteristic that is needed to measure physical condition. They decided there were more variables to consider, but that the heart rate can be indicative for this context about physical condition. The “noise” in their data urged them to consider the sources of variability. Also, the method of measuring was discussed, because some students noticed that other students did not apply the same methods. Some students tried to control the variability by repeated measurements, but again noticed more variability.

5.1 Measurement

As the analysis suggests, performing measurements seemed to stimulate students to consider variability in different relevant ways. We suggest that the students gained an understanding of “sensible” measurements by using measurement and modelling activities to find patterns, such as a trend. They found trends by representing, analysing and generalising their collected data in table formats as well as in figures. Their prior contextual knowledge helped them to acknowledge and deal with variability (e.g., “your body is not a machine”), but the teacher’s support was often needed to elicit students’ reasoning about variability.

5.2 Authentic Practices

We think that the authentic character of the activities supported students in modelling their measurement results. The transcripts suggest that students were not just solving a word problem, but considered variability to find an answer for a “real” problem. For most students, these activities based on authentic practices were successful in reaching the goal of supporting their reasoning about the ways to consider variability. Some students struggled with explaining variability in their measurement results, but our study suggests that the students gained awareness that you could use the model for sensible predictions, but that it does not precisely describe reality. The measurement experiences of the students, together with the class discussion, contributed to the students’ view that a model is a simplification of reality. For some students, this was difficult to see initially and some struggled with the variability of the data. However, the results of this study suggest that the teaching and learning activities generally supported students in learning to measure parameters of physical condition using modelling techniques and to reason about variability in valuable ways.

5.3 Limitation and future research

As a limitation of our study, we note that only one small group of students was involved. We thus see this study as a proof of principle that it is promising to base tasks in statistics education on authentic professional practices in which statistical techniques are used. We think that the measurement activities can be extended to support students in understanding other more sophisticated types of variability as well, such as sampling variability. In follow-up research (AUTHORS, 2013), we found that basing teaching and learning strategies on authentic professional practices may also help students to be motivated to learn and see the point of their learning and to help to develop rich multifaceted concepts. Also, we suggest that our strategy, based on authentic professional practices in which statistics is used, can help students to make connections between school subjects such as mathematics and biology.

References

Adams, T. L., & Harrell, G. (2003). Estimation at work. In D. H. Clements & G. Bright (Eds.), *Learning and teaching measurement* (pp. 229–244). Reston, VA: National Council of Teachers of Mathematics.

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Utrecht, the Netherlands: CD Bèta Press.

Barab, S. A., & Squire, K. D. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences, 13*(1), 1–14.

Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics, 45*, 35–65.

Ben-Zvi, D., & Aridor, K. (in press). Children's wonder how to wander between data and context. In D. Ben-Zvi & K. Makar, (Eds.), *Teaching and learning of statistics: International perspectives*. University of Haifa, Israel: Statistics Education Center.

Buys, K., & de Moor, E. (2005). Domain description measurement. In M. van den Heuvel-Panhuizen & K. Buys (Eds.), *Young children learn measurement and geometry* (pp. 15-36). Utrecht, the Netherlands: Freudenthal Institute (FIsme).

Cobb, G. W., & Moore, D. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly, 104,* 801–823.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46.

Conconi, F., Ferrari, M., Ziglio, G., Droghetti, P., & Codeca, L. (1987). Determination of the anaerobic threshold by a noninvasive field test in runners. *Journal of Applied Physiology, 52*, 869–873.

Cooper, B., & Harries, T. (2002). Children's responses to contrasting "realistic" mathematics problems: Just how realistic are children ready to be? *Educational Studies in Mathematics, 49*, 1–23.

delMas, R. C., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal*, *4*(1), 55–82.

Dierdorp, A., Bakker, A., Eijkelhof, H. M. C., & Van Maanen, J. A. (2011). Authentic practices as contexts for learning to draw inferences beyond correlated data. *Mathematical Thinking and Learning, 13*(1-2), 132–151.

Dierdorp, A., Bakker, A., Eijkelhof, H. M. C., & van Maanen, J. A. (2013). Betekenisvolle statistiek in beroepspraktijken als brug tussen wiskunde en natuurwetenschappen: evaluatie van een ontwerponderzoek. [Meaningful statistics in professional practices as a bridge between mathematics and science: An evaluation of a design research project.] *Pedagogische Studiën, 90*, 4–18.

Eijkelhof, H. M. C., & Krüger, J. (2009). *Improving the quality of innovative science teaching materials.* Paper presented at European Science Education Research Association conference (ESERA 2009).

Enderson, M. (2003). Using measurement to develop mathematical reasoning at the middle and high school levels. In D. H. Clements & G. Bright (Eds.), *Learning and teaching measurement* (pp. 271–281). Reston, VA: National Council of Teachers of Mathematics (NCTM).

English, L. D. (2009). Promoting interdisciplinarity through mathematical modeling. *ZDM The International Journal on Mathematics Education, 41*, 161–181.

Estepa, A., & Sanchez-Cobo, F.T. (1998, June). Correlation and regression in secondary school textbooks. In L. Pereira Mendoza, L. Seu, T. Wee, & W.K. Wong (Eds.), *Proceedings of the Fifth International Conference on the Teaching of Statistics* (vol. 2, pp. 671–676). Voorburg, the Netherlands: International Statistical Institute.

Fitzallen, N. (2012). Interpreting graphs: Students developing an understanding of covariation. In J. Dindyal, L. P. Cheng & S. F. Ng (Eds.), *Mathematics education: Expanding horizons* (Proceedings of the 35th annual conference of the Mathematics Education Research Group of Australasia). Singapore: MERGA.

Gellish, R. L., Goslin, B. R., Olson, R. E., McDonald, A., Russi, G. D., & Moudgil, G. V. (2007). Longitudinal modeling of the relationship between age and maximal heart rate. *The American College of Sports Medicine, 39*(5), 822–829.

Gooya, Z., Khosroshahi, L. G., & Teppo, A. R. (2011). Iranian students' measurement estimation performance involving linear and area attributes of real-world objects. *ZDM The International Journal on Mathematics Education*, *43*(5), 709–722.

Hjalmarson, M. A., Moore, T. J., & delMas, R. (2011). Statistical analysis when the data is an image: Eliciting student thinking about sampling and variability. *Statistics Education Research Journal, 10*(1), 15–35.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, *33*(4), 259–289.

Lee, H. S., & Butler-Songer, N. (2003). Making authentic science accessible to students. *International Journal of Science Education, 25*(1), 1–26.

Lehrer, R. (2003). Developing understanding of measurement. In J. Kilpatrick, W. G. Martin & D. E. Schifter (Eds.), *A Research Companion to Principles and Standards for School Mathematics* (pp. 179–192). Reston, VA: National Council of Teachers of Mathematics.

Lehrer, R., & Kim, M. J. (2009). Structuring variability by negotiating its measure. *Mathematics Education Research Journal, 21*(2), 116–133.

Lehrer, R., Kim, M., & Schauble, L. (2007). Supporting the development of conceptions of statistics by engaging students in modeling and measuring variability. *International Journal of Computers for Mathematical Learning, 12*(3), 195–216.

Lijnse, P. L., & Klaassen, K. (2004). Didactical structures as an outcome of research on teaching-learning sequences? *International Journal of Science Education, 26,* 534–554.

MacGillivray, H. (2004). Coherent and purposeful development in statistics across the education spectrum. In G. Burrill & M. Camden (Eds.), *Curricular Development in Statistics Education: International Association for Statistical Education 2004 Roundtable*. Voorburg, the Netherlands: International Statistical Institute.

Makar, K., & Confrey, J. (2007). Moving the context of modeling to the forefront: Preservice teachers' investigations of equity in testing. In W. Blum, P. Galbraith, H-W. Henn & M. Niss (Eds.), *Modelling and applications in mathematics education* (pp. 485–490). New York: Springer.

Moore, D. S. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). Washington, DC: National Academy Press.

Paulet, G., Gratas, A., Dassonville, J., & Rochcongar, P. (1981). Comparative study of four effort tests on nine-year-old children. *European Journal of Applied Physiology, 46*(1), 55–68.

Pedhazur, E. J., & Pedhazur Schmelkin, L. (1991). *Measurement, design, and analysis: An integrated approach* (1st ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pfannkuch, M. (2008, July). *Building sampling concepts for statistical inference: A case study.* Paper presented at the International Congress on Mathematical Education, Monterrey, Mexico.

Prins, G. T. (2010). *Teaching and learning of modelling in chemistry education. Authentic practices as contexts for learning.* Utrecht, the Netherlands: CD Bèta Press.

Prins, G. T., Bulte, A. M. W., Van Driel, J. H., & Pilot, A. (2008). Selection of authentic professional practices for chemistry education. *International Journal of Science Education, 30*(14), 1867–1890.

Rabinovich, S. G. (2005). *Measurement errors and uncertainties – theory and practice.* New York: Springer.

Reid, J., & Reading, C. (2008). Measuring the development of students' consideration of variation. *Statistics Education Research Journal, 7*(1), 40–59.

Reading, C., & Shaughnessy, J. M. (2004). Reasoning about variation. In D. Ben-Zvi and J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 201–226). Dordrecht, the Netherlands: Kluwer.

Roth, W. M., & Bowen, G. M. (2003). When are graphs ten thousand words worth? An expert/expert study. *Cognition and Instruction, 21*, 429–473.

Schwartz, D. L., Goldman, S. R., Vye, N. J., Barron, B. J., & Cognition and Technology Group at Vanderbilt (1998). Aligning everyday and mathematical reasoning: The case of sampling assumptions. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K-12* (1st ed., pp. 233–273). Mahwah, NJ: Lawrence Erlbaum Associates.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F.K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 2. pp. 957–1010). Charlotte, NC: Information Age.

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

Smith III, J. P., van den Heuvel-Panhuizen, M., & Teppo, A. R. (Eds.) (2011). Learning, teaching, and using measurement: Introduction to the issue. *ZDM The International Journal on Mathematics Education, 43*(5), 617–610.

Sorto, M. A. (2006, July). Identifying content knowledge for teaching statistics. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil [CDROM]. Voorburg, the Netherlands: International Statistical Institute.

Van den Akker, J. J. H., Gravemeijer, K., McKenney, S., & Nieveen, N. (2006). *Educational design research.* London: Routledge, Taylor & Francis.

Verschaffel, L., Greer, B., Van Dooren, W., & Mukhopadhyay, S. (Eds.). (2009). *Words and worlds: Modelling verbal descriptions of situations.* Rotterdam/Boston/Taipei: Sense Publishers.

Watson, J. M. (2007). The role of cognitive conflict in developing students' understanding of average. *Educational Studies in Mathematics, 65,* 21–47.

Westbroek, H. B., Klaassen, C. W. J. M., Bulte, A., & Pilot, A. (2010). Providing students with a sense of purpose by adapting a professional practice. *International Journal of Science Education, 32*, 603–627.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*, 223–265.

Chapter 4

SUPPORTING STUDENTS' CONCEPTUAL DEVELOPMENT FOR SHUTTLING BETWEEN CONTEXT AND STATISTICS: THE CASE OF SAMPLING

Abstract

This article focuses on how to support students (aged 16-18) to develop statistical concepts in such a way that they can shuttle between contextual and statistical spheres. Shuttling involves mathematization and linking results of statistical modelling or computation back to the contextual problem. To realize this shuttling in education, designers and teachers need to help students see the purpose of tasks and the utility of concepts in the shuttling process. To this end we based an instructional unit on authentic professional practices. Two case studies were carried out to assess the potential of two tasks from this unit to support students' concepts that underlie sampling – sample size, random process, distribution, confidence interval, and relationship between sample and population – and to find indications of what teachers should do to use this potential. Analysis of video recordings indicates students' coherent development of these concepts, suggesting the tasks have the potential for which they were designed. The students seemed aware of the purposes of the tasks and could apply their statistical knowledge, but tended to forget to shuttle back from the statistical to the contextual sphere. This suggests that teachers need to help students to link back to the contextual problem even if tasks are based on authentic professional problems and students see the task's purpose.

Keywords: *statistical reasoning, sampling, inferential reasoning, authentic practice, upper secondary education, purpose & utility*

1 Conceptual development for shuttling between contextual and statistical spheres

This paper deals with a fundamental challenge in design for mathematics and statistics education: to stimulate students to develop concepts in such a way that they can shuttle back and forth between a contextually phrased problem and mathematics or statistics. Many contextual problems can be solved with the help of mathematics or statistics. This requires a process of mathematization, from the "world of life" where people live and act to the "world of symbols" where reality is symbolized and symbols are manipulated (Freudenthal, 1991, pp. 41-42). But the results of mathematization (e.g., modelling) also need to be evaluated on their merits in the context. The literature on modelling (e.g., Galbraith & Stillman, 2006) and statistical thinking emphasizes the importance of such "shuttling between the contextual and statistical spheres" (Wild & Pfannkuch, 1999, p. 28) as a learning goal for mathematics and statistics education.

From the educational literature, however, we know that such shuttling is not easily promoted in students. Research into school statistics reports the problems that students have in drawing inferences which are contextually meaningful (Makar & Rubin, 2009). Transfer of school mathematics to out-of-school situations is known to be difficult (e.g., Nunes, Schliemann, & Carraher, 1993) and the use of realistic contexts does not necessarily have the desired effect (Boaler, 1993). Solution-oriented solving of story problems can lead to poor conceptual understanding of the domain (Hung & Jonassen, 2006). The substantive literature body on word problems (Palm, 2008; Verschaffel, de Corte, & Lasure, 1994) further shows that students often take a purely calculational approach and fall prey to suspension of sense-making when solving contextually phrased word problems. Ainley, Pratt and Hansen (2006, p. 25) even consider "the provision of authentic tasks inherently problematic." They propose to focus task design on purpose and utility rather than authenticity. A purposeful task is one that has a meaningful outcome for the student, and utility is the construction of meaning for the ways in which mathematical concepts are useful. This focus on purpose and utility is a way out of what they call a "planning paradox" in task design:

> … if teachers plan from tightly focused learning objectives, the tasks they set are likely to be unrewarding for the pupils, and mathematically impoverished. If teaching is planned around engaging tasks the pupils'

> activity may be far richer, but it is likely to be less focused and learning may be difficult to assess (2006, p. 24).

We propose that a focus on purpose and utility can go well together with authenticity. Research in science education has shown that a design approach in which tasks are based on authentic professional practices can help students see the purpose of what they do in classrooms and the utility of what they learn (Westbroek et al., 2010). However, there are potential drawbacks as well. Westbroek et al. based their instructional sequence in chemistry education on the professional practice of testing drinking and swimming water. In simulating this practice, students saw the purposes of what they did and the utility of what they learned, but they learned very little about chemistry. Such an approach may thus come at the expense of conceptual learning. This suggests that the planning paradox is not necessarily resolved by focusing on purpose and utility. Design should also focus on conceptual development, and this – we argue – adds particular requirements both to tasks and teaching.

A narrow focus on tight conceptual learning objectives may lead to an atomistic approach in which specific aspects of concepts are addressed in isolation (Bakker & Derry, 2011). Reform attempts that engage students in richer projects aim to address concepts more holistically in relation to each other. In such holistic approaches it may be more challenging, as Ainley et al. (2006) point out with their planning paradox, to steer what students might learn and assess what they have learned. What we wanted to investigate in our research is whether it is feasible, in principle, to have the best of different worlds. The aim of this paper is to examine whether realistic tasks, inspired by authentic professional practices can support students' conceptual understanding in such a way that they can shuttle between contextual and statistical spheres.

2 Theoretical background

2.1 Relation between concepts and shuttling

Shuttling as characterized above is an informal overarching term covering several processes. The shuttling from context to statistics can take different forms such as mathematization, modelling, or the application of statistical knowledge. Shuttling back involves judgment of the statistical outcomes' value for the contextual problem at hand: Is this result meaningful, useful, valid? Is there a range for which the model holds? Under what conditions does it make sense to use this model or technique?

Such shuttling requires a particular type of conceptual understanding (as opposed to procedural knowledge involved in calculational approaches). As Brandom (2000, p. 48) stated in line with his inferentialist theory:

> To grasp or understand (. . .) a concept is to have practical mastery over the inferences it is involved in—to know, in the practical sense of being able to distinguish, what follows from the applicability of a concept, and what it follows from.

The consequence of this view is that the dichotomy between the contextual and conceptual dissipates. For example, if someone understands the concept of sampling (and other concepts relevant to the context), it means she knows how to draw inferences with this concept in concrete situations. She knows when the concept can be applied and what the consequences are of applying it. Understanding thus also includes knowing about the concept's utility. As a consequence, the shuttling between what is conveniently summarized as context and statistics in fact requires a deep understanding of several concepts, some of which are statistical (cf. Bakker & Derry, 2011). This means they fall under norms that are common in statistical practice. From an inferentialist perspective, students thus need to broaden the scope of using and understanding statistical concepts beyond the statistical domain. It is for this reason that we focus in the following on conceptual development in the broad sense, thus including judgment about contextual problems. In short: shuttling is the main learning goal and conceptual development is the means to this end.

2.2 Concepts underlying sampling

We focus on sampling – an important yet somewhat neglected area in statistics education – in relation to other statistical concepts such as variability, correlation, regression and distribution. Sampling is considered a key aspect to the teaching of informal inferential reasoning (e.g., Saldanha & Thompson, 2003). Pfannkuch (2008, p. 1) argued: "When students are not aware of sampling their informal inferential reasoning is limited." We should emphasize two related points. First, concepts should not be studied in isolation (Bakker & Derry, 2011). Brandom (2000, p. 15-16) observes that concepts come in packages: "one cannot have any concepts unless one has many concepts. For the content of each concept is articulated by its inferential relations to other concepts. Concepts, then, must come in packages." The sampling we study in this article is closely linked to correlation, regression, variability and distribution, but also to the contextual concern of saving money by avoiding unnecessary big samples. Second, concepts are often multifaceted. Based on Wild and Pfannkuch's (1999) framework for statistical thinking, Pfannkuch (2008, p. 4) noted that students' statistical reasoning about sampling involves the following five underlying concepts: sample size (in relation to the law of large numbers), random process, distribution, informal confidence interval, and relationship between sample and population. An informal confidence interval is "a sense of the reasonably expected variability around the expected value" (Shaughnessy, 2006, p. 87).

One pedagogical approach related to sample size that informed our tasks was that of growing samples (Konold, & Pollatsek, 2002; Bakker, 2004, 2007). In this approach students start with a small sized sample and increase the sample size step by step. In this way they gradually develop their reasoning about sampling in relation to other statistical concepts such as distribution (Bakker & Derry, 2011) and become able to infer from a sample about a population (Ben-Zvi, Aridor, Makar, & Bakker, 2012).

2.3 Basing design on authentic professional practices

Basing tasks on situations from authentic professional practices has been studied in science education with some promising results (e.g., Jurdak, 2006; Prins, 2010). In mathematics and statistics education the potential of this approach has received much less attention so far. In an earlier publication (Dierdorp, Bakker, Eijkelhof, & Van Maanen, 2011) we argued that such an approach assisted students in making statistical inferences about an authentic problem, but also helped them to grasp ideas that underlie statistical inference such as coordinating individual and aggregate views on data sets, recognizing trends that extended beyond the data, making sense of variability, and using data as evidence when drawing inferences.

Given the shuttling challenges mentioned in Section 1, and the required conceptual development, we formulate the following research question:

What is the potential of tasks based on authentic professional practices to support students in developing concepts that underlie sampling in such a way that they can shuttle between contextual and statistical spheres?

3 Method

3.1 Setting

This research question is addressed through two case studies that were part of two 23-lesson design-based research cycles with one group of thirteen twelfth-grade students and one of sixteen eleventh-grade students. The data of this paper stem from a four-year design-based research project (Barab & Squire, 2004; Van den Akker, Gravemeijer, McKenney, & Nieveen, 2006) with six design cycles, with the aim to investigate how students can learn to draw inferences at school level beyond correlated real-world data, in such a way that they experience a link between mathematics and the natural sciences.

The case studies, from the fifth and sixth cycle, were situated in two pre-university education (VWO) schools. This educational track for Grades 7-12 (ages 12-18) is meant for students who have the option to go to university. Both groups had opted to study the school subject “Nature, Life, and Technology” (SLO, 2008) for Grades 11 and 12 (16-18-year-olds), which the Dutch government introduced – among other reasons - to foster better connections between mathematics and the natural sciences. The first group was taught at the start of Grade 12 by the first author. All lessons were video recorded. The second group from another school was taught by another teacher at the end of Grade 11, with the first author observing and interviewing students (video recorded).

The two sampling tasks that we discuss here covered three of the 23 lessons, each 50 minutes in both schools. Students used computers with both the educational software tool *Fathom* (Finzer, 2006) for learning statistics to simulate drawing samples and a spreadsheet (Excel). Dutch mathematics curricula for the pre-university track include the law of large numbers. Therefore, we can assume that students in grade 11 or higher have had the opportunity to learn about it. Correlation and regression are not part of the obligatory syllabus. Dutch curricula pay relatively little attention to statistics compared to curricula in countries such as USA, UK, Australia and New Zealand.

3.2 Sampling tasks

To stimulate students to develop concepts underlying sampling such that they could shuttle between contextual and statistical spheres we designed two tasks based on authentic professional practices: a task about heart rate and a task about sampling the height of dykes. We followed three task design principles: (1) the task aims at engaging students with activities based on an authentic professional practice (cf. Dierdorp et al, 2011; Prins, 2010); (2) the task follows the design suggestion of Galbraith and Stillman (2006) to pay special attention to back and forth linking of mathematical solutions to the context; and (3) the purpose of the tasks should be clear to the students and students should see the utility of concepts (Ainley et al., 2006) in such a way that they could apply them in contexts. With these general design principles we tried to overcome the aforementioned planning paradox and stimulate shuttling.

3.2.1 Heart rate task

To stimulate students' reasoning about sampling we drew on a professional practice of research on peak heart rates (*PHR*). Gellish et al. (2007) measured many people and found a different relationship between age and peak heart rate ($PHR = 207 - 0.7A$, with A as age) than the one typically used in sport physiotherapy ($PHR = 220 - A$). We provided the students with Gellish et al.'s data set of 908 measurements. In the Heart rate task we asked the students if we could do with a smaller sample: What smaller sample size would be sufficient to find a reliable formula that is close to the original formula based on Gellish et al.'s data set? We consider this a hybrid task in the sense that it is both contextual and statistical: It is motivated by the contextual need to save money but at the same time it is statistically formulated in terms of sample size. This hybridity was expected to support students' shuttling between contextual and statistical spheres. Moreover, we anticipated that saving money would be a clear purpose, also from a student perspective.

The task is based on the instructional idea of growing samples. The idea is that, starting with a small sample, students could experience the limitations of what they can infer about the whole population from this sample. Such an approach is helpful in supporting coherent reasoning about sampling, distribution and other statistical key concepts (Bakker, 2007; Ben-Zvi et al., 2012). The Heart rate task can be seen as a shrinking sample task because students started with a large data set and Gellish et al.'s regression line, and then investigated random sampling (with replacement) of decreasing sample sizes to find the smallest sample size that still produced a reliable model (see Figure 1). It is also important that the students discover that repeated samples are needed to find a trend and draw inferences.

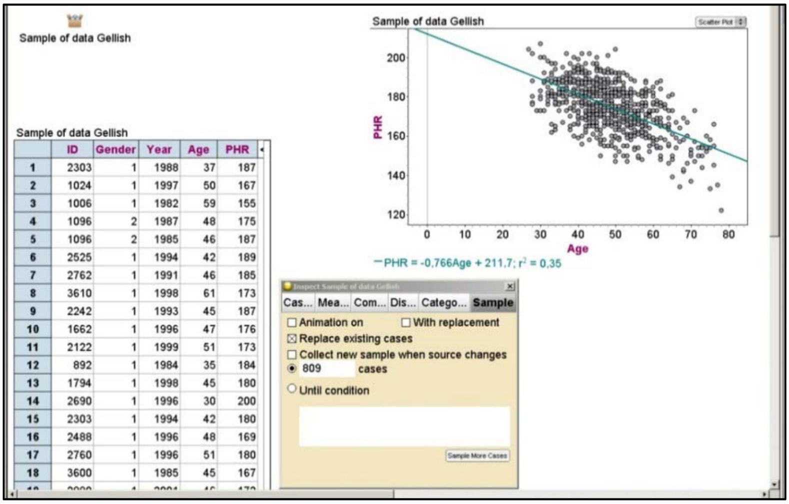

Figure 1. Fathom software with a sample of 908 peak heart rate measurements.

We conjectured that the question about how small a sample can be and still allow a sufficiently reliable inference can stimulate students to reason about all five aforementioned concepts underlying sampling (Section 2.2) and not just one aspect of sampling at a time, as happens in atomistic approaches to task design. We expected that students would become aware that small samples of the same size vary due to randomness and that increasing sample size would lead to more stability (law of large numbers). Further, we expected that drawing scatterplots would help students recognize the shapes of bivariate distributions (cf. Cobb, McClain, & Gravemeijer, 2003) to find trends and acknowledge the need to consider informal confidence intervals. Finally, we conjectured that the task's focus on how small the sample could be when trying to make a reliable inference would invite students to keep relating sample and population.

3.2.2 *Dyke sampling task*

The second task is inspired by the practice of monitoring the height of dykes. A dyke is an artificial construction to prevent flooding. Dyke monitoring is essential for the Netherlands because large parts of the country are below sea level. A persistent problem is that dyke heights decrease over time. If the height falls below a "critical value" high sea and river water levels become a danger.

Before the Dyke sampling task students were given authentic data, collected from helicopters and satellites for the detection and monitoring of surface deformation (quite an expensive way of monitoring dykes) and they modeled the data set with a linear function. The students had to draw an inference on the data set about when the height of a dyke should be increased, just as professional consultancy organizations for dyke monitoring do (see Dierdorp et al., 2011).

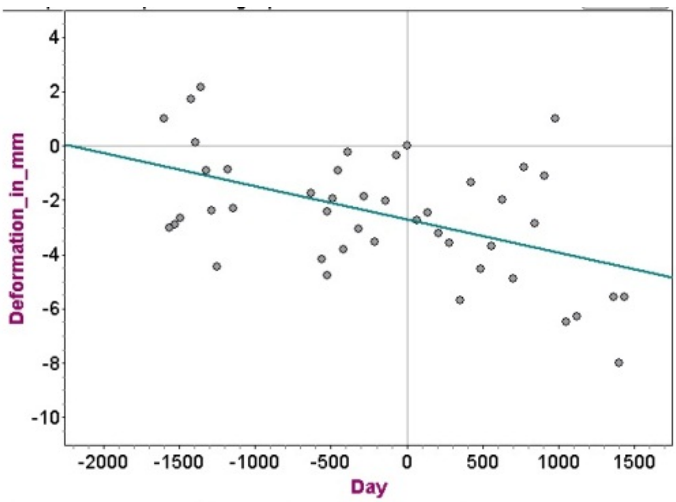

Figure 2. Scatterplot of deformations with 44 measurements during 3034 days (more than eight years). Day = 0 is used as a reference point for comparison with other dyke positions.

Like the Heart rate task, which preceded the Dyke sampling task, the latter is hybrid in the sense that it combines contextual and statistical concerns, thus stimulating students' reasoning about sampling while promoting the shuttling between contextual and statistical spheres. Pointing to the high cost of measurement, we asked the students what smaller sample would still have led to a reliable prediction of when the critical value will be reached. Because we had to simplify the professional practice we did not ask the students to consider the cost of constructing a higher dyke, or the cost of flooding. The students got 44 real data points of the deviations of a dyke location (Figure 2, from Delft University of Technology).

The students could change the sample size and get the corresponding scatterplot with a regression line and a formula. They had to decide themselves which number of measurements was required to find a reliable prediction based on a smaller sample which is close to the prediction obtained by the regression line for the complete set of 44 measurements. The students were not told what was meant by "close to" or "reliable," only that they had to save money by reducing the number of measurements.

To address the research question we sought an efficient way to check if our tasks could in principle support students' concepts underlying sampling in order to allow shuttling between context and statistics. To this end it seemed sufficient to use case studies of students working with these tasks. The advantage was that small numbers of students could be monitored more carefully than larger groups. Case study 1 focuses on concepts underlying sampling; case study 2 on the shuttling.

3.3 Case study 1

The first case study focused on a 12th-grade student, Rick (all names are pseudonyms), aged 17, with average marks. We chose Rick because he had average marks and he talked enough to be useful as an object of our study. While Rick worked on the tasks, we video recorded his activities and transcribed the spoken text. Data were also collected on the other students, which we occasionally draw on.

To identify which concepts underlying sampling Rick used in the interaction with the teacher or his peers, we divided the transcripts into 34 fragments of interactions about one element of reasoning between Rick and peer student or teacher. Each fragment consisted of several turns, i.e. the spoken text of a person which is not interrupted by another person as a turn (Chi et al., 2001). The fragments had an average time span of 15 seconds (*SD* = 8), and an average number of turns of 1.7 (*SD* = 0.6). Each fragment was coded with one or more codes referring to the five aforementioned concepts underlying sampling (Section 2). The inter-rater agreement measured with Cohen’s kappa was substantial (.69) according to Cohen (1960).

3.4 Case study 2

We focused the second case study on what teachers may need to do to use the task’s potential to stimulate students to shuttle between the contextual and statistical spheres in a case where they did not do this by themselves. For this second case study on another but similar school, Sean and Kars (grade 11, aged 17), were selected because their academic performance was similar to that of Rick.

This second case study focuses on the Dyke sampling task, during which the researcher observed and interviewed the students during the task. We video recorded the interaction and transcribed the spoken text. To analyse what teachers may need to do to stimulate students to shuttle between the contextual and statistical spheres we divided the transcripts into three phases, and identified the researcher’s attempts to help them shuttle back from statistics to context.

4 Results

4.1 Developing concepts underlying sampling (Case 1)

As we have argued before, the development of a concept should be understood in relation to other concepts (Bakker & Derry, 2011), but we qualitatively illustrate the five concepts underlying sampling one by one for reasons of readability. In 4.1.6 we provide a quantitative overview.

4.1.1 Sample size

Two aspects of sample size emerged as Rick worked through the task, coordinating and interpreting the statistics in non-trivial ways within the context of analyzing Heart rate data. First, we expected Rick to decrease sample size step by step, but after his start with the large data set, he jumped to a small sample size of 50 and increased step by step to the large set. From the data we have of the other students we know that all students grew rather than shrank their samples. Based on these observations we conjecture that growing samples is a more natural process for students than shrinking samples.

Second, when Rick drew samples of size 50 he noticed the variability in the coefficients of the regression lines when taking samples of this same size. He mentioned that when he drew bigger samples, he would expect a formula closer to the original formula of the population: "It [the regression line] is moving more towards the true regression line." Rick expected that when the sample size became big enough the regression coefficients would stabilize, possibly informally drawing on the law of large numbers. He also formulated it reversely as follows:

> Rick: If you sample 100 [instead of 500] cases, you get more variability. More deviations but the line will [still] have a negative slope. The intersection with the vertical axis will be different too. By changing the number of cases [sample size] you can monitor the effect on the regression line.

This fragment suggests that Rick may be aware of the effect of sample size on the variability of coefficients and was developing a sense of the law of large numbers.

4.1.2 Random process

To illustrate how Rick's conceptual understanding of sample size and law of large number became related to other concepts underlying sampling, we give one example on randomness. At first, he believed that a bigger sample size would lead to a more reliable regression line, but when he took a sample of size 100 he noticed that the regression coefficients deviated more from the original formula than those of the last sample of size 50. He then realized that a relatively small extension of sample size would not necessarily lead to a "better" formula and tried to explain this with the idea that the software sampled some higher peak heart rates even though it sampled randomly. The teacher tried to stimulate Rick to reflect on his former statement that a higher amount of cases would imply a formula for the regression line more similar to the original. Rick responded:

> Rick: But that is not always necessary because you always take a random group. So, you cannot expect it [the regression line] to come closer every step. But you could say that each time there is also variation in the difference

> according to the real [original] line. It [software] chooses different data points, but eventually those points get closer together. Gradually, the line will become more precise. It [the slope of the regression line] will be slightly above the -0.7.

Rick seemed to be clear about the fact that the randomness of the sample leads to unpredictable outcomes and that it is not necessarily the case that he would find a regression line more similar to that of the larger set when taking a slightly bigger sample. We see that Rick's simulations of repeated samples (cf. Shaughnessy, 2007) stimulated a growth in his understanding of how samples behave.

Another aspect shown in this excerpt was that Rick, like his peers, seemed to focus on one characteristic of the regression line: the slope. Initially, he was searching for a sample size which produced a slope close to the slope of the original regression line (the "real line"). Later he included the intercept of the regression line, which is contextually relevant in relation to the critical value. The observations gave us indications that he focused more on the statistics than the contextual interpretations. It seems that he treated the coefficients as isolated numbers without context, excluding realistic contextual considerations about what counts as close given the context.

4.1.3 Distribution

Two aspects of distribution came to the fore: that of a somewhat stable margin around a regression line and the overall decrease of that margin with larger sample sizes. First, in the context of the heart rate task, Rick and his fellow students decided that a client's peak heart rate must be within a margin of the regression line. When the teacher asked Rick if he expected a larger correlation when drawing a larger sample, he answered negatively and explained:

> Rick: You take new values each time. These are arbitrary and are not correlated to the regression line. When you start taking random values and do so the tenth time, taking random values, they are still random. Only the margin will become more colored. It will not be wider and will never become narrower and never become much wider.

This excerpt indicates Rick's sense of margin. He expected the margin to be about the same for each sample size. Only when drawing a sample with a bigger size did he mention that the margin became "more colored." He probably expected more points within the margin when the sample is larger, which suggests a stabilizing distribution.

Second, when Rick presented his results to the whole class, and the students looked at his graph of the slopes obtained by growing samples, they also considered the shape of the bivariate distribution (Figure 3), which he decided to call "a trumpet shape." He saw this shape in the Dyke sampling task as well. This time he did not draw the trumpet, but only mentioned: "You again see the trumpet shape". He used this shape to predict the values of other slopes when taking other sample sizes.

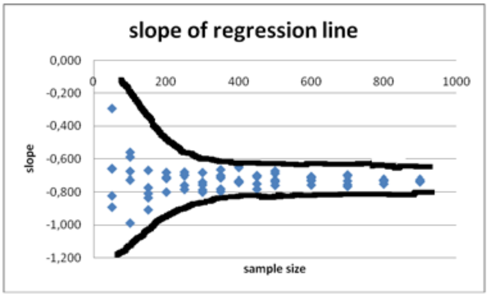

Figure 3. The "trumpet shape" distribution of the slopes of regression lines obtained by growing samples. This is a reconstruction of Rick´s drawing on the smart board.

4.1.4 Informal confidence interval

In both tasks, the concept of distribution seemed closely linked to that of an informal confidence interval. In the previous excerpt on the heart rate data Rick mentioned that with bigger samples the points were getting closer together. He entered the regression coefficients in a spreadsheet, and estimated a reasonable interval for each sample size (calling this the "margin"). He then found an acceptable margin for making a prediction for each sample size of 200, 250, 300, and up. In the next excerpt Rick considered a margin around an expected value:

Rick: Till 200 it is too varied. … It could be coincidence. You can build a safety margin, then you go to 250.

T: Why this margin?
Rick: Not everybody is the same.

He thought that a sample size of 200 would allow him to predict a reliable regression line but he explained that the margin of the slopes he found at size 200 could be small as the result of coincidence. To be safe, he suggested a sample size of one "step" bigger, 250. His response "Not everybody is the same" suggests that he thought of variability in people's heart rates, which suggests that the context helped him to think of sources of variability.

4.1.5 Relationship between sample and population

At this point we proceed with the Dyke sampling task. As with the Heart rate task, Rick seemed aware of the task's purpose: to find a smaller sample size in order to save money and still have a reliable formula for the regression line. He said to his fellow student Eline: "It [the Fathom software] plots the regression line. Then you are able to see how many points [measurements] can be saved and still find a reliable regression line." However, Eline was still not sure whether she understood the task. Rick tried to explain it to her, initially with no success. Rick started to show her that sometimes a sample with size 2 gives a regression line with a positive direction instead of the negative slope of the regression line of the complete data set. He used this fact to show that such a small sample was not sufficient to get a reliable linear relationship in the context of the data that it describes. Eline agreed that they needed to find a negative slope and that a sample with size 2 made no sense. Rick then showed her samples of size 4, 6, and 8, and compared the formulas of the corresponding regression lines with the formula of the regression line of the complete data set. After a while she asked him if he was in search of a sample size for which the average of the regression coefficients found by samples of a certain size was similar to the regression coefficients of the regression line of the complete set.

This episode about the reliability of a sample with a certain size indicates that Rick thought about the relationship between the sample and population distribution. He compared the results of every sample with the results of the original set and used words such as "similar," "real," "original" and "first outcomes" to refer to the data of the population. Some students had mathematical reasons for choosing a sample size of 22 (50% of the population), but Rick put forward contextual reasons, considering the deviations of the deformations between sample and population over time:

> Rick: It depends on how much you want to save. Sample sizes 36 have no big deviations any more. Then you save something around 18% and the deviation is…. That is acceptable when you consider five years, but it depends on how much you want to save. Saving 8 [= 44 – 36] times [helicopter flights] would give almost the same results [regression coefficients].

The observations indicated that other students also developed some understanding of the relation between the sample distribution and population distribution, but they did not link their solutions back to the context. Many only considered the trumpet shape distribution of the slope to consider the sample size for making a reliable inference. Only Rick (and with him Eline) shuttled between the contextual and statistical spheres: In their informal statistical inferences they included the number of flights which can be saved instead of only looking at the slope of the regression line. This suggests that Rick's conceptual understanding of sampling was richer than that of most peers'.

4.1.6 Quantitative overview

As a complement to the aforementioned qualitative examples, we now present a quantitative analysis of Rick's spoken text (Table 1). From the rather well balanced distribution of codes across the different concepts underlying sampling, we conclude that the tasks have the potential to stimulate students' coherent development of concept underlying sampling.

Table 1
Number of codes referring to concepts underlying sampling recognized in Rick's spoken text per task (coded fragments lasted 15 seconds on average).

Task	*Sample Size*	*Random process*	*Distribution*	*Confidence Interval*	*Sample/ Population*
Heart Rate	*3*	*4*	*4*	*3*	*3*
Dyke Sampling	*3*	*3*	*2*	*2*	*5*
Total:	*6*	*7*	*6*	*5*	*8*

4.1.7 Conclusions from Case 1

The question which smaller sample sizes would have been reliable enough in the context of determining heart rate and dyke height levels seemed to make sense to Rick and his peers, although some needed extra explanation. The tasks thus seemed purposeful. They also have the potential to assist students in developing concepts underlying sampling in relation to each other and in relation to meaningful contexts. However, most students did not by themselves link their results back to the context. In a subsequent teaching experiment, we therefore carried out another case study to gain further insight into how a teacher can use the task's potential to improve students' shuttling between the statistical and contextual spheres. We focus on the Dyke sampling task because it proved hardest to help students think through the consequences of their answers in contextual terms.

Table 2
Three phases identified during the Dyke sampling task from the second case study.

Phase	*Description*	*Students…*	*Duration (minutes)*
1	In the math world	…work independently on the task	20
2	Nudging to the context	…are interviewed by the researcher	19
3	Translating to the context	…work independently on the researcher's context question	7

4.2 Using the Dyke sampling task's potential to improve students' shuttling (Case 2)

From this case study we only present the Dyke sampling task, in which the researcher needed to put a great deal of effort in supporting Sean, who struggled with linking the statistical solution back to the context. We distinguished three phases in Sean's work on the task with his fellow student Kars (Table 2).

4.2.1 Phase 1: In the math world

During the first twenty minutes Sean and Kars worked independently on the Dyke sampling task. They drew a lot of samples before finding a suitable formula for the regression line and discussed which sample had produced an acceptable formula. Although all tasks were based on authentic professional contexts and the students seemed to see the purpose of the task, they tended to stay in the math world focusing on formulas. A typical interaction looked like the following discourse:

Sean: The slope [-0.00124, slope of the regression line at sample size 20] is almost the same [as the slope of the original formula, -.00123].

Kars: Yes, the *b* too [the *b* in the formula y = ax + b; -2.74 vs. -2.8].

4.2.2 Phase 2: Nudging to the context

After about twenty minutes, the researcher discovered that Sean and Kars were reasoning without referencing back to the contextual problem. He tried to refocus their attention on the contextual meaning of their decisions in several ways. For example, the following question was intended to nudge them to consider the contextual meaning of their reasoning:

R: When this [regression value] is -10 [millimeters] we need to heighten the dyke (see also Figure 2). How many measurements do you need to get the same result as with the original regression line? [Phase 2, 10:40]

However, Sean and Kars stayed focused on the slope of the regression line and did not use the context. Often the slope was different only in the fourth decimal, and they thought that the formula obtained by sample size 30 (H = -0.00123d – 2.8) was close enough to the original formula (H = -0.00149d – 2.9). They judged this purely on the basis of the formula, whereas the researcher hoped they would think through the contextual consequences of the differences between the regression formulas. However, they did not.

4.2.3 Phase 3: Translating to the context

At the end of phase 2, the researcher again tried to narrow down the possibilities by asking more specifically about the difference in days between the prediction based on the regression line with sample size 30 and the prediction based on the original formula based on a sample of 44. Only then did they realize that the context asked for a more precise approach:

R: What would such a small difference in coefficients mean in terms of the prediction?

Sean: It differs a lot.

Kars: 330 [days], that is a lot.

Sean: Yes, a lot indeed. It [formula] seemed to differ slightly but it differs too much.

Because Sean and Kars were no longer satisfied with sample size 30, they decided to take samples of size 35, 40, and 42, each five times. For each sample they also calculated the difference between the corresponding calculated predicted days and the day when the original formula would predict the critical value of -10 mm. They also calculated the average of differences for each sample size. They decided that in this context sample size 40 was acceptable. They were disappointed that they only saved four measurements, but when the researcher asked how much money would be involved in skipping four helicopter flights, they were more satisfied.

This last case study suggests that the students were focused primarily on the mathematical concepts of the tasks. The researcher repeatedly had to emphasize the contextual problem to stimulate the shuttling back to the contextual sphere. He had to ask about specific contextual consequences of difference in the formula obtained by their sample and the original formula to encourage them to make a meaningful inference within the context. This last kind of support proved necessary for students such as Sean and Kars who at first were satisfied with their mathematical solution.

5 Conclusion and discussion

The aim of this paper was to examine whether realistic tasks, inspired by authentic professional practices, can support students' conceptual understanding in such a way that they can shuttle between contextual and statistical spheres. In line with Ainley et al. (2006) we assumed that such tasks needed to provide students with purpose and utility. It seems possible to use authentic problems from professional practices to design tasks that are purposeful from a student perspective and lead students to see the utility of what they learn (Westbroek et al., 2010). This may help students apply what they have learned. However, it is not self-evident that students develop rich conceptual understanding from authentic tasks because designers seem to have less control about what students learn conceptually. Yet such conceptual understanding seems necessary for students to link contextual and statistical considerations.

In response to the research question we conclude that the analyses show that the realistic sampling tasks, inspired by authentic professional practices, are rich and focused enough to stimulate reasoning about the concepts underlying sampling in a balanced way and in relation to each other. This seems an advantage over atomistic approaches to statistics education that deal with aspects of concepts one by one (cf. Bakker & Derry, 2011) and this seems to address Ainley et al.'s (2006) concern that engaging tasks are often less focused.

In Case 1, we argued explicitly that the task suitably assisted students (in this case, Rick) to develop the main concepts underlying sampling (cf. Pfannkuch, 2008). In both cases presented, we had asked students to consider how small a sample can be so that the inference is still reliable. This required them to reason about sample size and the relation between sample and population. The issue of randomness came up as students wrestled with sampling variability in comparing samples with the same size to judge if inferences are reliable. When comparing samples of the same size, students generated informal confidence intervals and interpreted bivariate distributions. These findings align with earlier findings at the middle school level that tasks based on growing samples have the potential to stimulate students to reason about multiple facets of distribution and uncertainty (Bakker, 2007; Ben-Zvi et al., 2012).

In line with Brandom's (2000, p. 48) inferentialist perspective on concepts, conceptual understanding is "to have practical mastery over the inferences it is involved in—to know, in the practical sense of being able to distinguish, what follows from the applicability of a concept, and what it follows from." This implies that conceptual understanding includes the ability to shuttle between contextual and statistical spheres. Both cases showed this to be challenging, or at least unfamiliar to students in school mathematics. It seems that reasoning within either the statistical or contextual world is easier than making connections. Developing tasks that have a clear purpose and lead to concepts with utility is not sufficient. For the students involved the purpose of the tasks was clear: to find a smaller sample size to save money and still have a reliable formula for the regression line. It seemed that they did see the utility of sampling to find such a reliable formula.

Because for some students the tasks were not sufficient to support their shuttling back to the context, we explored in a second case study what types of questions teachers may need to ask. Asking specific questions about practical consequences of mathematical issues (e.g., what would such a small difference in coefficients mean in terms of the prediction?) appeared to be one element which assisted the students in refocusing on the context sphere. It is of course possible to include such more specific questions in the instruction of the task, for example by stating the costs of a helicopter flight and asking students how much they expect to save by using smaller samples. However, we think it may be beneficial for the richness of students' reasoning if they first think through many options themselves. In their study, at least, Bakker and Derry (2011) argued that sixth-grade students had developed rich concepts of variation which would be unlikely if the teacher had focused on measures of centre in well-structured problems.

The small scale of our case studies puts limitations on the generalizability of our results. Yet they do suggest that the designed tasks have the potential to support students' reasoning of sampling when making authentic inferences. Moreover, we think that the data analysis method we used in the first case study (Table 1) would be useful in other contexts in which designers or researchers want to know if tasks have the potential to stimulate reasoning with a particular coherent set of concepts. This seems especially important when using authentic or realistic tasks in which we seem to have less control over what is learned conceptually.

We addressed a persistent design challenge and do not claim to have solved it. More research is needed to investigate how to support students' conceptual development needed to shuttle between contextual and statistical spheres. More specifically, we suggest investigating teachers' scaffolding of students' shuttling.

6 Acknowledgements

We thank Katie Makar and Dani Ben-Zvi for their helpful comments on earlier versions of this manuscript.

References

Ainley, J., Pratt, D., & Hansen, A. (2006). Connecting engagement and focus in pedagogic task design. *British Educational Research Journal, 32*(1), 23–38.

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal, 3*(2), 64–83.

Bakker, A. (2007). Diagrammatic reasoning and hypostatic abstraction in statistics education. *Semiotica, 164*, 9–29.

Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning, 13*(1&2), 5–26.

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences, 13*(1), 1–14.

Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM – The International Journal on Mathematics Education, 44*, 913–925. doi: 10.1007/s11858-012-0420-3

Boaler, J. (1993). Encouraging the transfer of 'school' mathematics to the 'real world' through the integration of process and content, context and culture. *Educational Studies in Mathematics, 25*, 341–373.

Brandom, R. B. (2000). *Articulating reasons: An introduction to inferentialism*. Cambridge, MA: Harvard University Press.

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25*, 471–533.

Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction*, *21*, 1–78.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46.

Dierdorp. A., Bakker. A., Eijkelhof. H. M. C., & Van Maanen, J. A. (2011). Authentic practices as contexts for learning to draw inferences beyond correlated data. *Mathematical Thinking and Learning, 13* (1&2), 132–151.

Finzer, W. (2006). *Fathom Dynamic Data™ Software* (Version 2.03) [Computer software]. Emeryville, CA: Key Curriculum Press. Available: http://www.keypress.com/fathom.

Freudenthal, H. (1991). *Revisiting mathematics education, china lectures.* Dordrecht: Kluwer.

Galbraith, P., & Stillmann, G. (2006). A framework for identifying student blockages during transitions in the modeling process. *Zentralblatt für Didaktik* *der Mathematik, 38*, 143–162.

Gellish, R. L., Goslin, B. R., Olson, R. E., McDonald, A., Russi, G. D., & Moudgil, V. K. (2007). Longitudinal modeling of the relationship between age and maximal heart rate. *The American College of Sports Medicine, 39*, 822–829.

Hung, W., & Jonassen, D. H. (2006). Conceptual understanding of causal reasoning in physics. *International Journal of Science Education, 28*, 1601–1621.

Jurdak, M. E. (2006). Contrasting perspectives and performance of high school students on problem solving in real world situated, and school contexts. *Educational Studies in Mathematics, 63*(3), 283–301.

Konold, C., & Pollatsek, S. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*, 259–289.

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

Nunes, T. , Schliemann, A. D., & Carraher, D. W. (1993). *Street mathematics and school mathematics.* Cambridge, UK: Cambridge University Press.

Palm, T. (2008). Impact of authenticity on sense making in word problem solving. *Educational Studies in Mathematics, 67*(1), 37–58.

Pfannkuch, M. (2008, July). Building sampling concepts for statistical inference: A case study. Paper presented at the International Congress on Mathematical Education, Monterrey, Mexico.

Prins, G. (2010). *Teaching and learning of modeling in chemistry education.* Utrecht, the Netherlands: CD-Bèta Press.

Saldanha, L., & Thompson, P. (2003). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51*, 257–270.

Shaughnessy, J. M. (2006). Reseach on students'understanding of some big concepts in statistics. In G. F. Burrill & P. C. Elliott (Eds.), *Thinking and reasoning with data and chance* (pp. 87–98). Reston, VA: National Council of Teachers of Mathematics.

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

SLO (2008). *Outline of a new subject in the sciences.* Enschede: Stichting Leerplan Ontwikkeling.

Van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (2006). *Educational design research*. London: Routledge, Taylor & Francis.

Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modeling of school arithmetic word problems. *Learning and Instruction, 4,* 273–294.

Westbroek, H., Klaassen, K., Bulte, A., & Pilot, A. (2010). Providing students with a sense of purpose by adapting a professional practice. *International Journal of Science Education, 32*, 603–627.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*, 223–265.

Chapter 5

MEANINGFUL STATISTICS IN PROFESSIONAL PRACTICES AS A BRIDGE BETWEEN MATHEMATICS AND SCIENCE: AN EVALUATION OF A DESIGN RESEARCH PROJECT

Abstract

Creating coherence between mathematics and science, and making these school subjects meaningful are still topical challenges. This study investigated how students made meaningful connections between mathematics, statistics, science and applications when they engaged in a specially developed unit. This unit is based on professional practices in which mathematical, statistical and scientific knowledge is used. The central question is to what extent professional practices can serve as meaningful contexts for senior high school students (aged 16-17) to help them make connections between mathematics, statistics, science and professional practices. Surveys on the opinions of students (388 before and 27 after completing two chapters of the unit) on the educational strategy, and student work are used to answer this question. The analysis of responses to surveys shows that students consider an educational strategy based on authentic professional practices meaningful. The results indicate that an educational strategy based on professional practices can help students to make connections between mathematics, statistics, science and professional practices.

Keywords: *coherence, statistical reasoning, inferential reasoning, authentic professional practice, upper secondary education, meaningful education*

1 Introduction

The purpose of this study is to determine whether senior high school students learn to see the connections between mathematics, statistics, science and some professional practices through an instructional unit based on professional practices in which statistics is used. Statistics is seen here as a possible bridge between mathematics and some of the natural sciences, because many scientists frequently apply statistical and mathematical techniques (Erickson, 2002). Statistics is a mathematical science, but not part of mathematics (Moore & Cobb, 2000), although statistics in most countries is part of the mathematics curriculum of secondary education (Gattuso, 2006). Statistics could, therefore, function as a bridge, so as to connect a part of mathematics with the sciences. Our question was: Would students also recognize this, and would it improve their learning?

It is important that students experience coherence between different sciences. Mathematics and statistics provide the tools by which quantitative relationships in the natural sciences can be modelled, calculated, represented, and predicted, and the natural sciences provide relevant contexts in which mathematical and statistical knowledge can be applied (Davison, Miller, & Metheny, 1995). Therefore, international and Dutch national committees (AAAS, 1989; cTWO, 2007; Boersma et al., 2007; NCTM, 2000; NiNa, 2010; Apotheker et al., 2010; NRC, 1996; Stuurgroep-NLT, 2007) advocate the integration of mathematics and the natural sciences, where possible.

However, making the relationship between science and contexts clear to students appears to be difficult (Berlin & White, 2012). Often, students in secondary education do not experience the meaningfulness of science because they experience the learning within school subjects as a train with individual wagons, whose windows are blinded and only the driver on the engine knows where the journey goes to (Claxton, 1991). A report on alignment between mathematics and physics (Van de Giessen, Hengeveld, Van der Kooij, Rich, & Sonneveld, 2007), speaks in this context of "sectarianism". This compartmented thinking also makes it difficult for students to recognize what they have learned as meaningful, and apply it to another subject or in other contexts (Bransford, Brown, & Cocking, 2000). More curricular coherence is needed. Berlin and Lee (2005) provide an overview of the U.S. efforts to curricular cohesion in the period 1970-2001. In this period the number of publications on curricular coherence increased significantly, but the authors conclude that there is a need of more insight in how more cohesion can be achieved.

The literature provides no univocal definition of curricular coherence (Hurley, 2001), well understandable from the various manifestations of a curriculum — intended, implemented and attained curriculum (Van den Akker, 2009). The intended curriculum includes the vision that lies at the basis of the curriculum, but also the formal texts in which the curriculum is described. Within an intended curriculum coherence refers to aligning the curriculum between or within different school curricula. Intended cohesion must be reflected in the implemented curriculum that includes the interpretations of teachers and the resulting practice of education. A curriculum refers to the learning outcomes achieved and the experiences of the students. As several researchers (e.g., Newmann et al, 2001; Rudduck & Wallace, 1994) point out, there is very little research on coherence in the attained curriculum, in other words: how curricular coherence for students works (Frykholm & Glasson, 2005) and how students make and observe connections.

Furner and Kumar (2007) underline the recommendations of Berlin and White (1992) and Sunal and Furner (1995) for the integration of mathematics and science by promoting overlapping subject matter and supporting students to search for patterns in data to get a more meaningful view of scientific phenomena. Sunal and Furner also stress the importance that school bridges the gap between school practice and extracurricular life. Bennett, Lubben, and Hogarth (2007) provide clear evidence that "context-based" education contributes to the making of meaningful education. The affective reactions and motivation of their students gave them indications that this strategy leads to better understanding of science education, a finding that is supported by Scott, Ametller and Mortimer (2011). They investigated how teachers and students make connections in meaningful interactions between teaching and learning of scientific concepts, but their study had no special attention to mathematics.

Considering these challenges to achieve coherence and meaningful science education, renewal committees for science and mathematics education in the Netherlands advise using the so-called concept-context approach: students are expected to learn concepts in meaningful contexts. These contexts may be based on authentic practices, such as scientific or professional practices in which scientific knowledge is applied. The use of these practices is a possible way to make science education meaningful.

Meaningful scientific concepts and contexts involve various aspects of affective, cognitive and meta-cognitive nature. It is important that students see the purpose ("need to know") of what they need to learn (Westbroek, 2005) and become motivated by the context to engage in the instructional unit (Prins, 2010). Engagement in the sense of motivated behaviour, according to the learning motivation model of Eccles et al (1993), is determined by the value that students assign to their task (cf. Volman, 2011). Furthermore, students recognize their education as more meaningful if it is authentic and if they are able to apply the concepts learnt (Boersma et al, 2007; Volman, 2011). Clarke (1988) argues that students experience lessons more meaningful when dealing with authentic contexts such as sporting contexts. Although several mono-disciplinary units based on authentic professional practices exist for biology and chemistry (Prins, 2010; Westra, 2008), there is little research on how the relationship between school subjects can be promoted by materials that are based on professional practices where knowledge from different sciences is integrated.

In the current study, we define coherence between mathematics, statistics, science and professional practices for the attained curriculum as the ability of students to make sense of the contexts so that they can apply scientific and mathematical knowledge when solving authentic problems. This fits the definition of the intended curriculum in which the relationship between mathematics and science is seen as an interdisciplinary mix, in which connections between these disciplines are placed while the subject specific concepts remain recognizable (Lederman & Niess, 1997).

In this article we test our assumption that education is meaningful for students and that students can apply their learning in other disciplines or in other contexts when it was based on problems in authentic professional practices. We assume a reciprocal relation between coherence and meaning: when students observe coherence between different science subjects, these subjects will have more meaning for them, and vice versa: meaning helps them to see coherence. We now first give an overview of the efforts in the Netherlands to facilitate coherence between the sciences and then discuss the rationale for the choice of an instructional unit based on professional practices.

2 Theoretical backgrounds

2.1 Coherent teaching in the Netherlands

One of the objectives of the Dutch government during the past decade was to promote greater coherence between school subjects, but curricular coherence is still fragile and needs continued attention (Boersma, Bulte, Krüger, Peter, & Seller, 2010; Nieveen, Handelzalts, & Van Eekelen, 2011). The introduction of a new curriculum for upper secondary school (1998) intended, among many other aims, to establish such coherence. In 1999, the Sonata project (Coherent Education in Science and Technology) started to monitor good practice in the field of coherent teaching in lower secondary and in the science stream of upper secondary (Geraedts, Boersma, Huijs, & Eijkelhof, 2001). At the end of this project in 2004 the team concluded that intrinsic coherence in educational practice was still a “white spot”. The Sonata project was continued in the Salvo project (Coherent Learning in Secondary Education), which among other things sought to develop exemplary material for an on-going learning strand about the relationships between quantities - a theme that lends itself to a coherent approach (Mooldijk & Sonneveld, 2010). A recent initiative to foster interdisciplinary cohesion is the introduction of Nature, Life and Technology (NLT), a new three year course in addition to physics, chemistry, biology and mathematics (NLT Steering Committee, 2007).

Some renewal committees for science and mathematics education have published documents in which they tried to make visible the link between the new syllabuses (Boersma et al, 2010). Within these committees, there is little attention to the possible role of statistics in creating coherence, although statistics seems appropriate to let students experience such coherence. cTWO, the committee for mathematics, only states (2007) that it is important for students to learn that mathematics is indispensable in engineering and science and that it is closely intertwined with everyday life. Here statistics is mentioned as one of the integrating factors. Furthermore, in the committee reports of mathematics and physics, cTWO and NiNa (Van de Giessen et al, 2007; NiNa, 2010) not much is written about statistics. Rather, modelling is the preferred tool for students to experience the coherence between the various disciplines.

2.2 Professional practices as a basis for an instructional unit

Authentic practices are "en vogue" as an inspiration for teaching and learning strategies. Educational research about science and mathematics shows how one can develop teaching materials based on authentic scientific or professional practices (eg, Lee & Songer, 2003; Westbroek, 2005). Lee and Songer used the practice of weather forecasting, and Westbroek used the practice of testing the water quality to involve students in learning chemistry. Westbroek used a profession that made the featuring chemical concepts for students more meaningful. A disadvantage of this teaching strategy is that so much emphasis will lie on the professional aspects, that little school knowledge is learnt. Conversely, in mathematics education contexts, if they already exist, usually are subordinated to the learning concepts (Boaler, 1993).

For a good balance the steering committee "New Chemistry" advises to use the concept-context approach. It states that major developments in scientific research as well as in industry (e.g., nanotechnology) are important since they can offer students a meaningful curriculum based on interdisciplinary contexts (Pharmacist et al, 2010). The Committee "Renewal Biology Education" (CVBO) in its concept-context approach explicitly opted for "social practices", such as scientific or professional practices, as an inspiration for teaching learning strategies (Boersma et al, 2007). This approach is founded on the cultural-historical activity theory (Van Oers, 1987).

For educational purposes, professional practices are first educationalized because the authentic acts are often too complicated for students. Also, professional practice and educational practice have different purposes. In a professional practice one may want to test or optimize a process, while in a teaching practice students have to learn something.

We have, just as Westra (2008), chosen to offer students the concepts within different educationalized practices, so that they learn to apply the concepts in several different contexts. The underlying idea is, to reveal to students the relationship between mathematics, statistics, science within the contexts in which these disciplines function. In the present study, correlation and regression are chosen as statistical concepts because these are applicable in science and the students have such a level that they can master them. Moreover, they receive little attention in the curricula at the upper secondary level. During the especially designed instructional unit students themselves come to the idea that they need statistical techniques such as correlation and regression models in order to solve the real problems that they are facing in the authentic professional practices (Dierdorp, Bakker, Eijkelhof, & Maanen, 2011) They need to handle both mathematical and scientific concepts. The instructional unit is called: Statistics as a bridge between mathematics and the natural sciences. In the unit there is a focus on the understanding of the new concepts, which at the same time highlight the connections between mathematics, statistics and the respective professional practices.

2.3 Connections between mathematics, statistics, science and professional practice

In accordance with the concept-context approach we have chosen to not only investigate disciplinary coherence, but also coherence between disciplines and their areas of application (contexts). Disciplinary coherence is visible in some disciplines such as mechanics, a part of physics with a strong mathematical foundation. Modelling is often an interdisciplinary activity, which is also recommended to students to clarify the coherence between science subjects (Van de Giessen et al, 2007). As a premise that underlies our investigation, we suspect that an educational strategy in which statistics functions in a professional practice, will work as a bridge between mathematics and some science (Figure 1); we see it as a promising approach that has received little attention in the literature. Although there exist direct connections between mathematics and the natural sciences (e.g., through differential equations), we focus in this article, as we did in the instructional unit, on a more indirect relationship, generated by the application of statistics within professional practice. In this section we discuss the relationships that are central to this article.

Figure 1. Statistics used in a professional practice as a bridge between mathematics and natural science (bridge metaphor).

2.3.1 Mathematics and Statistics (MS)

Mathematicians generally strive to abstract the mathematics from the contexts, so that underlying structures become clear, while statisticians precisely need to analyse the data (Rossman, Chance, & Medina, 2006). Mathematics looks for statements or theoretical underpinnings of a phenomenon, and seeks, through statistical analysis, for significance (Moore & Cobb, 2000). Statistics, for example in the form of correlation and regression, can be used to derive a mathematical model. Mathematical procedures such as calculation methods or solving equations should in turn involve the statistics to substantiate or justify.

2.3.2 Statistics and Natural Science (SN)

As mentioned, correlation and regression often are used for the analysis of data collected in a scientific experiment, with the purpose of eventually finding a model. In such situations, statistics provides a bridge between mathematics and natural science. One example in our study is the determination of the moment in the training of an athlete when the metabolism turns from the aerobic to the anaerobic phase. Data on the pulse of the athlete are collected while the training effort increases, and subsequently statistical techniques such as correlation and regression are used to create a mathematical model with the purpose of giving optimal training advice (Gellish, Goslin, Olson, McDonald, & Moudgil, 2007).

2.3.3 Professional Practice and Statistics (PS)

Statistics is used in many professional practices. One example in our study is the analysis of dyke heights. All dykes sag continuously. In order to predict when a dyke should be raised, regularly data about the height of a dyke are collected with helicopters and satellites. Regression is then used to create a mathematical model which helps to minimize the risk of flooding. Another example is the Dutch Metrology Institute, the former institute of weights and measures, which works with regression models for calibrating measuring instruments. Comparison of the characteristics (correlation coefficient and regression line) of a measuring instrument of unknown quality with the characteristics of a calibrated instrument provides the Institute with a proof of the quality level of the tested instrument.

2.3.4 Mathematics and Professional Practice (MP)

In the previous examples of professional practices there was a clear link with mathematics. Galbraith and Stillmann (2006, p. 150) write: “In some cases, an adequate response requires arguments that integrate mathematical knowledge with the impact of this knowledge in the real situations to justify interpretation.” It is important that students understand how a model works and how mathematical concepts, such as solving equations, play a role in creating or interpreting the models.

2.3.5 Professional Practice and Natural science (PN)

If a professional scientist has collected experimental data and has found a mathematical model that represents these data, they must be interpreted in order to draw conclusions. In the example of dyke subsidence it is important to understand that landslides or erosion can be responsible for this subsidence. Also, a natural phenomenon such as heavy rain provides much variability around a model. These aspects will have to be estimated and discussed within a professional advice. It is important that students understand that nature does not exactly behave as the model does, and that there is always variability (Wild & Pfannkuch, 1999).

We note that the boundaries between the five links are not always sharp. Mathematical statistics, for example, will be located somewhere between mathematics and statistics, and scientific researchers are at the same time practitioners.

2.4 Research questions

To investigate whether a teaching strategy based on professional practices enables students to make meaningful connections between mathematics, statistics, science and professional practice, we have designed an instructional unit, based on three professional practices: a sports physiologist, an official of Public Works and Water Management, and an analyst at the Dutch Metrology Institute. In these three professional environments the professionals work with statistical techniques to analyse their data. To investigate to what extent the unit contributes to achieving our goals we propose the following main question:

> *To what extent can professional practices serve as meaningful contexts to show students connections between mathematics, statistics, science and professional practices?*

To answer this question we raise two new questions:

RQ1: *How meaningful do students find an instructional unit that is based on professional practices in which statistics plays a role?*

RQ2: *To what extent are students able to check within the instructional unit the connections between mathematics, statistics, science and professional practices?*

3 Method

3.1 Educational unit

The study presented here, is part of a larger project, which works along the lines of developmental research (Barab & Squire, 2004; Van den Akker, Gravemeijer, McKenney, & Nieveen, 2006). It consists of six design cycles with professional practices as a basis for teaching learning strategies. For the selection of the three professional practices we started with a preliminary study, which included a literature review and interviews with experts, practitioners, teachers and students. Criteria for the selection of practices were (cf. Prins, Bulte, Pilot, & Van Driel, 2008):

- the practitioner uses correlation and regression in data modelling
- the actions of at least one professional practice are educationalized into teaching activities, in which students can perform a short, relevant experiment of their own
- for students, the professional practice is easy to recognize
- students can see the relevance of the knowledge that the practitioner uses

Based on each of the three professional practices we designed a section for the instructional unit. Section 1 (H1) is about the professional practice of a sports physiologist, Section 2 (H2) is about the monitoring of dyke heights and Section 3 (H3) concerns the calibration of measuring instruments.

3.2 Participating students

To answer Research Question 1, we examined 415 (198 + 190 + 12 + 15) students (see Table 4). In 2008 they were all 5-VWO students from two schools (2x A and B) and in 2011, two groups which studied at the school for Nature, Life and Technology (NLT) (school A and C). The group (N = 15) of School A went through the instructional unit beginning 6 VWO. One group (school C, N = 16) was examined to answer RQ2. VWO stands for the Dutch secondary school type , grades 7-12 (5 VWO = grade 11, 6 VWO = grade 12), that gives entrance to selected university studies, the selection depending on the courses taken in the final three years of VWO. To measure the knowledge of the students, we have asked two groups, prior to administering the unit, if they could write a formula in which a regression line and the related correlation were calculated. On both questions they answered unanimously “no”. This was expected because the subject correlation and regression was not a mandatory part of the VWO curriculum. On the other hand, the students had already learnt in their mathematics program about statistical measures such as mean, median and they could calculate the standard deviation of a data set.

3.3 Measurement Instruments

To answer Research Question 1, we used surveys. Prior to the completion of the instructional unit, we asked students (N = 198) in a survey (Enq_1) their opinion about a possible teaching-learning strategy which would be based on professional practices. We asked questions about the aspects that we thought to be meaningful for students (see Table 5). We asked for example if such a teaching-learning strategy would help the students to understand the usefulness of the course material (item 1); also if it gave them motivation to learn (item 2), and if they expected that the techniques they learnt would be applicable in other subjects (item 3 and 4) and if they find it useful when the course material goes across multiple school disciplines (item 5). The students could express their views on a five point scale from “strongly disagree” to “strongly agree”. In the next cycle the survey was improved (Enq_2) and shortly after administered to a new group of students in 5 VWO (N = 190). We give the results of this second survey as far as the items are relevant to the research questions in this article and indicate whether the students think an instructional unit based on professional practices to be meaningful.

Furthermore, we have asked the final two groups of students (N = 15, 6 VWO, resp. N = 16, 5 VWO) after the completion of H1 (Enq_3) and H2 (Enq_4) if they think to have acquired more insight in the use of statistics in the presented professional practices (item 6) and whether they expect the exercises in the instructional unit to be authentic (item 7 and 8). For organizational reasons, we could not ask the same questions about Chapter 3. Of four students some data are missing. These are reported in Table 4, with a negative number. In the analysis, we therefore studied 12 + 15 = 27 students.

To answer Research Question 2 (about the connections between mathematics, statistics, science and professional practices), we asked students during the instructional unit to fully elaborate on paper their answers of the assignments. We have collected and analysed all this written work, which we call, from now on, the students' work. Next, in a posttest after the completion of the instructional unit, we analysed on the basis of the students' work if the students mastered the material and could explain the connections. The posttest involved all students. Within the framework of the bridge metaphor, we examine five possible connections between mathematics, statistics, science and professional practice (Figure 1). Table 1 shows which links relates to which tasks in the posttest.

Table 1
Possible connections MS, MP, PS, PN, and SN to be made in the posttest (see Figure 1 for the meaning of the codes).

Task	MS	MP	PS	PN	SN
Task 1	x	x	x	x	x
Task 2	x				
Task 3	x		x		
Task 4	x		x	x	x

To answer Research Question 2 (about the connections between mathematics, statistics, science and professional practices), we asked students during the instructional unit to fully elaborate on paper their answers of the assignments. We have collected and analysed all this written work, which we call, from now on, the students' work. Next, in a posttest after the completion of the instructional unit, we analysed on the basis of the students' work if the students mastered the material and could explain the connections. The posttest involved all students. Within the framework of the bridge metaphor, we examine five possible connections between mathematics, statistics, science and professional practice (Figure 1). Table 1 shows which links relates to which tasks in the posttest.

In the analysis of the students' work and the posttest, we investigated whether the instructional unit contributes to the links that students perceive between mathematics (M), Statistics (S), professional practice (P) and natural science (N). We developed an analytical model (see Figure 1 and Tables 2 and 3) for analysing the written response. With this we have distinguished the students work of the last group (N = 16) about elements M, S, P and N (Table 2) and the students answers coded with MS, MP, PS, PN and/or SN (Table 3).

3.4 Data analysis

In the analysis of the surveys we compared the same relative frequency of relevant items at Enq_2, Enq_3, and Enq_4. Enq_2 also contained an open question, where the students were invited to explain why they do, or do not, appreciate professional practices as the basis of the curriculum. The answers are coded as positive, negative or inconclusive. The inter-rater reliability was measured using Cohen's kappa (Cohen, 1960) and proved to be very high (.91).

Table 2
Definitions of the considered elements.

Element	Examples
M	The student indicates that calculations must be carried or used. The student performs a calculation. The student gives a mathematical relationship between two quantities. The student uses formulas.
S	The student interprets or refers to a graph or table using the data. The student uses a statistical term (e.g., mean, variation, SD, residual, etc.). The student mentions a methodological aspect (e.g., representativeness, keeping certain conditions constant, etc.).
P	The students mention an observed physical variable (e.g., weight, age). The student makes a link between two physiological variables (e.g., heart rate and intensity of training).
N	The students mention a quantity from a natural (e.g., weight, age). The student performs a connection between two scientific properties (e.g., heart rate and intensity of training).

Multiple codes per response were possible, what was needed given the research question, but what affected the inter-rater reliability (Cohen's kappa), which was measured on the basis of 150 student responses. The measured value of kappa (.61) is considered by Cohen (1988) as still substantial. The first task of the unit was similar to the first task of the posttest for a comparison between relationships that students make at the beginning and end of the instructional unit. The work of two students was not complete. These are not included in the analysis.

Multiple codes per response were possible, what was needed given the research question, but making it difficult to get a high kappa. On the basis of 150 student responses, we have the inter-rater reliability measured by Cohen's kappa. The measured value (.61) by Cohen (1988) is seen as substantial. The first task of the unit was similar to the first task of the posttest for a comparison between relationships that students make at the beginning and end of the instructional unit. The work of two students is not complete. These are not included in the analysis.

Table 3
Codes used to code students' work and task 1 from the posttest.

Code	Explanation
MS	The student involves mathematical techniques in the interpretation of the graphical representation of the data.
MP	The student involves mathematical techniques in formulating a sport physiological advice.
PS	The student uses statistical techniques to formulate a sport physiological advice.
PN	The student works with a physical/biological variable in the sports physiological advice.
SN	The student explains statistical results from the underlying physical/biological aspects.

Table 4
Numbers of students surveyed in the analyses of 2008 and 2011 (in brackets the number of students in the group and the number of students whose information was not complete).

		RQ1				RQ2	
School	Year	Enq_1	Enq_2	Enq_3	Enq_4	Student work/posttest	Posttest
A	2008 (June)	198					
A, B	2008 (Nov.)		190				
A	2011			12 (15 – 3)	12 (15 – 3)		
C	2011			15 (16 – 1)	15 (16 – 1)	14 (16 – 2)	16
	Total:	198	190	27	27	14	16

4 Results

4.1 RQ1: To what extent is education based on a professional meaningful for students?

The first item of the survey was: "With a professional practice as a basis for the curriculum I see the usefulness of the curriculum" of Enq_2 (N = 190). The response shows that prior to the unit most students (+ and + + added: 53 + 11 = 64%) expected to see the usefulness (Table 5). 123 respondents out of 190 gave an optional explanation. Of the explanations 103 were positive, 12 negative and 8 unclear. Examples of positive explanations are as follows:

- "Because I often can't see the usefulness of certain subjects."
- "You see at least its usefulness [curriculum] in it."
- "It [the professional practice] gives better understanding of your ultimate goal."

Table 5

Percentage score of 8 items from Enq_2 (N = 190), Enq_3 (N = 27) and Enq_4 (N = 27). In this table the abbreviations "-", "-", "0", "+" and "+ +" refer to "Strongly disagree" to "strongly agree".

	Item	Aspect meaningfulness	Question-naire	--	-	0	+	++
1	With a professional practice as a basis for the curriculum I see the usefulness of the curriculum.	useful	Enq_2	2	5	30	53	11
			Enq_3	0	0	26	67	7
			Enq_4	0	15	11	63	11
2	If the techniques of the curriculum can be used by professionals, it motivates me to learn these techniques.	motivation	Enq_2	2	8	24	51	15
			Enq_3	4	4	37	48	7
			Enq_4	4	7	33	52	4
3	I think I might be able to use the technique to determine a regression line in more school subjects.	apply	Enq_2	4	4	19	56	16
			Enq_3	7	7	15	52	19
			Enq_4	4	7	4	70	15
4	I think that the technique of correlation and regression is needed within each of the natural sciences.	apply	Enq_2	2	7	46	37	8
			Enq_3	0	15	15	59	11
			Enq_4	4	11	19	56	11
5	I find it interesting when I learn something about another subject, such as biology or physics, when I follow a statistics unit.	connection	Enq_2	14	24	29	27	6
			Enq_3	4	22	26	37	11
			Enq_4	7	11	37	37	7
6	I have gained more insight into how statistics can be used by a sports physiologist / official of Ministry Water management	connection	Enq_3	0	11	11	67	11
			Enq_4	0	4	15	74	7
7	I think determining a threshold point in the practice of a physiologist is done the same way as in the lessons.	authenticity	Enq_3	0	37	33	26	4
8	I think that the determination of the date on which action should be taken to heighten a dyke is in reality the same as in the lessons.	authenticity	Enq_4	15	19	37	26	4

One negative coded response was: “No, I think it is not effective.” Another student also made explicitly clear that coherence was important to her: “I would like to see professional practices as the basis for curriculum, because now I can easily make connections between different subjects.” We concluded that the results of Enq_2 indicates that most students, prior to the unit, see the point of using professional practices. Because this also was an outcome of Enq_1, we had sufficient reasons to develop the instructional unit.

From Enq_3 it appeared that students who went through the instructional unit, considered the teaching learning strategy also as useful after the completion of H1 of the instructional unit. These students (67% + 7% = 74%) were even more positive (+ and + + added). Even after H2, which taught the mathematical foundation of correlation and regression, the majority (74%) of the students gave still a positive response (Enq_4). A binomial sign test to assess responses per student to Enq_3 and Enq_4 yielded $p = 0.60$. So we do not observe a change which is statistically significant. This indicates that the opinions of students did not change significantly, not even after the mathematical-theoretical approach of H2 which had the risk that students would not directly see its usefulness.

The answers to the items 2 to 8 show a positive inclination. Only the item about the authenticity of the lesson activities related to professional practice (items 7 and 8) is scored in a wider range. This finding is consistent with our observation that the students were aware of the fact that the professional practices are educationalized. A sports physiologist does not limited herself/himself to just the heart rate, as it is presented in the instructional unit, but she/he in preparing an advice also involves other variables.

4.2 Research Question 2: Relations between the natural sciences and professional practices

4.2.1 Comparison of the connections made by students

In this section we discuss the extent to which students have learnt to make connections. The first task of the instructional unit and of the posttest was:

> Sports physiologists often determine the threshold point of the heartbeat of their clients.
> a. What do we mean in this case, with threshold?
> b. Why is it important to measure this threshold point?

Comparison of the responses to this task shows that the students after the completion of the instructional unit, often saw a link between science disciplines and professional practice more clearly than at the start of the unit (Table 6). Only the relationship between mathematics and the professional practice (MP) shows just a negligible increase. At the beginning, the relationship between mathematics and statistics (MS) is not detected, and after working through the unit it clearly was. In this task, we assumed that the students would develop a mathematical model for the measurements (MS). Seven students did so. Three of them made a drawing such as Jan's in Figure 2.

Table 6
Number of students who submitted links between mathematics (M), professional practices (P), Statistics (S) and the natural sciences (N), for the preliminary task of the instructional unit and during the posttest (N = 16).

Code	Introductional task	Posttest
MS	0	7
MP	2	3
PS	5	13
PN	4	13
SN	1	12

Three students explicitly linked the model shown in Figure 2 with the advice that a sports physiologist can give to her/his client (MP). Most students (13) associated statistical techniques (e.g. regression line for the linear portion) with the physiological advice (PS). These thirteen involved the scientific terms aerobic and anaerobic combustion with their advice (PN). They wrote that if training takes place too often in the anaerobic section (above the threshold, "omslagpunt" in Figure 2), the threshold point goes down, in which case acidification starts earlier. So an athlete should train slightly below the threshold point. Twelve students linked their physiological advice with the data obtained during the testing of clients (SN).

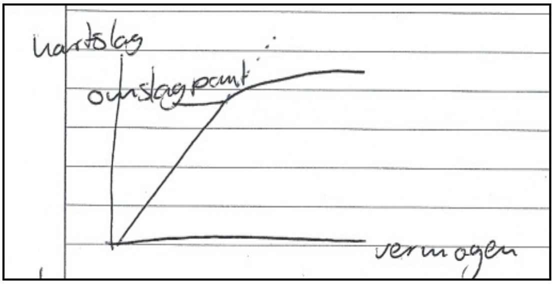

Figure 2. Jan's illustration of the model for the determination of the threshold point between aerobic and anaerobic metabolism.

4.2.2 The extent to which students made connections in the posttest

In Table 7 the post tests are the average student scores per order. The students were sufficiently able to apply concepts from the instructional unit: the average score was 64%.

Table 7
Connections MS, MP, PS, PN and SN made in the tasks of the posttest and scores of the final group (N =16).

Task	MS	MP	PS	PN	SN	Max. score	Mean Score (SD)	% Score
Task 1	x	x	x	x	x	2	1,1 (0,6)	55
Task 2	x					9	6,3 (1,8)	69
Task 3	x		x			16	11,0 (3,3)	69
Task 4	x		x	x	x	11	6,1 (2,3)	55

Task 2 was about the relationship between mathematics and statistics. The task assessed what students had learnt in H2 about the mathematical background of correlation and regression. In this task, we asked the students to describe the sequence of steps by which the least squares method determines the coefficients of the regression line. Eight students did this perfectly, four made a small mistake, and no one had a completely wrong answer. Also we asked them to explain the least squares method. This was the trickiest part of task 2 (mean score 53%). Furthermore we presented in task 2 to the students in a random order steps from the method of finding a normal equation. The students had learned to draw a scatterplot of points (d_i, H_i) with d for day and H for deviation and represent a regression line by

$\hat{H} = a \cdot d + b$, with a and b as regression coefficients.

Further, they had learned to calculate the distance from a point to the regression line (residual) with: $a \cdot d_i + b - H_i$. Using the least squares method the students were able to find the best fit for the regression line by minimizing the sum of residual squares:

$$S = \sum_{1}^{n} (a \cdot d_i + b - H_i)^2$$

Derivations to a and b gave them a system of two normal equations

$$\begin{cases} a\sum d_i^2 + b\sum d_i = \sum d_i H_i \\ a\sum d_i + n \cdot b = \sum H_i \end{cases}$$

The students had learned to calculate all the sigma's and were able to solve the system.

In task 2 we asked them to write the steps for finding a normal equation in the correct order and to explain each step. Most students were able to write the steps in the proper order, but their explanations sometimes were rather weak (See Figure 3).

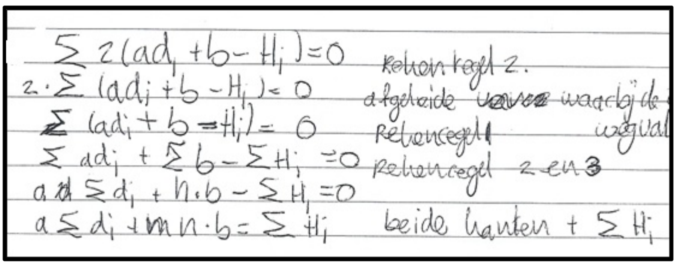

Figure 3. Part of a student's response on task 2 of the posttest.

In task 3 the students were given 44 measurements of dyke-deformations. To test whether students could explain the MS-relation, they had to calculate the time where raising the dyke could no longer be avoided. Apart of some sloppiness all students could calculate the correlation. Except for one student, all students could determine the formula for the regression line.

In task 3, the relation with the professional practice was very important (PS). Because, when only the regression line is used for the calculation of the day when the embankment is to be incremented, then it is too late: there are also points below the regression line. The assignment required that students would take a safety margin. Five did not. A student used Excel for this margin (see Figure 4).

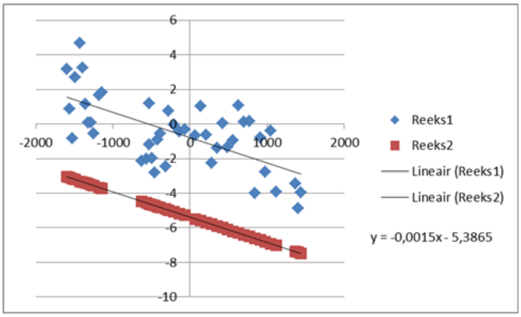

Figure 4. Excel output by one of the students. The upper line is the regression line. The lower line is constructed by subtracting three times the SD of the residuals from the regression values (According to a discussion about the rules of thump for a normal distribution).

The other students adapted a well-known formula. By applying statistical know how to a mathematical model (regression line), the students showed that they also saw the connection between mathematics and statistics (MS). The fourth task coincided with the chapter on the calibration of measuring instruments. The students were given a new context in which, for nine samples of blood serum, the concentration of iron and the corresponding absorption were given. Except for one student, they all could determine the calibration line (regression line). To emphasize the relation with the natural sciences as these are taught at school, points were deducted if incorrect letters for the variables were used. Also, the students were asked to determine the iron concentration of a patient of which the measured absorbance of a blood sample was given. Thirteen students did this correctly. The other three had problems in connecting the scientific terms to the statistical concept of regression line (SN and PN).

In task 4 students were also facing a new context for them: atomic absorption spectrometry. The students were given the data (the amount of added silver and the absorption) of nine blood samples. Most students (15) were able to determine the regression line, but only two students were able to perform the steps necessary to calculate the concentration of the original sample. In this task the students were required to combine at the same time mathematical, statistical and natural sciences with professional techniques (MS, PS, PN and SN).

5 Conclusion and discussion

In answer to the first Research Question, we conclude that many students experience an instructional unit based on professional practices as meaningful. Both surveys, one prior to the experiments and the other after the completion of the unit, indicates that students widely appreciated such an educational strategy. They indicated that they see the need for a curriculum that is based on professional practices, that such an instructional strategy motivates them, and that they think to be able to apply what they had learnt also in other disciplines. The opinions of the students were more divided about the authenticity of the teaching material, which is well understandable because the professional practices were first educationalized, especially by taking away some of the complexity.

In the study presented here, we argued that coherence is not enough to understand relationships. When making connections, it is also important that students master the concepts and are able to apply them. In answer to our second Research Question, we conclude that the students after the completion of the unit clearly made more connections between mathematics, statistics, natural science and professional practice than at the start of the module. The relationship between mathematics and the professional practice is recognized in the students' answers not so strongly as the other connections. This is understandable because in this instructional unit the connection between mathematics and the professional practice (MP) was in accordance with the design, which presents the connection primarily via statistics; there was no a direct connection designed between mathematics and the professional practice. We assume that our analysis model can be easily adapted for connections between other school subjects and professional practices. Also, we conclude that our instructional unit helped the students to apply the new concepts in other disciplines. The analysis of the posttest with tasks in which various connections were involved, shows that students were able to apply the learned concepts. They were also able to make required connections. In reaching this conclusion, we must remember that the measured inter-rater reliability is acceptable although still quite low (0.61). We attribute this to the fact that for each item multiple codes were possible. Furthermore, the boundaries between the five areas are not always clear (2.3). Also, note that the other coder/researcher had only the transcripts and had not been involved in acquiring the data.

The answer to the main question is that we designed an example of an instructional unit based on professional practices which was not only meaningful for most of the students, but also after the completion of the unit, caused a relevant increase in the number of times that students made connections between mathematics, statistics, natural science and some professional practices. Although the literature (e.g., Berlin & White, 2011) indicates that it is difficult to make the relationship between science and contexts clear to the students, the design on the basis of professional practices, with a focus on meaning from a student's perspective, seems a promising direction to take. Further research is needed to investigate whether and when professional practices in other situations can help to improve curricular coherence. This also asks for comparative research. Also, further research along the lines of Dam, Janssen and van Driel (2012) is needed in order to implement education based on authentic professional practices and to better understand the role of the teacher.

References

AAAS (1989). *Project 2061: Science for all Americans*. Washington, DC: American Association for the Advancement of Science.

Apotheker, J., Bulte, A., Kleijn, E. de, Van Koten, G., Meinema, H., & Seller, F. (2010). *Scheikunde in de dynamiek van de toekomst over de ontwikkeling van scheikunde in de school van de 21^e^ eeuw.* Enschede: SLO.

Barab, S. A., & Squire, K. D. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences, 13*, 1–14.

Bennett, J., Lubben, F., & Hogarth, S. (2007). Bringing science to life: a synthesis of the research evidence on the effects of context-based and STS approaches to science teaching. *Science Education, 91*, 347–370.

Berlin, D. F., & Lee. H. (2005). Integrating science and mathematics education: Historical analysis. *School Science and Mathematics, 1*, 15–24.

Berlin, D. F., & White, A. L. (1992). Report from the NSF/SSMA Wingspread Conference: A network for Integrated Science and Mathematics Teaching and Learning. *School Science and Mathematics, 92*, 340-342.

Berlin, D. F., & White, A. L. (2012). A longitudinal look at attitudes and perceptions related to the integration of mathematics, science, and technology education. *School Science and Mathematics, 1,* 20–30.

Boaler, J. (1993). Encouraging the transfer of 'school' mathematics to the 'real world' through the integration of process and content, context and culture. *Educational Studies in Mathematics, 25*, 341–373.

Boersma, K. Th., Van Graft, M., Harteveld, A., De Hullu, E., De Knecht-van Eekelen, A., … Mazereeuw, M., (2007*). Leerlijn biologie van 4 tot 18 jaar. Uitwerking van de concept-context benadering tot doelstellingen voor het biologieonderwijs.* Utrecht, the Netherlands: NIBI.

Boersma, K., Bulte, A., Krüger, J., Pieters, M., & Seller, F. (2010). *Samenhang in het natuurwetenschappelijk onderwijs voor havo en vwo.* Utrecht, the Netherlands: Stichting Innovatie van Onderwijs in Bètawetenschappen en Technologie (IOBT).

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Clarke, D. (1988). *The mathematics curriculum and teaching program professional development package: Assessment alternatives in mathematics*. Carlton, Vic.: Curriculum Corporation.

Claxton, G. (1991). *Educating the enquiring mind: The challenge for school science.* New York: Harvester Wheatsheaf.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46.

cTWO (2007). *Rijk aan betekenis. Visie op vernieuwd wiskundeonderwijs.* Utrecht, the Netherlands: Commissie Toekomst Wiskunde Onderwijs.

Dam, M., Janssen, F. J. J. M., & Van Driel, J.H. (2012). Concept-contextonderwijs leren ontwerpen en uitvoeren – een onderwijsvernieuwing praktisch bruikbaar maken voor docenten. *Pedagogische Studiën*, *90*(2), 63–77.

Davison, D. M., Miller, K. W., & Metheny, D. L. (1995). What does integration of science and mathematics really mean? *School Science and Mathematics, 95*(5), 226–230.

Dierdorp. A., Bakker. A., Eijkelhof. H. M. C., & Van Maanen, J. A. (2011). Authentic practices as contexts for learning to draw inferences beyond correlated data. *Mathematical Thinking and Learning, 13*, 132–151.

Eccles, J. S., Midgley, C., Wigfield, A., Miller Buchanan, C., Reuman, D., Flanagan, C., & Iver, D. M. (1993). Development during adolescence. The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist, 48*(2), 90–101.

Erickson, T. (2002). *Technology, statistics, and subtleties of measurement: Bridging the gap between science and mathematics.* Paper presented at the 6th International Conference on Teaching Statistics (ICOTS-6), Cape Town: South Africa.

Frykholm, J., & Glasson, G. (2005). Connecting science and mathematics instruction: Pedagogical context knowledge for teachers. *School Science and Mathematics, 105*(3), 127–141.

Furner, J. M., & Kumar, D. D. (2007). The mathematics and science integration argument: A stand for teacher education. *Eurasia Journal of Mathematics, Science & Technology Education*, *3*(3), 185–189.

Galbraith, P., & Stillmann, G. (2006). A framework for identifying student blockages during transitions in the modeling. *Zentralblatt für Didaktik der Mathematik, 38,* 143–162.

Gattuso, L. (2006). Statistics and mathematics. Is it possible to create fruitful links? In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil.*

Gellish, R. L., Goslin, B. R., Olson, R. E., McDonald, A., Russi, G. D., & Moudgil,

G. V. (2007). Longitudinal modeling of the relationship between age and maximal heart rate. *The American College of Sports Medicine*, *39*(5), 822–829.

Geraedts, C. L., Boersma, K. T., Huijs, H. A. M., & Eijkelhof, H. M. C. (2001). *Ruimte voor sonate.* Delft, the Netherlands: Stichting Axis.

Giessen, C. van de, Hengeveld, T., Van der Kooij, H., Rijke, K., & Sonneveld, W. (2007). *Eindverslag van Werkgroep Afstemming Wiskunde-Natuurkunde.* Utrecht: cTWO and NiNa.

Hurley, M. (2001). Reviewing integrated science and mathematics: The search for evidence and definitions from new perspectives. *School Science and Mathematics, 101*, 259–268.

Lee, H.S, & Songer, N.B. (2003). Making authentic science accessible to students. *International Journal of Science Education, 25*, 923–948.

Lederman, N. G., & Niess, M. L. (1997). Integrated, interdisciplinary, or thematic instruction? Is this a question or is it questionable semantics? *School Science and Mathematics, 97*(2), 57–58.

Mooldijk, A., & Sonneveld, W. (2010). Coherent education in mathematics and physics: The theme of proportionality in mathematics and physics. In N. Valadines (Eds.), *Trend in Science and Mathematics Education (TiSME)* (pp. 43–50). Cyprus: Cassoulides.

Moore, D. S., & Cobb, G. W. (2000). Statistics and mathematics: tension and cooperation. *The American Mathematical Monthly, 107*, 615–630.

NCTM (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

NRC, National Research Council (1996). *National science education standards*. Washington, DC: National Academy Press.

Newmann, F. M., Smith, B., Allensworth, E., & Bryk, A. S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis, 23*, 297–321.

Nieveen, N., Handelzalts, A., & Van Eekelen, I. (2011). Naar curriculaire samenhang in de onderbouw van het voortgezet onderwijs. *Pedagogische Studiën, 88*, 249–265.

NiNa, Commissie Vernieuwing Natuurkundeonderwijs havo/vwo (2010). Nieuwe natuurkunde, advies-examenprogramma's voor havo en vwo. Amsterdam: Nederlandse Natuurkundige Vereniging.

Prins, G. T. (2010). *Teaching and learning of modeling in chemistry education*. Utrecht, the Netherlands: CD-Bèta Press.

Prins, G., Bulte, A. M., Pilot, A., & Van Driel, J. (2008). Selection of authentic modelling practices as contexts for chemistry education. *International Journal of Science Education, 30*, 1867–1890.

Rossman, A., Chance, B., & Medina, E. (2006). Some key comparisons between statistics and mathematics, and why teachers should care. In G. F. Burrill (Ed.), *Thinking and reasoning with data and chance: Sixty-eighth annual yearbook of the National Council of Teachers of Mathematics* (pp. 323–333). Reston, VA: NCTM.

Rudduck, J. H. S., & Wallace, G. (1994) Coherence and students experiences of learning in secondary schools. *Cambridge Journal of Education, 24*, 197-211.

Scott, P., Mortimer, E., & Ametller, J. (2011). Pedagogical link-making: A fundamental aspect of teaching and learning scientific conceptual knowledge. *Studies in Science Education, 47*, 3–36.

Stuurgroep-NLT. (2007). *Contouren van een nieuw bètavak*. Enschede: Stuurgroep-NLT.

Sunal, D., & Furner, J. M., (1995, March). *Teaching mathematics and science not mathematics or science*. Paper presented at the NASA Teacher Enrichment Program Mid-Year Workshop at the Marshall Space Flight Center, Huntsville: Alabama.

Van den Akker, J. J. H. (2009). Curriculum design research. In T. Plomp & N. Nieveen (Eds.), *An introduction to educational design research* (pp. 37–51). Enschede: SLO.

Van den Akker, J. J. H., Gravemeijer, K. P. E., McKenney, S., & Nieveen, N. (2006). *Educational design research*. London: Routledge, Taylor & Francis.

Van Oers, B. (1987). *Activiteit en begrip. Proeve van handelings-psychologische didactiek.* Amsterdam, the Netherlands: VU Uitgeverij.

Volman, M. (2011). *Kennis van betekenis. Betrokkenheid als kwaliteit van leerprocessen en Leerresultaten.* Oratie Universiteit van Amsterdam.

Westbroek, H. B. (2005). *Characteristics of meaningful chemistry education - The case of water quality.* Utrecht, the Netherlands: CD-Bèta Press.

Westra, R. (2008). *Learning and teaching ecosystem behaviour in secondary education. Systems thinking and modelling in authentic practices*. Utrecht, the Netherlands: CD-Bèta Press.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*, 223–265.

Chapter 6

CONCLUSION AND DISCUSSION

1 Introduction

The purpose of this research project was to investigate how to support upper-secondary school students' learning and understanding of statistical correlation and regression models in such a way that they can apply the techniques in new situations and experience coherence between mathematics and the natural sciences. In line with this purpose our main research question was:

> *What are characteristics of a valid and effective teaching and learning strategy to teach students about correlation and regression in such a way that they experience coherence between mathematics and the natural sciences?*

In order to collect data for answering this question we designed a teaching and learning strategy for upper-secondary school students (Grade 11 and 12, aged 16-18 years) and investigated trials at three similar Dutch schools (pre-university track: VWO). Doing so, we wanted to contribute to a school subject that was a new one at that time in the Netherlands: NLT (Nature, Life and Technology; see Eijkelhof & Kruger, 2009). Three teachers volunteered to participate in this study. The background of the first teacher (also the principal researcher) is mathematics and biology, the second teacher is a mathematician and the third teacher's background is physics and mathematics. We tested the evolving teaching and learning strategy in four studies.

Our first study (Chapter 2) sets out to investigate how to support students' informal inferential reasoning when they learn about correlation and regression in the context of use by professionals. Furthermore, we focus the next two studies each on a specific aspect necessary for reasoning about informal inferential reasoning, such as variability and measurement (Chapter 3) and students' ability to reference back and forth ("shuttling") between the context and statistics when engaged in sampling tasks (Chapter 4). From here, in relation to informal inferential reasoning we zoom out in our fourth study (Chapter 5) to study more broadly to what extent students experience the coherence between mathematics and the natural sciences, with statistical concepts and modelling techniques as the bridging tools. In sections 6.2 to 6.5 we summarize the main findings of these studies and discuss the contribution to the knowledge base with respect to the support of students' learning and understanding of correlation and regression. In the last sections (6.6 to 6.11) we focus on considerations, limitations and future research suggestions which we could not address in Chapters 2 to 5.

2 How to support students to make inferences beyond correlated data? (Chapter 2)

The research question for our first study was: *How does a teaching and learning strategy based on authentic practices support students in making statistical inferences about authentic problems with the help of correlation and linear regression?* The focus of this study was the support of students' ability to draw informal inferences. To analyse this ability we used the students' written work. In a matrix we compared the observed responses by students with the hypothesized well informed responses. In this manner we investigated to what extent the students were able to draw informal inferences.

The results indicated that our teaching and learning strategy inspired by authentic professional practices had advantages and also produced new challenges. Students generally saw the practical value of correlation and regression and many of them were able to draw realistic inferences. We concluded that a teaching and learning strategy with statistical core ideas embedded in authentic professional practices can help students to make more and better informal statistical inferences. Another advantage of our strategy was that it helped students to understand that practitioners in many fields use these techniques in their daily professional work and to realize that they could use these techniques in many situations themselves at school or in a future academic study.

However, in the more complex context of monitoring heights of dikes students struggled to see the need to account for variability. It is a challenge for the designer to appropriately modify the level of complexity of the authentic problems and tools for student use. It also appeared necessary to provide the teacher with enough information, because the teacher's own knowledge of the context is important in supporting the reasoning of students. For example, inferences about when to heighten dikes depend on more variables than just the regression results on dike deformations. It is necessary to indicate clearly that these measurements are only part of the total issue.

3 Supporting students to reason about variability (Chapter 3)

In the second study we asked: *How do secondary students consider variability within measurement activities based on authentic professional practices?* In this study we investigated how students in Grade 12, aged 17-18, consider variability using context-rich measurement tasks. To investigate whether students demonstrated key components of reasoning about variability: noticing and acknowledging variability, measuring and modelling variability, explaining variability and using investigative strategies, we adapted a framework from Wild and Pfannkuch, (1999).

We concluded that the measurement activities supported students to reason about the four components of variability as described in our framework and that the use of measurement tasks based on authentic professional practices is a promising design strategy to scaffold students' statistical reasoning about variability.

4 Supporting students in developing the concepts that underlie sampling (Chapter 4)

In the third study we asked: *What is the potential of tasks based on authentic professional practices to support students in developing concepts that underlie sampling in such a way that they can shuttle between contextual and statistical spheres?* This study focused on the aim that students should (a) see the purpose of tasks and the utility of concepts and (b) be able to mathematize authentic problems and apply concepts in context. We designed two activities based on the authentic practice of a physiotherapist and carried out two case studies to test the potential of these tasks to support students' concepts that underlie sampling and to investigate what teachers should do to exploit this potential. An additional persistent challenge is to support students to shuttle between contextual and statistical spheres. Analysis of video recordings focused on five concepts that underlie sampling (inspired by Pfannkuch, 2008), which are important for students' statistical reasoning about sampling: sample size, random process, distribution, intuitive confidence interval and relationship between sample and population.

The analyses showed that in order to support students' shuttling between contextual and statistical spheres, realistic sampling tasks inspired by authentic professional practices can be helpful. This suggests an advantage over atomistic approaches to provide students with statistical concepts one by one (cf. Bakker & Derry, 2011).

We concluded that it seems possible, not only in science education (e.g., Westbroek, Klaassen, Bulte, & Pilot, 2010) but also in statistics education, to use authentic problems from professional practices to design tasks that are purposeful from a student perspective and lead students to see the utility of what they learn. This may help students to apply what they have learned in new situations. However, it is not self-evident that students develop rich conceptual understanding from authentic tasks because designers seem to have less control about what students learn conceptually. Yet such conceptual understanding seems necessary for students to link contextual and statistical considerations. Wild and Pfannkuch (1999) studied professional statisticians' reasoning and noted that these professionals are involved in a "continual shuttling backwards and forwards between thinking in the context sphere and the statistical sphere" (1999, p. 228). When students are involved with contextual authentic questions data analysis pushes them into the statistical sphere where students can develop statistical concepts. However, the analysing results propels them back to the context to find a meaning for those results. Such shuttling helps the students to see the tasks' purpose and the concepts' utility and contributes to their ability to mathematize authentic problems and apply concepts in meaningful contexts.

5 Coherence between school subjects (Chapter 5)

The research question in our final study was: *To what extent can professional practices serve as meaningful contexts to show students connections between mathematics, statistics, science and professional practices?* This study was looking for ways to promote coherence in the learning outcomes: how can students integrate knowledge from mathematics and the natural sciences when they practice informal inferential reasoning about solutions of authentic problems and how do they value such an approach? Such experience of coherence is important for students to appreciate contexts as meaningful to learn statistical concepts and shuttle between contextual and statistical spheres (see 6.4). In particular Ben-Zvi and Garfield (2010) claim, with regard to informal inferential reasoning, that statistics can be viewed as an intermediary discipline that connects mathematics and science. We developed an analytical model to investigate students' ability to integrate mathematics, statistics, science and professional practices and analysed students' responses to questionnaires and tasks before and after the lessons.

This analysis showed that our designed instructional unit is found meaningful by most students. The unit motivated them and they believed to be able to apply the learned concepts also in other school subjects. We also argued that meaningfulness alone is not sufficient for students to make connections between school subjects. For making connections, it is also important that students master the concepts and are able to apply them. We suggest that it is important to monitor more often whether students are able to make connections and we assume that our method of analysing can be adapted for measuring connections that students make between other school subjects and other contexts. Here we note that monitoring on the basis of this model is difficult because the boundaries of the school subjects are not always very clear.

6 Answer to the main research question

Based on the four studies we conclude that our strategy to teach students about correlation and regression seems valid and effective. It seems valid because we designed the strategy in line with prevailing epistemological ideas of the involved school subjects (e.g., mathematics: calculate standard deviation, statistics: produce a formula for the regression line, biology: aerobic respiration, geometry: reasons for subsidence, physics: operation of a thermometer). It seems effective because the involved students learned to solve real-world problems by correctly using correlation and regression models. They also appeared to understand the concepts and process of modelling and were able to combine mathematical and statistical techniques with concepts of the natural sciences when solving real-world problems. In Table 1, we give our design characteristics for a teaching and learning strategy and refer to evidence in our studies.

Table 1
Characteristics of our teaching and learning strategy and research questions in corresponding Chapters 2.

	Characteristics of the teaching and learning strategy	RQ1 (Ch 2)	RQ2 (Ch 3)	RQ3 (Ch 4)	RQ4 (Ch 5)
1	The strategy is inspired by reality, such as by authentic professional practices.	x	x	x	x
2	The strategy's level of the authentic problems' complexity and tools is modified for students' use.	x	x	x	x
3	The strategy is focused on informal inferential reasoning.	x	x	x	x
4	The strategy engages students to learn about variability and measurement.	x	x		
5	The strategy engages students with their own measurements.	x	x		
6	The strategy provokes students to shuttle between contextual and statistical spheres by means of tasks with a clear purpose and teacher directions.	x		x	
7	The strategy provides students with realistic sampling tasks.	x	x	x	x
8	The strategy is multi-disciplinary.	x			x
9	The strategy activates students' feelings of coherence between the involved school subjects.	x			x

Our study built upon the work of researchers (Edelson & Reiser, 2006; Prins, 2010; Westbroek, 2005; Westra, 2008) who have studied how to inspire science education by authentic professional practices. The choices we made were that we focused on statistics and informal inferential reasoning and investigated senior secondary school students. In their research on informal inferential reasoning Pratt, Ainley, Kent, Levinson, Yogui and Kapadia (2011) elaborated on the element of risk when students learn about statistics. With regard to informal inferential reasoning, we underestimated how complex risk is when operating within the context of dyke maintenance. By educationalizing this authentic professional practice we probably over-simplified the practice. Issues such as the acceptability of flooding risks, the considerations of safety margins already included in critical values defined by the Ministry of Transport and for example what variability in sea water levels we can expect, are still open to elaboration in future research.

Another contribution to the knowledge of educational strategies inspired by authentic professional practices is that we found evidence for a possibility to overcome the problems with respect to the planning paradox. A conclusion of Ainley, Pratt and Hansen (2006) was that engaging context-rich tasks are often less focused on mathematics. In our study we argue that tasks about statistics based on authentic professional practices can have the potential of being engaging for students while at the same time developing concepts relevant for their school subjects.

In line with Brandom's (2000, p. 48) inferentialist perspective on concepts, conceptual understanding can be characterized as follows: "To grasp or understand a concept is to have practical mastery over the inferences it is involved in—to know, in the practical sense of being able to distinguish, what follows from the applicability of a concept and what it follows from." From this perspective conceptual understanding includes the ability to shuttle between contextual and statistical spheres. Our study showed that this shuttling is challenging, or at least unfamiliar to students in school mathematics. It seems that reasoning within either the statistical or contextual world is easier than making connections. Developing tasks that have a clear purpose and lead to concepts with utility is not sufficient: the teacher is needed to lead the process of shuttling.

7 Design research

The purpose of the design research (see Chapter 1: 7.1) of this thesis was to test and improve our conjectures formulated in the Hypothetical Learning Trajectory, but also to develop an understanding of how the strategy works. In terms of Nieveen, McKenney and Van den Akker (2006) our study is not a validation study but a development study. In the first two columns of Table 2 we summarize the educational engineering research cycle based on Nieveen, et al., (2006). In the last column we mention where to find the items in this dissertation.

Table 2
Educational research cycle based on Nieveen et al. (2006).

	Development study	Evidence in
Design aim	To solve educational problems	Chapter 1
Quality focus of design	Practicality of intervention	Chapters 2, 3, 4 and 5
Knowledge claim	Broadly applicable design principles	Section 6.6
Methodological emphasis	Iterative development with formative evaluation in various user settings	Chapters 2, 3, 4 and 5
Practical contribution	Implemented interventions in several contexts/classrooms	Chapters 2, 3, 4 and 5

A design research approach is a constant struggle to develop both educational materials and a theory. The design of the educational materials is only a side product, but to investigate how it works it has to be developed first. When the iterative process reveals how it works or should work, adaptations are necessary. Thus, when testing our theory we implicitly tested our educational materials.

In future research it may be easier to investigate our theory because the educational materials will need less attention, so the focus can rather be on theory development regarding our teaching and learning strategy. In the next sections we give suggestions for future research.

8 The practicality of our study

The possible impact of our study for educational practice is multiple. Its scientific findings are directly applicable to educational practice. The practicality (Nieveen, McKenney, & Van den Akker, 2006) of our study implies an effective intervention: we have developed an instructional unit and a research based student test that are realistically usable in the setting of secondary schools. Also, we developed a set of design characteristics (see Table 1) as criteria for designers of similar teaching and learning strategies. Our unit helps students to understand and reason about correlation and regression and provokes them to reason about topics from the natural sciences. To understand and apply the statistics, the students had to use the mathematical knowledge and skills they had already acquired and also to expand their mathematical knowledge (see Chapter 5). In addition, students needed knowledge from the natural sciences. For example, in the unit's chapter about the sport physiologist (see Chapter 3) the students needed to understand the difference between aerobic and anaerobic respiration. During one of the macrocycles two students who also took biology as a school subject volunteered to give a detailed presentation about this difference. This presentation is an example of how the instructional unit may provoke students to be engaged and involved in several knowledge domains.

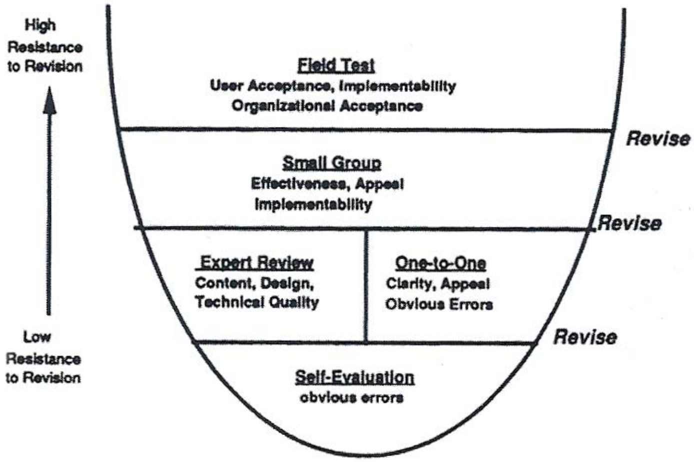

Figure 1. Layers of formative evaluation (Tessmer, 1993).

As a limitation for the practicality of this study one should take into account that we describe research results from three schools and focus on two schools and two teachers. The scale of our implementation seems small, due to our pragmatic approach. In scaling up from a small target group to a larger audience, we followed the layers of formative evaluation of Tessmer (1993) (see Figure 1).

After our self-evaluation we had expert reviews at several moments and a one-to-one trial with two students before testing the materials in a small group. Only after, did we have several cycles in complete classes. The reasons for not involving more teachers is that the researcher had an intense research focus and that the teaching inspired by authentic practices requires an extensive preparation of the teacher.

Future research in relation to the practicality is recommended. We suggest that an extensive accompanying instruction set for teachers should be developed and tested and that the teaching and learning strategy should be tested in more schools.

9 Researcher, designer and teacher integrated in one person

The main researcher of this thesis had a complicated task. He was supposed to have an "inner triangulated" view, because he was at the same time both researcher and designer, as well as teacher in some studies. Especially at the start of the project it was a challenge to have a research focus rather than a teaching focus. It was necessary to make a mind shift from "How can I teach the student …" to "How can I investigate how the student …" In every macrocycle that followed, this shift became easier. The researcher initially acted as teacher to investigate whether the ideas worked, but in all cases another teacher was deployed in the next macrocycle. An advantage of the complicated role of the researcher was that lines were short and that the researcher had an optimal focus during the next macrocycle when the other teacher was teaching.

A limitation of the approach to integrate the roles of researcher, designer and teacher in one person is that his influence becomes large and that results might depend in an implicit manner on the characteristics of the teacher-researcher (e.g. the fact that he was enthusiastic about the project and that he was better informed than an average teacher could be, even after considerable instruction and training.)

Future research with more teachers involved should answer the question of whether our strategy works for other teachers. Teacher preparation seems to be important because knowledge of authentic professional practices is not trivial for each teacher. The teachers involved in our study already had much knowledge of the authentic practices, but spent a lot of time in preparing the lessons. As mentioned before, we suggest that an extensive accompanying instruction set is required.

10 Educational change

The research presented in this thesis has connections with educational change planned by the Dutch government. Within this context of curricular innovation, we wanted to design a new way of teaching and learning statistics and simultaneously show students coherence between school subjects. The project was funded through DUDOC, which was established in 2007 by the Science and Technology Platform with the support of the Ministry of Education (Knippels, Goedhart, & Plomp, 2008). The DUDOC board consisted of representatives of the Dutch curriculum innovation committees of mathematics and science subjects and was led by an independent Chairman. DUDOC gave twenty teachers the possibility to do educational research with a PhD as the intended outcome. DUDOC specifically targeted teachers because the Ministry wanted to bridge the gap between the academic world and school practice. Educational change should receive an impulse from these twenty teacher researchers. However, it is not easy to change the daily organization and culture of schools. Fullan (2007, pp. 291) claimed:

> It will be necessary to build an instructional system that is based on personalization (connecting to the unique needs of each student) and precision (connecting in a way that is geared specifically to the student's needs in a timely fashion).

Fullan also stressed that the change must not be a fast one. Our project fits this statement well; we made several small and in-depth contributions and expect more to be needed before change at a broader level can be observed.

We designed an instructional unit based on authentic professional practices to motivate and show students why they have to learn the concepts, through not only offering exercises, but also making the students reason about the concepts to provide better understanding. Several students used the concepts of correlation and regression learned in our unit in tasks from other school subjects. Nowadays, upper secondary school students in the Netherlands have to do a small (i.e. 80 hours) research project. We experienced that six students of the last investigated group (N = 16) used the correlation and regression techniques from our instructional unit for their research. In this timely fashion our instructional unit met the needs of several students.

With respect to educational change, a limitation of our study is that the conclusions formulated in this Chapter 6 show how our intervention "works" for one particular case. One may wonder whether the conclusions about our rather specific strategy can be generalized, for example to other parts of mathematics and also to other school systems and other countries. Changing societies may need other approaches. Our new way of teaching correlation and regression in secondary school was validated by triangulation involving literature, expert reviews, students' opinions, teacher's conduct and observations made by the researchers. For example, we asked a physiotherapist to review our instructional unit chapter about his profession. Also, we discussed the instructional unit with expert statistician Prof.dr. Richard Gill. He explained that in real forecasting the margin around the regression line is not linear but hyperbolic, an theorem discussed and proven in Appendix A. As a simplification of the authentic practice we decided a linear margin would be acceptable as a first step. As a next step, the teacher can discuss the fact that safety margins for the far future have to be larger than for the near future. We expect students to know this from weather predictions, but the proof of the hyperbolic shape transcends the secondary school curriculum.

Future research may concern the practical issue of the length (19 lessons) of our instructional unit. In view of the Dutch mathematics curriculum, it was too long as a chapter in a mathematics textbook and therefore we decided to design it as an instructional NLT unit. At the time, NLT was a rather new school subject. Another advantage of housing the units within NLT is that we were able to involve several subjects from the natural sciences, since that is precisely the key philosophy of NLT. Now, we could ask students a lot of "Why," "What" and "How" questions. Initially, many students did not like this because they were not used to these questions, but afterwards they claimed that discussions (provoked by why, what and how) were very informative. The instructional unit has reached a final version, but still contains tasks which may require further investigations. For use in mathematics classes we would like to know whether an adapted, shorter version would also work.

Another direction of future research could be to find out whether our combination of authentic practices and integration with the natural sciences may also be possible in other branches of mathematics education. We think that there are challenging possibilities in mathematical modelling and optimization.

11 Education based on authentic professional practices

In our study, statistics used in authentic professional practices acts as a bridge for students between mathematics and the natural sciences (Chapter 5). Students experienced that the authentic contexts urged them to look beyond the data (Chapter 2), explain variability (Chapter 3), or look more closely at the results when sampling (Chapter 4). The fact that the authentic contexts were based on professional practices was a stimulus to learn for students. Many students said or wrote: “Now I know why I have to learn this” (Chapter 5). In this way we addressed a pressing problem in mathematics education: the perceived irrelevance and lack of motivation for the subject (Claxton, 1991). Our solution is in line with Makar and Confrey (2007) who argue that students who engage in statistical inquiry with a compelling purpose, such as modelling experiences with authentic data, gain a deeper understanding of data analysis, the context itself and modelling.

A challenge in basing teaching strategies on authentic professional practices is that the practices must be educationalized, yet still have to give the students the feeling that the activities are authentic. In our strategy most activities are authentic, but in reality professionals go much further, in the questions they ask, the instruments they use and in the depth of their investigations. The tasks in our activities are small and less complicated compared to real authentic professional practice. The reason for this is that professionals and students have different aims. Students have to learn something and professionals want to obtain results that count. Professional practices have a cultural history with its own knowledge and values and students are confronted with the practice only once.

Another limitation of our instructional unit based on authentic professional practices was that it was carefully designed to realize the bridge of statistics and we have only worked out one way of realizing a bridge between mathematics and the natural sciences. We did not investigate the direct connection between mathematics and the natural sciences, nor using another field as a bridge between mathematics and a natural science. Also, we analysed student work to investigate whether the instructional unit supported the students to identify links between mathematics (M), Statistics (S), professional practice (P) and natural science (N). We developed an analytical model (see Figure 2). With this we categorized the students’ work about elements M, S, P and N and coded the connections that students made between the four elements with MS, MP, PS, PN and SN.

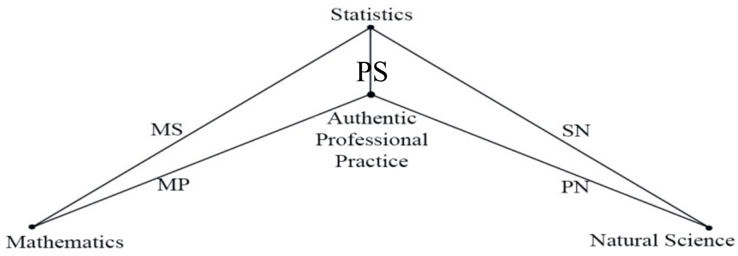

Figure 2. Statistics used in a professional practice as a bridge between mathematics and natural science (bridge metaphor from Dierdorp et al., 2013).

However, the boundaries of the elements M, S, P and N for our bridge metaphor were not always clear. Sometimes it was difficult to attribute a code because of the apparent overlap of the four elements. For example, defining the difference between mathematics and statistics is not trivial. Furthermore, we did not investigate the direct connection between mathematics to the natural sciences. We did not do this, because we were interested in how students experience statistics as an intermediary between mathematics and the natural sciences.

Future research to involve the direct connection between mathematics and the natural sciences would be interesting, for instance to replace statistics with another topic, such as mechanics and investigate whether our model fits the new situation, or to replace mathematics and natural sciences. Our model might fit the new situation, but research has to be done to investigate this. Often it seems that every participant (students, teachers, curriculum specialists, even governmental officials) acknowledges that a huge gap exists between school subjects like mathematics and natural science, but we presume that research such as ours can help to fill the gaps.

We designed a strategy that provoked or inspired students to learn about statistics and that stimulated them to use it in other practices. Our study shows that such a strategy works to teach students statistical techniques, that they can learn to understand the mathematical background, use mathematical tools and that the natural sciences offer powerful contexts to evoke students' interests to learn and reason about statistics. Maybe our metaphor is a bridge too far to be used in such new situations, but we think that we found evidence that statistics can bridge the gap between mathematics and the natural sciences as it were a bridge over troubled water (cf. Simon & Garfunkel, 1970, track 1).

References

Ainley, J., Pratt, D., & Hansen, A. (2006). Connecting engagement and focus in pedagogic task design. *British Educational Research Journal*, *32*(1), 23–38.

Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning, 13,* 5–26.

Ben-Zvi, D., & Garfield, J. (2010). Introducing the emerging discipline of statistics education. *School Science and Mathematics,108*, 355–361.

Brandom, R. B. (2000). *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, MA: Harvard University Press.

Claxton, G. (1991). *Educating the enquiring mind: The challenge for school science.* New York: Harvester Wheatsheaf.

Dierdorp, A., Bakker, A., Eijkelhof, H. M. C., & Van Maanen, J. A. (2013). Betekenisvolle statistiek in beroepspraktijken als brug tussen wiskunde en natuurwetenschappen: evaluatie van een ontwerponderzoek. *Pedagogische Studiën, 90*(2), 4–18.

Edelson, D. C., & Reiser, B. J. (2006). Making authentic practices accessible to learners: Design challenges and strategies in R. Keith Sawyer, *Cambridge Handbook of the Learning Sciences.* Cambridge University Press.

Eijkelhof, H. M. C., & Krüger, J. (2009). *Improving the quality of innovative science teaching materials*. Paper presented at European Science Education Research Association conference (ESERA 2009).

Fullan, M. (2007). *The Meaning of Educational Change* (4th ed.). New York: Teachers College.

Pfannkuch, M. (2008, July). *Building sampling concepts for statistical inference: A case study.* Paper presented at the International Congress on Mathematical Education, Monterrey, Mexico.

Pratt, D., Ainley, J., Kent, P., Levinson, R., Yogui, C., & Kapadia, R. (2011) Informal inferential reasoning as a contextualised risk-based enterprise. *Mathematical Thinking and Learning, 13*(4), 322–345.

Prins, G. T. (2010). *Teaching and learning of modeling in chemistry education*. Utrecht, the Netherlands: CD-Bèta Press.

Knippels, M.C.P.J., Goedhart, M., & Plomp, T. (2008). Docenten in onderzoek - het DUDOC-programma. *Tijdschrift voor Didactiek der Bèta-Wetenschappen, 25*(1-2), 51–69.

Makar, K., & Confrey, J. (2007). Moving the context of modeling to the forefront: Preserviceteachers' investigations of equity in testing. In W. Blum, P. Galbraith, H. W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education.* New York: Springer.

Nieveen, N., McKenney, S., & Van den Akker, J. (2006). Educational design research: the value of variety. In Van den Akker, Gravemeijer, K., McKenney, S., & Nieveen, N. (Eds.) (2006) *Educational Design Research* (pp. 151–158). London: Routledge.

Simon, P., & Garfunkel, A. (1970). Bridge over troubled water. On *Bridge over troubled water* [CD]. Washington DC: Columbia Records.

Tessmer, M. (1993). *Planning and conducting formative evaluations.* London: Kogan Page.

Westbroek, H. B. (2005). *Characteristics of meaningful chemistry education - The case of water quality.* Utrecht, the Netherlands: CD-Bèta Press.

Westbroek, H. B., Klaassen, K., Bulte, A., & Pilot, A. (2010). Providing students with a sense of purpose by adapting a professional practice. *International Journal of Science Education 28,* 1063–1086.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*, 223–265.

Westra, R. (2008). *Learning and teaching ecosystem behaviour in secondary education. Systems thinking and modelling in authentic practices*. Utrecht, the Netherlands: CD-Bèta Press.

Summary

Samenvatting

Acknowledgements

Curriculum vitae

Overview of publications related to this thesis

FIsme Scientific Library

Appendix A

Apeendix B

Summary

One of the key challenges in mathematics and science education in secondary school is to realise coherence between these school subjects. The idea explored in this PhD thesis is that statistical modelling is one of the possible ways to let students experience the connections between mathematics on which statistical modelling techniques are based, and scientific contexts in which they can be applied. One set of techniques that receive rather little attention in education and in the literature is that of correlation and regression, despite their frequent use in many disciplines and fields of work.

Many professionals use correlation and linear regression models to draw inferences about real-world situations. However, for secondary school students, the relevance of these techniques is less obvious. When taught – in the Netherlands it is not mandatory – it is often limited to the techniques practised with unrealistically "clean" data sets. Learning is then focused on the mathematical apparatus and isolated numbers, whereas the *raison d'être* of the technique is its usefulness in process analysis. So, coherence is directly visible when correlation and regression are presented as the mathematical domain in which processes that are observed in the sciences have their natural place. Because students increasingly need to do small science research projects, it is relevant for them to learn and understand correlation and regression models and to apply such models in new situations and different school subjects.

The purpose of this PhD project was to find out how to support upper-secondary school students' learning and understanding of correlation and regression models in such a way that they can apply the techniques in new situations and experience coherence between mathematics and the natural sciences. We addressed the following main research question: "What are characteristics of a valid and effective teaching and learning strategy to teach students about correlation and regression in such a way that they experience coherence between mathematics and the natural sciences?" By "valid" we mean that the strategy must be in line with prevailing epistemological ideas of the school subjects involved, such as calculate standard deviation (mathematics), understand aerobic metabolism (biology), or produce a formula for a regression line (statistics). We consider the strategy effective if students can correctly apply correlation and regression models to solve real-world problems.

Statistical modelling with correlation and regression in scientific or professional practices involves several steps. Variables related to the phenomena at stake need to be measured. The data form a sample that can be analysed by means of statistical techniques. With understanding of modelling techniques, sampling, variability and related statistical concepts, students can model the phenomena, and then shuttle back to the implications for the professional practices. By experiencing such steps students can learn about the modelling cycle.

Basing our strategy on professional practices in which science and statistical modelling are used, we assumed that students could learn to see the connections between these professional practices, mathematics, statistics and science. In line with recent research in science education we expected that a teaching and learning strategy becomes meaningful for students when such strategy is based on authentic professional practices, because in relevant situations theory and application are not separated. Therefore, we decided to design a teaching and learning strategy based on authentic professional practices.

Recent research in statistics education focuses on statistical inference, defined as probabilistically expressed generalisations from data. It is acknowledged that informal statistical inference is important, not only for future citizens and employees, but also as preparation for more formal inferential reasoning. In line with this research we started the teaching and learning strategy with informal inferential reasoning and worked towards more formal techniques in the second half of the strategy. In drawing informal inferences, the focus is on reasoning and conceptual understanding and less on statistical procedures. An explicit link between statistical inferences and tendency (regression line) is the acknowledgement of a mechanism or tendency that extends beyond the data at hand as an important idea to draw inferences. We focus on tendencies that can be modelled by means of correlation or regression.

To answer the research question we designed a teaching and learning strategy based on authentic professional practices. In our investigations we focused on four related aspects. After a broad focus on informal inferential reasoning (1), we zoomed in on specific concepts required for inferential reasoning: variability in measurement data (2) and sampling (3). Measurement and sampling are also important interfaces between mathematics and science. Last we zoomed out and focused more broadly on the coherence between mathematics, statistics, science and professional practices (4). We tested our teaching and learning strategy in four studies, each with its own research question:

1. How does a teaching and learning strategy based on authentic practices support students in making statistical inferences about authentic problems with the help of correlation and linear regression?

2. How do secondary students consider variability within measurement activities based on authentic professional practices?

3. What is the potential of tasks based on authentic professional practices to support students in developing concepts that underlie sampling in such a way that they can shuttle between contextual and statistical spheres?

4. To what extent can professional practices serve as meaningful contexts to show students connections between mathematics, statistics, science and professional practices?

In this thesis we formulated a set of selection criteria for authentic practices related to correlation and regression modelling activities. On the basis of these criteria we chose the professional practices of sports physiologists, researchers who monitor dyke heights, and calibrators of measurement devices. As part of the teaching and learning strategy we designed three chapters for an instructional unit, each inspired by one of the three authentic professional practices. The intention of the first chapter (measuring physical condition) was to introduce conceptual ideas in an informal way. It stimulated students' motivation to learn about correlation and regression to predict students' own threshold point between aerobic and anaerobic metabolism in order to improve their physical condition. The intention of the second chapter (monitoring dyke heights) was to engage students in a more formal approach to the concepts and techniques in order to predict the moment to heighten dykes to prevent flooding, and the third chapter (calibrating) to give the students the possibility to apply what they had learned in a new situation calibrating measurement devices. We tested different aspects of our strategy in the four studies (Chapters 2 to 5).

Students’ informal inferential reasoning

Chapter 2 reports on supporting students to draw *informal inferences* using correlation and regression models. We developed a data analysis matrix to analyse this ability and used the students’ written work and transcripts of student interactions. In the matrix we compared the observed student responses with the hypothesized responses. In this manner we investigated how our strategy supported students in making informal inferences about authentic problems with the help of correlation and regression models.

The findings indicate that our teaching and learning strategy inspired by authentic professional practices has advantages but also produces new challenges. Most students saw the practical value of correlation and regression, and many of them were able to draw realistic inferences. The teaching and learning strategy with statistical core ideas embedded in authentic professional practices seemed effective to help students to make informal statistical inferences. Another advantage of the strategy was that it helped students to understand that practitioners in many fields use these techniques in their professional work and to realise that they could use these techniques in many situations themselves at school or in a future academic study. A challenge for students was that they did not know how to deal with the variability that should be taken into account when making predictions. For example, they often forgot to consider the variability when predicting the moment to heighten the dyke, by only using the regression line. Considering the variability around the regression line is crucial given the risk of flooding. We therefore saw a need to pay more attention to variability in the next cycle of design-based research.

Students’ reasoning about variability through measurement activities

Chapter 3 presents our investigation of how students reason about *variability* when engaged with measurement tasks using correlation and regression models. An advantage of involving students in measurement activities is that it invites them to make connections between the real world and the world of data. Also, measurement activities seem suitable for teaching about variability in interdisciplinary contexts.

We analysed students' spoken utterances to get insight into the ways how students in Grade 12, aged 17-18, consider variability when engaged in context-rich measurement activities. To investigate how students reason about variability, we adapted a framework of Wild and Pfannkuch (1999) based on how statisticians consider variability: noticing and acknowledging variability, measuring and modelling variability, explaining variability, and using investigative strategies.

We concluded that students considered variability in all of these ways, though of course in less advanced ways than statisticians would do. We proposed that the authentic character of the activities supported students in modelling their measurement results and analyse these with the help of techniques such as finding a trend in order to find a meaningful answer for an authentic problem and consider the role of variability in several ways.

Students' conceptual development of sampling for shuttling between context and statistics

Chapter 4 focuses on a key aspect of informal inferential reasoning and a way to connect professional contexts, mathematics and science: sampling. Many professionals are engaged with sampling in their professional practices. However, sampling is another somewhat neglected area in statistics education. We investigated if students would see the purpose of sampling tasks and the utility of concepts. Also, we investigated if students would be able to mathematize authentic problems and apply concepts in context. We designed two sampling tasks requiring correlation and regression models, based on an authentic professional practice of a sports physiologist. We carried out two case studies to test the potential of these tasks to support students' concepts that underlie sampling, and to investigate what teachers should do to exploit this potential. We investigated if students made the connection forth and back (shuttling) between the real world (contextual sphere) and data (statistical sphere).

The analyses showed that in order to support students' shuttling between contextual and statistical spheres, realistic sampling tasks inspired by authentic professional practices can be effective. Such shuttling helped the students to see the tasks' purpose and the concepts' utility, and contributes to students' ability to mathematize authentic problems and apply concepts in meaningful contexts. However, shuttling back seemed rather a challenge for students. The teacher subtly tried to stimulate shuttling back, but was unsuccessful and had to ask explicitly for the meaning of their answers in relation to the context. After that the students acknowledged the importance of shuttling back. Despite the authentic nature of the problems used in the unit, it still seems necessary for teachers to support students in shuttling back.

Coherence between mathematics and the natural sciences

Chapter 5 reports on how students integrate knowledge from mathematics and the natural sciences when they practise informal inferential reasoning about solutions of authentic problems, and how they value such an approach. As mentioned before we assumed that a teaching and learning strategy becomes meaningful for students when such a strategy is based on authentic professional practices. However, in this study we suggested that that meaningful alone is not sufficient for students to make connections between school subjects. For making connections, it is also important that students master the concepts and are able to apply them. We investigated more broadly to what extent students experienced the coherence between mathematics and the natural sciences, with statistical concepts and modelling techniques as the bridging tools.

We developed an analytical model to investigate students' ability to integrate mathematics, statistics, science and professional practices, and analysed students' responses to questionnaires and tasks before and after the lessons. We concluded that our designed instructional unit as part of the teaching and learning strategy is found meaningful by most students. The strategy motivated them and they showed to be able to apply the learned concepts also in other school subjects.

Conclusion and discussion

Our conclusion to the main research question is that our teaching and learning strategy for correlation and regression seems valid and effective. It seems valid because we designed the strategy in line with prevailing epistemological ideas of the involved school subjects. It seems effective because the involved students learned to solve real-world problems by correctly using correlation and regression models. They also appeared to understand the modelling process and statistical concepts involved, and were able to combine mathematical and statistical techniques with concepts of the natural sciences when solving real-world problems.

We think that students who engage in education of correlation and regression models based on authentic professional practices with a compelling purpose can gain an understanding of data analysis and modelling. We argued that tasks about statistics based on such practices can have the potential to be engaging for students while they develop concepts relevant for their school subjects.

Limitations and future research

As a limitation of this research one should take into account that we described research results of three schools and that we have to be careful with generalizations. We suggest that extensive accompanying instruction guidelines for teachers need to be developed, and that the teaching and learning strategy should be tested in more schools. Future research with more teachers involved should give an answer to the question if our strategy works for other teachers and what support they need. It would further be interesting to investigate if a similar design strategy would work for other parts of mathematics (e.g., mathematical modelling and optimization), other school systems and other countries.

Our strategy based on authentic professional practices inspired students to learn about statistics and stimulated them to use it in other situations, such as school science projects. Our study shows that such strategy can be effective to teach students statistical techniques, that they can learn to understand the mathematical background, use mathematical tools, and that the natural sciences offer powerful contexts to motivate students to learn and reason about statistics. We hope that future research will make learning for students more and more meaningful and interesting in such a way that they experience more coherence between school subjects.

Samenvatting

Een belangrijke uitdaging voor de wiskunde en de natuurwetenschappelijke vakken in het voortgezet onderwijs is het realiseren van samenhang tussen deze disciplines. In dit proefschrift hebben we onderzocht of statistisch modelleren een manier is om leerlingen samenhang te laten ervaren tussen wiskunde, waarop statistische modelleertechnieken zijn gebaseerd, en de natuurwetenschappelijke contexten waarin ze kunnen worden toegepast. Sommige van zulke technieken, die nog weinig aandacht hebben gekregen in de literatuur, hebben betrekking op correlatie en regressie, hoewel deze beide technieken gebruikt worden binnen tal van disciplines en beroepen**.**

In de praktijk worden correlatie en lineaire regressiemodellen veelvuldig gebruikt om conclusies te trekken over onderzoek in de praktijk. Voor middelbare scholieren is de relevantie van deze technieken echter minder evident. Als deze technieken al onderwezen worden – hetgeen in Nederland niet verplicht is – wordt het onderwijs hierin vaak beperkt tot het leren van de technieken, waarbij vaak gebruikgemaakt wordt van onrealistische "schone" datasets. Leren is dan vaak gericht op technieken, terwijl de *raison d'être* van de techniek gelegen is in de uitkomst van het modelleerproces. Als de technieken die nodig zijn voor het leren van correlatie en regressie worden gepresenteerd als het wiskundige domein waarin de in de natuurwetenschappen waargenomen processen hun natuurlijke plaats hebben, lijkt het mogelijk dat voor de leerlingen samenhang zichtbaar wordt. Mede omdat de leerlingen op school steeds vaker kleine wetenschappelijke onderzoeksprojecten uitvoeren, is het voor hen van belang om over correlatie en regressie te leren en dergelijke technieken toe te passen in nieuwe situaties en verschillende schoolvakken.

Het doel van dit promotieonderzoek was inzicht te verkrijgen in de wijze waarop scholieren uit de bovenbouw van het vwo ondersteund kunnen worden om over correlatie en regressie te leren op een zodanige wijze dat ze de technieken in nieuwe situaties kunnen toepassen en samenhang tussen wiskunde en de natuurwetenschappen kunnen ervaren. De hoofdonderzoeksvraag die we stelden, was:

> *Wat zijn de kenmerken van een valide en effectieve onderwijsleerstrategie om leerlingen correlatie en regressie te onderwijzen op een zodanige*

wijze dat ze samenhang tussen wiskunde en de natuurwetenschappen ervaren?

Met ‘valide’ bedoelen we dat de strategie in overeenstemming moet zijn met de epistemologische ideeën die binnen de betrokken disciplines gangbaar zijn, zoals bijvoorbeeld het berekenen van de standaarddeviatie (wiskunde), het begrijpen van aerobe stofwisseling (biologie), of het produceren van een formule voor een regressielijn (statistiek). Wij beschouwen de strategie als effectief als leerlingen de juiste correlatie- en regressiemodellen kunnen toepassen bij het oplossen van reële problemen.

Statistisch modelleren in een al dan niet wetenschappelijke beroepspraktijk omvat verschillende stappen. Variabelen die verband houden met de verschijnselen waar het om gaat, moeten worden gemeten. De gegevens vormen een steekproef en kunnen worden geanalyseerd met behulp van statistische technieken. Door gebruik van modelleringstechnieken, steekproeven, variabiliteit en andere relevante statistische begrippen, kunnen leerlingen de verschijnselen modelleren, en vervolgens terugkoppelen naar de gevolgen voor de beroepspraktijk. Door dergelijke stappen te zetten, kunnen leerlingen iets leren over een modelleercyclus.

We hebben onze strategie gebaseerd op beroepspraktijken waarin wetenschappelijke en statistische modellen worden gebruikt en zijn ervan uitgegaan dat leerlingen hierdoor worden gestimuleerd om samenhang te leren zien tussen deze beroepspraktijken, wiskunde, statistiek en natuurwetenschappen. In lijn met recent onderzoek in het natuurwetenschappelijk onderwijs hebben we aangenomen dat een dergelijke strategie, waarbij theorie en toepassing in relevante situaties niet van gescheiden zijn, voor leerlingen betekenisvol wordt. Daarom hebben we besloten om een onderwijsleerstrategie te ontwerpen die gebaseerd is op authentieke beroepspraktijken.

Recent onderzoek in statistiekonderwijs richt zich op “statistische inferenties” (gevolgtrekkingen), gedefinieerd als probabilistische generalisaties op basis van data. Informele statistische inferenties zijn niet alleen belangrijk voor toekomstige burgers en werknemers, maar ook als voorbereiding op een meer formele inferentiële wijze van redeneren. Aansluitend op dergelijk onderzoek zijn wij in onze onderwijsleerstrategie ermee begonnen leerlingen te laten kennismaken met “informeel inferentieel redeneren” en hebben we toegewerkt naar meer formele technieken in de tweede helft van de strategie. Bij het opstellen van informele inferenties, ligt de focus op kunnen redeneren en op begripsmatig begrip, en minder op statistische procedures. In educatief onderzoek naar informeel inferentieel redeneren spreekt men dan ook van de noodzaak om een mechanisme of een trend te herkennen waarbij men “voorbij” de data kijkt als een belangrijk idee om conclusies te trekken. Dit betekent dat je conclusies trekt die niet meteen in de data te zien zijn, maar wel hieruit te destilleren zijn. Wij richten ons in dit onderzoek op trendverschijnselen die kunnen worden gemodelleerd door middel van correlatie of regressie.

Om de onderzoeksvraag te kunnen beantwoorden hebben we een onderwijsleerstrategie ontworpen op basis van authentieke beroepspraktijken, waarbij het mogelijk was om onderdelen van de beroepspraktijk te didactiseren en te laten aansluiten bij begrippen van het wiskundig en natuurwetenschappelijk curriculum van de middelbare school. In ons onderzoek hebben we ons gericht op vier aspecten. Na een brede focus op informeel inferentieel redeneren (1), hebben we ingezoomd op specifieke begrippen die nodig zijn voor inferentieel redeneren: variabiliteit in meetgegevens (2) en steekproeven nemen (3). Meten en steekproeven nemen zijn activiteiten op het raakvlak tussen de wiskunde en de natuurwetenschappen. Ten slotte hebben wij uitgezoomd door ons meer in het algemeen op de samenhang te richten tussen wiskunde, statistiek, natuurwetenschap en beroepspraktijken (4). We hebben onze onderwijsleerstrategie getest in vier studies, die elk een eigen onderzoeksvraag hadden:

1. Op welke manier kan een onderwijsleerstrategie die is gebaseerd op authentieke beroepspraktijken, leerlingen ondersteunen bij het maken van statistische inferenties over authentieke problemen waarbij gebruikgemaakt wordt van correlatie en regressie?

2. Hoe houden vwo-scholieren rekening met variabiliteit als zij meetactiviteiten uitvoeren die gebaseerd zijn op authentieke beroepspraktijken?

3. Wat is het potentieel van taken op basis van authentieke beroepspraktijken om leerlingen te ondersteunen in het ontwikkelen van begrippen die ten grondslag liggen aan steekproeven, op een zodanige manier dat ze kunnen ‘pendelen’ tussen context en statistiek?

4. In hoeverre kunnen beroepspraktijken als betekenisvolle contexten fungeren om leerlingen verbanden te laten leggen tussen wiskunde, statistiek, natuurwetenschappen en beroepspraktijken?

In dit proefschrift hebben we een aantal selectiecriteria geformuleerd voor authentieke beroepspraktijken waarin men gebruikmaakt van correlatie- en regressiemodellering. Op basis van deze criteria hebben we ervoor gekozen om binnen onze onderwijsleerstrategie de hoofdstukken van het in de klas te gebruiken en special daarvoor ontworpen lesmateriaal te baseren op de beroepspraktijken van sportfysiologen (1), onderzoekers die dijkhoogten monitoren (2), en laboranten die meetapparaten kalibreren (3). De bedoeling van het eerste hoofdstuk van het lesmateriaal (meten van fysieke conditie) was om de begrippen op een informele manier te introduceren. Het had als doel om de motivatie van leerlingen om te leren over correlatie en regressie te bevorderen, waarbij ze hun eigen drempelwaarde tussen het aerobe en anaerobe metabolisme moesten voorspellen met als doel hun fysieke conditie te verbeteren. De bedoeling van het tweede hoofdstuk (monitoren van dijkhoogten) was om leerlingen te betrekken bij een meer formele benadering van de begrippen en technieken om het moment te kunnen voorspellen waarop dijken verhoogd moeten worden om overstromingen te voorkomen, en het derde hoofdstuk (kalibreren) had als doel om de leerlingen datgene wat ze geleerd hadden in een nieuwe situatie (kalibreren meetapparaten) te laten toepassen. De verschillende aspecten van onze strategie hebben we in vier studies getest (Hoofdstukken 2 tot en met 5).

Informeel inferentieel redeneren van leerlingen

Hoofdstuk 2 beschrijft hoe leerlingen worden geholpen om aan de hand van correlatie- en regressiemodellen informele inferenties te trekken. We hebben een data-analyse matrix ontwikkeld om dit vermogen te analyseren en hebben zowel het schriftelijk werk van de leerlingen gebruikt als de transcripties van hun interacties. In de matrix hebben we de feitelijke antwoorden van leerlingen vergeleken met die welke we hadden verwacht. Op deze manier hebben we onderzocht hoe onze strategie leerlingen ondersteunt bij het maken van informele inferenties over authentieke problemen met behulp van correlatie- en regressiemodellen.

De bevindingen wijzen erop dat onze onderwijsleerstrategie geïnspireerd op authentieke beroepspraktijken voordelen biedt, maar ook een nieuwe uitdaging betekent. De meeste leerlingen zagen de praktische waarde van correlatie en regressie voor een beroepspraktijk, en velen van hen schenen in staat om realistische conclusies te trekken. De onderwijsleerstrategie leek effectief om leerlingen te helpen informele statistische inferenties te maken. Een ander voordeel van de strategie was dat leerlingen tot het inzicht kwamen dat deskundigen op veel terreinen in hun professionele werk gebruikmaken van deze technieken en dat ze zelf deze technieken in veel situaties op school of in een toekomstige academische studie zouden kunnen gebruiken. Een uitdaging voor de leerlingen was dat ze niet wisten hoe ze moesten omgaan met de variabiliteit waarmee rekening moet worden gehouden bij het maken van voorspellingen. Zo vergaten ze bijvoorbeeld vaak om bij het voorspellen van het moment waarop een dijk verhoogd moest worden, rekening te houden met de variabiliteit en gebruikten ze alleen de regressielijn bij hun voorspellingen. Rekening houden met variabiliteit rond de regressielijn is cruciaal vanwege het risico van overstromingen, maar is vanwege het authentieke karakter van het probleem nogal complex. We hebben het dan ook noodzakelijk geacht om vanaf de tweede ronde waarin we het lesmateriaal met leerlingen uittestten meer aandacht te besteden aan variabiliteit.

Redeneren van leerlingen over variabiliteit veroorzaakt door meetactiviteiten

In hoofdstuk 3 onderzoeken we hoe leerlingen in hun redeneringen omgaan met variabiliteit wanneer zij aan taken werken waarbij zij metingen uitvoeren en de verkregen data analyseren met correlatie- en regressiemodellen. Leerlingen meetactiviteiten laten uitvoeren heeft het voordeel dat het hen in staat stelt om relaties te leggen tussen de reële natuurwetenschappelijke contexten (“real world”) en de wereld van de data. Ook lijken meetactiviteiten geschikt te zijn om in interdisciplinaire contexten variabiliteit te onderwijzen.

Wij hebben de mondelinge uitspraken van leerlingen geanalyseerd om inzicht te krijgen in de manieren waarop leerlingen met variabiliteit omgaan tijdens context-rijke meetactiviteiten. Om dit te onderzoeken hebben wij het raamwerk bewerkt van Wild en Pfannkuch (1999), dat gebaseerd is op hoe statistici variabiliteit zien: “aandacht voor en erkenning van variabiliteit”, “meting en modellering van variabiliteit”, “verklaring van variabiliteit”, en “gebruik van onderzoeks-strategieën”. We hebben dit onderzocht omdat bij informeel inferentieel redeneren het voor leerlingen belangrijk en misschien zelfs noodzakelijk is om goed om te gaan met variabiliteit. Onze verwachting was dat door het authentieke karakter van de activiteiten de leerlingen beter in staat zouden zijn hun meetresultaten te modelleren en te analyseren met behulp van technieken, zoals het vinden van een trend om een betekenisvol antwoord te vinden voor een authentiek probleem. Het zou hen ook helpen om op verschillende manieren de rol van variabiliteit te bestuderen. We hebben geconcludeerd dat leerlingen, hoewel natuurlijk op een minder hoog niveau dan statistici, op alle aspecten van het raamwerk in hun redeneren rekening hielden met variabiliteit.

De begripsmatige ontwikkeling van leerlingen om steekproeven te trekken teneinde tussen context en statistiek te kunnen 'pendelen'

Hoofdstuk 4 richt zich op een essentieel aspect van het informeel inferentieel redeneren en op een methode om professionele contexten, wiskunde en natuurwetenschappen met elkaar te verbinden: steekproeven nemen. Veel deskundigen houden zich in hun vak met steekproeven bezig. Maar ook steekproeven vormen een enigszins verwaarloosd gebied binnen het statistiekonderwijs. We hebben onderzocht of leerlingen het doel van steekproeftaken en het nut van de onderliggende begrippen zouden zien. Ook hebben we onderzocht of leerlingen in staat zouden zijn om authentieke problemen te mathematiseren en de begrippen in een context te kunnen toepassen. We hebben twee steekproeftaken ontworpen, gebaseerd op een authentieke beroepspraktijk van een sport-fysioloog waarbij correlatie- en regressiemodellen gebruikt moesten worden.

We hebben twee case studies uitgevoerd om na te gaan wat het potentieel is van deze taken om leerlingen steun te bieden bij het leren van begrippen die ten grondslag liggen aan steekproeven, en om te onderzoeken wat leraren moeten doen om dit potentieel te benutten. We hebben onderzocht of de leerlingen 'pendelden', dat wil zeggen de relatie legden tussen de realiteit ("contextual sphere") en de data ("statistiscal sphere") en omgekeerd.

De analyses tonen aan dat om het pendelen van leerlingen tussen contextuele en statistiek te ondersteunen, realistische steekproeftaken geïnspireerd op authentieke beroepspraktijken effectief kunnen zijn. Dergelijk pendelen hielp de leerlingen om het doel van de taken en het nut van de begrippen te zien, en droeg bij aan het vermogen van de leerlingen om authentieke problemen te mathematiseren en de begrippen in betekenisvolle contexten toe te passen. Voor de leerlingen was de terugweg bij het pendelen niet gemakkelijk. De onderzoeker probeerde subtiel deze terugkoppeling te stimuleren, maar zijn poging mislukte en hij moest uitdrukkelijk vragen naar de betekenis van hun antwoorden in relatie tot de context. Daarna zagen de leerlingen het belang in van het terugkoppelen. Ondanks het authentieke karakter van de opdrachten in het lesmateriaal, lijkt het nog steeds noodzakelijk dat leraren leerlingen ondersteunen bij het terugkoppelen van hun antwoorden naar de context.

Samenhang tussen wiskunde en de natuurwetenschappen

Hoofdstuk 5 doet verslag van de manier waarop leerlingen kennis uit de wiskunde en de natuurwetenschappen integreren, als ze oefenen met informeel inferentieel redeneren met betrekking tot oplossingen van authentieke problemen. Ook beschrijft dit hoofdstuk hoe ze een dergelijke aanpak waarderen. Zoals al eerder gezegd zijn we ervan uitgegaan dat een onderwijsleerstrategie voor leerlingen betekenisvol gemaakt kan worden, als zij gebaseerd is op gedidactiseerde authentieke beroepspraktijken. We gebruikten ”betekenisvol” als term om relevante affectieve en metacognitieve aspecten van leren samen te nemen (nut, motivatie, toepassing, zien van verbanden, authenticiteit). In deze studie hebben we bovendien de gedachte geopperd dat een betekenisvolle strategie alleen niet voldoende is om leerlingen verbanden te laten leggen tussen disciplines. Daarvoor is het ook belangrijk dat leerlingen de begrippen beheersen en in staat zijn om ze toe te passen. Ook onderzochten we meer in het algemeen in hoeverre leerlingen de samenhang ervoeren tussen wiskunde en de natuurwetenschappen, met statistische begrippen en modelleertechnieken als instrumenten voor het leggen van verbanden.

We hebben een analytisch model ontwikkeld om het vermogen te onderzoeken verbanden te leggen tussen wiskunde, statistiek, natuurwetenschappen en beroepspraktijken, en hebben voor en na de lessen de antwoorden van leerlingen geanalyseerd die betrekking hadden op vragenlijsten en opdrachten uit het lesmateriaal. Onze conclusie was dat het door ons ontworpen lesmateriaal als onderdeel van de onderwijsleerstrategie door de meeste leerlingen als betekenisvol werd ervaren. De strategie motiveerde hen en ze lieten zien de geleerde begrippen te kunnen toepassen, ook bij andere disciplines.

Conclusie en discussie

Onze conclusie ten aanzien van de hoofdonderzoeksvraag is dat onze onderwijsleerstrategie voor het onderwijs van correlatie en regressie valide en effectief lijkt. Zij lijkt valide, omdat het ontwerp van de strategie in lijn is met de gangbare epistemologische ideeën van de betrokken disciplines. Zij lijkt effectief omdat de betrokken de leerlingen geleerd hebben om authentieke problemen op te lossen door correct gebruik te maken van correlatie- en regressiemodellen. De leerlingen lijken ook het modelleerproces en de gebruikte statistische begrippen te begrijpen en zijn in staat om bij het oplossen van echte problemen, wiskundige en statistische technieken te combineren met de begrippen uit de natuurwetenschappen. Wij denken dat leerlingen door het leren van correlatie- en regressiemodellen op basis van authentieke beroepspraktijken met een aansprekend doel inzicht in data-analyse en modellering kunnen krijgen. We hebben betoogd dat opdrachten met betrekking tot statistiek op basis van dergelijke praktijken het potentieel hebben om leerlingen te motiveren wanneer ze relevante begrippen voor de betrokken disciplines ontwikkelen.

De kenmerken die ertoe bij lijken te dragen dat leerlingen samenhang tussen wiskunde en de natuurwetenschappen ervaren zijn:

1. De strategie is gebaseerd op authentieke beroepspraktijken. Hiertoe dienen wel, voordat ze in het onderwijs ingezet kunnen worden, onderdelen van de beroepspraktijken gedidactiseerd te worden (specifiek in ons geval betrof dit de selectie en redactie van problemen die met correlatie en regressie onderzocht kunnen worden).
2. De strategie biedt de leerlingen lesmateriaal dat ze uitnodigt om inferentieel te redeneren, zowel formeel als informeel (zie bijvoorbeeld Hoofdstuk 2).
3. De strategie biedt de leerlingen een balans tussen de te leren begrippen en de onderliggende context. Een aandachtspunt daarbij is het pendelen tussen de begrippen en de context (zie Hoofdstuk 3).
4. De strategie bevat activiteiten voor leerlingen waarbij zij zelf data verzamelen en analyseren (zie bijvoorbeeld Hoofdstuk 4).

Beperkingen en vervolgonderzoek

Bij dit onderzoek moet worden aangetekend dat de onderzoeksresultaten slechts drie scholen betreffen en dat we voorzichtig zijn met generalisaties. Wij stellen voor om voorafgaand aan mogelijk invoering in het onderwijs eerst nauwkeurige begeleidende instructies voor leraren te ontwikkelen., en de strategie eerst in meer scholen te testen. Toekomstig onderzoek met meer deelnemende leraren moet een antwoord geven op de vraag of onze strategie werkt voor andere leraren en welke steun zij nodig hebben. Het zou verder interessant zijn om te onderzoeken of een soortgelijk ontwerpstrategie zou werken voor andere delen van de wiskunde (bijvoorbeeld bij wiskundig modelleren en optimalisering), andere schoolsystemen en andere landen.

Onze strategie inspireerde leerlingen om te leren over statistiek en stimuleerde hen om de geleerde begrippen te gebruiken in andere situaties, zoals bij projecten voor natuurwetenschappelijke vakken. Onze studie laat zien dat een dergelijke strategie effectief kan zijn om leerlingen statistische technieken te onderwijzen, dat ze wiskundige achtergronden leren begrijpen, wiskundige technieken leren gebruiken, en dat de natuurwetenschappen geschikte contexten bieden om hen te motiveren om over statistiek te leren en te redeneren. We hopen dat toekomstig onderwijs het leren van leerlingen in toenemende mate betekenisvol en interessant maakt zodat ze meer het nut van de te leren begrippen inzien, beter in staat zijn deze begrippen toe te passen en meer samenhang tussen de disciplines op school ervaren.

Acknowledgements

My research years for this thesis have been a wonderful experience, albeit also a very difficult period for personal reasons. The joy of doing research, the opportunity to develop myself, working with very inspiring people and working on an interesting problem, helped me to finish this journey. Luckily I got help from many people and organizations, some of which I will mention.

First and foremost I offer my sincerest gratitude to my research supervisors, Jan van Maanen, Harrie Eijkelhof and Arthur Bakker for their patient guidance, enthusiastic encouragement and useful critiques of this research work. Special thanks should be given to Arthur, my daily research supervisor, for his professional guidance and valuable support. Without his enormous effort to support me in many ways – professional and personal – it would not have been possible for me to finish the research project in the way I did. I learned so much from him that it is impossible to mention everything, but his excellent view of research and outstanding writing skills were of great value. Jan helped me a lot with his textual suggestions and broad supervision, and Harrie's contribution to the relevant aspects of the natural sciences was essential. Many thanks to the three of you.

My gratitude also extends to the organizations that made this research possible. First, Platform Bètatechniek for funding this research project. Furthermore, I thank the DUDOC Program and the people engaged with this program. This includes Tjeerd Plomp for his guidance and scientific advice, Marie Christine Knippels and Gjalt Prins for their organizational support, all Dudoc peers, and all participants who shared their knowledge and support. Further, I want to thank the administration of my school, College Hageveld, for offering me this opportunity, and the schools involved in this research: Het Nieuwe Eemland (teacher George Molenaar), Kai Munk College (teacher Piet Post) and specially Fioretticollege where my teaching and learning strategy was tested many times by Jan Wanders, who spent a lot of time to make the strategy a success. Many thanks to him. Last, but certainly not least I thank all my colleagues at the Freudenthal Institute. I cannot mention everybody but I do mention my inspiring talks with Aad Goddijn and Martin Kindt, Jaap den Hertog, also for being a musical partner (Jaap, I miss you), the many social talks with Ank van der Heiden, Ellen Komproe and Mariozee Wintermans and Mark Uwland for his help with English texts. Thanks for Irma van Leeuwen for her help with appendix A. Further, one person that should certainly be on my list is Nathalie Kuijpers who made so many corrections in my writings. Without her patience it should have taken me another year to finish this thesis. My gratitude also goes to my office mates at the university, especially Angeliki Kolovou and Michiel Veldhuis. Michiel was not only my carpool mate, but also statistics adviser and interrater, just as Anneke Verschut, of some of my data. Many thanks for this.

I would also like to extend my thanks to the many experts who advised me, among them Leen van de Oever (NIBI), Fred Jansen (ICLON), Sybren Welbedacht (College Hageveld), Rens Swart (STOWA), Richard Gill (Leiden University), and Bert Nijdam (Statistics educator). My colleagues at the Freudenthal Institute who were willing to listen to my research talks, also advised me and above all inspired me. Just as did all the SRTL participants from whom I learned a lot, particularly Dani Ben-Zvi (University of Haifa), Katie Makar (University of Queensland) for their support and detailed advice.

I am also grateful to Delft University of Technology for their data of dyke deviations and allowing me to use it for educational use. My thanks are especially for Boudewijn Possel who suggested the idea and gave me the data. Further thanks go to Ronald Gellish (Oakland University) for allowing to use his data for educational use.

Also, thanks for my many friends and family, for their warmth and kindness, and their role in helping me. Although there are too many to mention all of them, I will mention Ton Epskamp for his endless support, Jo Bemelmans for his advice and support, Djowin Zschuschen for his motivational support and do the housekeeping when I was not able to do it, Diana Dierdorp, Anton Dierdorp and Jara Hofman for helping when possible.

Before I finish my incomplete list with thanks, I want to thank all students for allowing me to observe them, publish about them, making video records of them and actually do my research. Thank you all!

Finally, I want to thank my son Tim for being the light of my life and inspiring me to finish this project. Also special thanks for my parents for their support, love and encouragement throughout my study and whole life. Unfortunately I miss Tim and my father at the completion of my PhD project. I dedicate this book to the three of you and will love you forever.

Curriculum vitae

Adri Dierdorp was born on April 6, 1958 in Velsen, the Netherlands. In 1977 he started a teacher education course at Ubbo Emmius (Groningen), learning to teach mathematics and biology. After he graduated in 1982, he worked as a mathematics and biology teacher at several schools. He decided to specialize in mathematics and attended MOB mathematics to be able to teach the highest level of secondary school students, and graduated in 1985. From 1986 until the present day he has been working as a teacher (mathematics, biology and ICT) at College Hageveld in Heemstede, a VWO (pre-university track) school. During this last period he has also done some computer programming for several institutions, programming educational software, teaching system engineering, and performed a number of management tasks such as being chairman at eleven primary school boards. In 2007 Adri started a PhD project at the Freudenthal Institute for Science and Mathematics education at Utrecht University under supervision of Prof. dr. Jan van Maanen, Prof. dr. Harrie Eijkelhof and dr. Arthur Bakker.

Overview of publications related to this thesis

Refereed journal articles

Dierdorp, A., Bakker, A., Eijkelhof, H. M. C., & van Maanen, J. A. (2011). Authentic practices as contexts for learning to draw inferences beyond correlated data. *Mathematical Thinking and Learning, 13(1&2),* 152-151.

Dierdorp, A., Bakker, A., Eijkelhof, H. M. C., & van Maanen, J. A. (2013). Betekenisvolle statistiek in beroepspraktijken als brug tussen wiskunde en natuurwetenschappen: evaluatie van een ontwerponderzoek. [Meaningful statistics in professional practices as a bridge between mathematics and science: An evaluation of design research]. *Pedagogische Studiën, 90,* 4-18.

Manuscript under revision

Dierdorp, A., Bakker, A., Ben-Zvi, D., & Makar, K. (submited). Supporting students' considerations of variability using measurement activities based on authentic practices.

Dierdorp, A., Bakker, A., van Maanen, J. A., & H. M. C. Eijkelhof (submitted) Supporting students' conceptual development for shuttling between context and statistics: The case of sampling.

Selected papers presented at conferences

Dierdorp, A., Bakker, A., Maanen van, J.A., Eijkelhof, H.M.C.(2010, July). *Educational versions of authentic practices as contexts to teach statistical modeling.* Paper presented at ICOTS 8 Ljubliana: Slovenia.

Dierdorp, A., Bakker, A., Maanen van, J.A., Eijkelhof, H.M.C.(2009, May). *Basing an instructional unit about statistical modelling on authentic scientific practices*. Paper presented at ORD, Leuven: Belgium.

Dierdorp, A., Bakker, A., Maanen van, J.A., Eijkelhof, H.M.C.(2009a, August). *Authentic practices as contexts for learning to draw conclusions from correlated data*. Poster presentation presented at Esera, Istanbul: Turkey.

Dierdorp, A., Bakker, A., Maanen van, J.A., Eijkelhof, H.M.C.(2009b, July). *Authentic practices as contexts for learning to draw conclusions from correlated data*. Poster presentation presented at SRTL6, Brisbane: Australia.

Dierdorp, A., Bakker, A., Eijkelhof, H. M. C., & van Maanen, J. A. (2011b, February). *Supporting students ability to reason about variability with regard to measurement activities inspired by authentic practices*. Poster presentation presented at CERME, Rzeszow: Poland.

FIsme Scientific Library
(formerly published as CD-β Scientific Library)

78 Dolfing, R. (2013). *Teachers' Professional Development in Context-based Chemistry Education. Strategies to Support Teachers in Developing Domain-specific Expertise.*
77. Mil, M.H.W. van (2013). *Learning and teaching the molecular basis of life.*
76. Antwi, V. (2013). *Interactive teaching of mechanics in a Ghanaian university context.*
75. Smit, J. (2013). *Scaffolding language in multilingual mathematics classrooms.*
74. Stolk, M.J. (2013). *Empowering chemistry teachers for context-based education. Towards a framework for design and evaluation of a teacher professional development programme in curriculum innovations.*
73. Agung, S. (2013). *Facilitating professional development of* Madrasah *chemistry teachers. Analysis of its establishment in the decentralized educational system of Indonesia.*
72. Wierdsma, M. (2012). *Recontextualising cellular respiration.*
71. Peltenburg, M. (2012). *Mathematical potential of special education students.*
70. Moolenbroek, A. van (2012). *Be aware of behaviour. Learning and teaching behavioural biology in secondary education.*
69. Prins, G.T., Vos, M.A.J. & Pilot, A. (2011). *Leerlingpercepties van onderzoek & ontwerpen in het technasium.*
68. Bokhove, Chr. (2011). *Use of ICT for acquiring, practicing and assessing algebraic expertise.*
67. Boerwinkel, D.J. & Waarlo, A.J. (2011). *Genomics education for decision-making. Proceedings of the second invitational workshop on genomics education, 2-3 December 2010.*
66. Kolovou, A. (2011). *Mathematical problem solving in primary school.*
65. Meijer, M. R. (2011). *Macro-meso-micro thinking with structure-property relations for chemistry. An explorative design-based study.*
64. Kortland, J. & Klaassen, C. J. W. M. (2010). *Designing theory-based teaching-learning sequences for science. Proceedings of the symposium in honour of Piet Lijnse at the time of his retirement as professor of Physics Didactics at Utrecht University.*

63. Prins, G. T. (2010).*Teaching and learning of modelling in chemistry education. Authentic practices as contexts for learning*.
62. Boerwinkel, D. J. & Waarlo, A. J. (2010). *Rethinking science curricula in the genomics era. Proceedings of an invitational workshop*.
61. Ormel, B. J. B. (2010). *Het natuurwetenschappelijk modelleren van dynamische systemen. Naar een didactiek voor het voortgezet onderwijs*.
60. Hammann, M., Waarlo, A. J., & Boersma, K. Th. (Eds.) (2010). *The nature of research in biological education: Old and new perspectives on theoretical and methodological issues – A selection of papers presented at the VIIth Conference of European Researchers in Didactics of Biology*.
59. Van Nes, F. (2009). *Young children's spatial structuring ability and emerging number sense*.
58. Engelbarts, M. (2009*). Op weg naar een didactiek voor natuurkunde-experimenten op afstand. Ontwerp en evaluatie van een via internet uitvoerbaar experiment voor leerlingen uit het voortgezet onderwijs*.
57. Buijs, K. (2008). *Leren vermenigvuldigen met meercijferige getallen*.
56. Westra, R. H. V. (2008). *Learning and teaching ecosystem behaviour in secondary education: Systems thinking and modelling in authentic practices*.
55. Hovinga, D. (2007). *Ont-dekken en toe-dekken: Leren over de veelvormige relatie van mensen met natuur in NME-leertrajecten duurzame ontwikkeling*.
54. Westra, A. S. (2006). *A new approach to teaching and learning mechanics*.
53. Van Berkel, B. (2005). *The structure of school chemistry: A quest for conditions for escape*.
52. Westbroek, H. B. (2005). *Characteristics of meaningful chemistry education: The case of water quality*.
51. Doorman, L. M. (2005). *Modelling motion: from trace graphs to instantaneous change*.
50. Bakker, A. (2004). *Design research in statistics education: on symbolizing and computer tools*.
49. Verhoeff, R. P. (2003). *Towards systems thinking in cell biology education*.
48. Drijvers, P. (2003). *Learning algebra in a computer algebra environment. Design research on the understanding of the concept of parameter*.
47. Van den Boer, C. (2003). *Een zoektocht naar verklaringen voor achterblijvende prestaties van allochtone leerlingen in het wiskundeonderwijs*.

46. Boerwinkel, D.J. (2003). *Het vormfunctieperspectief als leerdoel van natuuronderwijs. Leren kijken door de ontwerpersbril.*
45. Keijzer, R. (2003). *Teaching formal mathematics in primary education. Fraction learning as mathematising process*.
44. Smits, Th. J. M. (2003). *Werken aan kwaliteitsverbetering van leerlingonderzoek: Een studie naar de ontwikkeling en het resultaat van een scholing voor docenten*.
43. Knippels, M. C. P. J. (2002). *Coping with the abstract and complex nature of genetics in biology education – The yo-yo learning and teaching strategy*.
42. Dressler, M. (2002). *Education in Israel on collaborative management of shared water resources.*
41. Van Amerom, B.A. (2002). *Reinvention of early algebra: Developmental research on the transition from arithmetic to algebra*.
40. Van Groenestijn, M. (2002). *A gateway to numeracy. A study of numeracy in adult basic education.*
39. Menne, J. J. M. (2001). *Met sprongen vooruit: een productief oefenprogramma voor zwakke rekenaars in het getallengebied tot 100 – een onderwijsexperiment*.
38. De Jong, O., Savelsbergh, E.R. & Alblas, A. (2001). *Teaching for scientific literacy: context, competency, and curriculum.*
37. Kortland, J. (2001). *A problem-posing approach to teaching decision making about the waste issue.*
36. Lijmbach, S., Broens, M., & Hovinga, D. (2000). *Duurzaamheid als leergebied; conceptuele analyse en educatieve uitwerking*.
35. Margadant-van Arcken, M. & Van den Berg, C. (2000). *Natuur in pluralistisch perspectief – Theoretisch kader en voorbeeldlesmateriaal voor het omgaan met een veelheid aan natuurbeelden*.
34. Janssen, F. J. J. M. (1999). *Ontwerpend leren in het biologieonderwijs. Uitgewerkt en beproefd voor immunologie in het voortgezet onderwijs*.
33. De Moor, E. W. A. (1999). *Van vormleer naar realistische meetkunde – Een historisch-didactisch onderzoek van het meetkundeonderwijs aan kinderen van vier tot veertien jaar in Nederland gedurende de negentiende en twintigste eeuw*.
32. Van den Heuvel-Panhuizen, M. & Vermeer, H. J. (1999). *Verschillen tussen meisjes en jongens bij het vak rekenen-wiskunde op de basisschool – Eindrapport MOOJ-onderzoek*.

31. Beeftink, C. (2000). *Met het oog op integratie – Een studie over integratie van leerstof uit de natuurwetenschappelijke vakken in de tweede fase van het voortgezet onderwijs*.
30. Vollebregt, M. J. (1998). *A problem posing approach to teaching an initial particle model*.
29. Klein, A. S. (1998). *Flexibilization of mental arithmeticsstrategies on a different knowledge base – The empty number line in a realistic versus gradual program design*.
28. Genseberger, R. (1997). *Interessegeoriënteerd natuur- en scheikundeonderwijs – Een studie naar onderwijsontwikkeling op de Open Schoolgemeenschap Bijlmer*.
27. Kaper, W. H. (1997). *Thermodynamica leren onderwijzen*.
26. Gravemeijer, K. (1997). *The role of context and models in the development of mathematical strategies and procedures*.
25. Acampo, J. J. C. (1997). *Teaching electrochemical cells – A study on teachers' conceptions and teaching problems in secondary education*.
24. Reygel, P. C. F. (1997). *Het thema 'reproductie' in het schoolvak biologie*.
23. Roebertsen, H. (1996). *Integratie en toepassing van biologische kennis – Ontwikkeling en onderzoek van een curriculum rond het thema 'Lichaamsprocessen en Vergift'*.
22. Lijnse, P. L. & Wubbels, T. (1996). *Over natuurkundedidactiek, curriculumontwikkeling en lerarenopleiding*.
21. Buddingh', J. (1997). *Regulatie en homeostase als onderwijsthema: een biologie-didactisch onderzoek*.
20. Van Hoeve-Brouwer G. M. (1996). *Teaching structures in chemistry – An educational structure for chemical bonding*.
19. Van den Heuvel-Panhuizen, M. (1996). *Assessment and realistic mathematics education*.
18. Klaassen, C. W. J. M. (1995). *A problem-posing approach to teaching the topic of radioactivity*.
17. De Jong, O., Van Roon, P. H. & De Vos, W. (1995). *Perspectives on research in chemical education.*
16. Van Keulen, H. (1995). *Making sense – Simulation-of-research in organic chemistry education*.
15. Doorman, L. M., Drijvers, P. & Kindt, M. (1994). *De grafische rekenmachine in het wiskundeonderwijs*.
14. Gravemeijer, K. (1994). *Realistic mathematics education*.

13. Lijnse, P. L. (Ed.) (1993). *European research in science education*.
12. Zuidema, J. & Van der Gaag, L. (1993). *De volgende opgave van de computer*.
11. Gravemeijer, K, Van den Heuvel Panhuizen, M., Van Donselaar, G., Ruesink, N., Streefland, L., Vermeulen, W., Te Woerd, E., & Van der Ploeg, D. (1993). *Methoden in het reken-wiskundeonderwijs, een rijke context voor vergelijkend onderzoek*.
10. Van der Valk, A. E. (1992). *Ontwikkeling in Energieonderwijs*.
9. Streefland, L. (Ed.) (1991). *Realistic mathematics education in primary schools*.
8. Van Galen, F., Dolk, M., Feijs, E., & Jonker, V. (1991). *Interactieve video in de nascholing reken-wiskunde.*
7. Elzenga, H. E. (1991). *Kwaliteit van kwantiteit*.
6. Lijnse, P. L., Licht, P., De Vos, W. & Waarlo, A. J. (Eds.) (1990). *Relating macroscopic phenomena to microscopic particles: a central problem in secondary science education*.
5. Van Driel, J. H. (1990). *Betrokken bij evenwicht*.
4. Vogelezang, M. J. (1990). *Een onverdeelbare eenheid*.
3. Wierstra, R. F. A. (1990). *Natuurkunde-onderwijs tussen leefwereld en vakstructuur*.
2. Eijkelhof, H. M. C. (1990). *Radiation and risk in physics education*.
1. Lijnse, P. L. & De Vos, W. (Eds.) (1990). *Didactiek in perspectief.*

Appendix A

REAL FORECASTS AND HYPERBOLIC SAFETY MARGINS

1. **Future forecasts**

In our instructional unit we intended students to consider a safety margin around the regression line when predicting the moment to heighten a dyke (chapter two of the instructional unit). Because some students remembered the "rules of thumb" for a normal distribution, the students were initially inclined to take a linear margin. We decided that a linear margin was an acceptable way for students to consider the risks involved. However, when we discussed this with statistician Prof. dr. Richard Gill, he explained that in real forecasting the margin around the regression line is not linear but hyperbolic, a theorem discussed and proven in this appendix. As a simplification of the authentic practice we decided a linear margin would be acceptable as a first step. As a next step, the teacher can discuss the fact that safety margins for the far future have to be larger than for the near future. We expect students to know this from weather predictions, but the proof of the hyperbolic shape transcends the secondary school curriculum.

Time is the undependable variable when making future forecasts based on the regression line. The magnitude of the forecast error depends on whether the formula of the regression line is known at x_t with $1 \le t \le T$ and when the forecast is to be made at X with $X > x_T$. In our case of dyke monitoring we get a regression line $H_t = a_0 + b_0 d_t + e_t$ with $1 \le t \le 44$ H for deviations of dyke heights and d for days at moment t. For $t > 44$ we use the estimated equation $\overline{H_t} = a_1 + b_1 d_t$.

The forecast error with $t > 44$ is:

$$H_t - \overline{H_t} = (a_0 + b_0 d_t + e_t) - (a_1 + b_1 d_t) = (a_0 - a_1) + (b_0 - b_1) d_t + e_t$$

The variance of this forecast error is

$$\sigma_f^2 = E(H_t - \overline{H_t})^2 = E[(a_0 - a_1) + (b_0 - b_1) d_t + e_t]^2 =$$

$$E[(a_0 - a_1)^2 + (b_0 - b_1)^2 d_t^2 + e_t^2 + 2(a_0 - a_1)(b_0 - b_1) d_t + 2(a_0 - a_1) e_t + 2(b_0 - b_1) d_t e_t] =$$

$$E(a_0 - a_1)^2 + E(b_0 - b_1)^2 d_t^2 + E(e_t^2) + 2E[(a_0 - a_1)(b_0 - b_1) d_t] + 2E[(a_0 - a_1) e_t] + 2E[(b_0 - b_1) d_t e_t]$$

While we use the assumptions:

$$E(e) = 0$$

$$E(e^2) = \sigma^2 \text{ (from } H!\text{)}$$

$$E(d_t) = d_t$$

we can rewrite the variance to: $\sigma_f^2 = V(a_1) + V(b_1)d_t^2 + 2 \cdot Cov(a_1, b_1)d_t + \sigma^2$

Johnson, Johnson, and Buse (1987, p. 384) claim that after tedious but straightforward manipulations can be written as

$\sigma_f^2 = E(H_t - \overline{H_t})^2 = \sigma^2 \left[1 + \frac{1}{T} + \frac{\left(d_t - \overline{d}\right)^2}{\sum_1^T \left(d_t - \overline{d}\right)^2} \right]$ $T = 44$ the number of measurements.

2. **Manipulations**

To produce the manipulations by ourselves we use the definitions from Rice (1995, p. 513):

$$V(a_1) = \frac{\sigma^2 \sum_{t=1}^{T} d_t^2}{T\sum_{t=1}^{T} d_t^2 - \left(\sum_{t=1}^{T} d_t\right)^2}$$

$$V(b_1) = \frac{T \cdot \sigma^2}{T\sum_{t=1}^{T} d_t^2 - \left(\sum_{t=1}^{T} d_t\right)^2}$$

$$Cov(a_1, b_1) = \frac{-\sigma^2 \sum_{t=1}^{T} d_t}{T\sum_{t=1}^{T} d_t^2 - \left(\sum_{t=1}^{T} d_t\right)^2}$$

So, $\sigma_f^2 = V(a_1) + V(b_1)D_t^2 + 2\cdot Cov(a_1, b_1)D_t + \sigma^2$

$$= \frac{\sigma^2 \sum_{t=1}^{T} d_t^2}{T\sum_{t=1}^{T} d_t^2 - \left(\sum_{t=1}^{T} d_t\right)^2} + \frac{T \cdot \sigma^2}{T\sum_{t=1}^{T} d_t^2 - \left(\sum_{t=1}^{T} d_t\right)^2} \cdot d_t^2 +$$

$$2 \cdot \frac{-\sigma^2 \sum_{t=1}^{T} d_t}{T\sum_{t=1}^{T} d_t^2 - \left(\sum_{t=1}^{T} d_t\right)^2} d_t + \sigma^2$$

$$= \sigma^2 \left[\frac{\sum_{t=1}^{T} d_t^2}{T\sum_{t=1}^{T} d_t^2 - \left(\sum_{t=1}^{T} d_t\right)^2} + \frac{T \cdot d_t^2}{T\sum_{t=1}^{T} d_t^2 - \left(\sum_{t=1}^{T} d_t\right)^2} + \frac{-2 \cdot \sum_{t=1}^{T} d_t}{T\sum_{t=1}^{T} d_t^2 - \left(\sum_{t=1}^{T} d_t\right)^2} + 1 \right]$$

$$= \sigma^2 \left[\frac{\sum_{t=1}^{T} d_t^2 + T \cdot d_t^2 - 2 \cdot d_t \sum_{t=1}^{T} d_t}{T\sum_{t=1}^{T} d_t^2 - \left(\sum_{t=1}^{T} d_t\right)^2} + 1 \right]$$

$$= \sigma^2 \left[\frac{\sum_{t=1}^{T} d_t^2 + T(d_t^2 - \frac{2}{T} \cdot d_t \sum_{t=1}^{T} d_t)}{T\sum_{t=1}^{T} d_t^2 - \left(\sum_{t=1}^{T} d_t\right)^2} + 1 \right]$$

$$= \sigma^2 \left[\frac{\sum_{t=1}^{T} d_t^2 + T\left[\left(d_t - \frac{1}{T} \cdot \sum_{t=1}^{T} d_t) \right)^2 - \left(\frac{\sum_{t=1}^{T} d_t}{T} \right)^2 \right]}{T\sum_{t=1}^{T} d_t^2 - \left(\sum_{t=1}^{T} d_t \right)^2} + 1 \right]$$

$$= \sigma^2 \left[\frac{\sum_{t=1}^{T} d_t^2 + T\left(d_t - \frac{\sum_{t=1}^{T} d_t)}{T} \right)^2 - \frac{\left(\sum_{t=1}^{T} d_t \right)^2}{T}}{T\sum_{t=1}^{T} d_t^2 - \left(\sum_{t=1}^{T} d_t \right)^2} + 1 \right]$$

$$= \sigma^2 \left[\frac{\sum_{t=1}^{T} d_t^2 + T\left(d_t - \overline{d} \right)^2 - \frac{\left(\sum_{t=1}^{T} d_t \right)^2}{T}}{T\sum_{t=1}^{T} d_t^2 - \left(\sum_{t=1}^{T} d_t \right)^2} + 1 \right]$$

$$= \sigma^2 \left[1 + \frac{\sum_{t=1}^{T} d_t^2 - \frac{\left(\sum_{t=1}^{T} d_t \right)^2}{T} + T\left(d_t - \overline{d} \right)^2}{T\left(\sum_{t=1}^{T} d_t^2 - \frac{\left(\sum_{t=1}^{T} d_t \right)^2}{T} \right)} \right]$$

$$= \sigma^2 \left[1 + \frac{\sum_{t=1}^{T} d_t^2 - \frac{\left(\sum_{t=1}^{T} d_t \right)^2}{T}}{T\left(\sum_{t=1}^{T} d_t^2 - \frac{\left(\sum_{t=1}^{T} d_t \right)^2}{T} \right)} + \frac{T\left(d_t - \overline{d} \right)^2}{T\left(\sum_{t=1}^{T} d_t^2 - \frac{\left(\sum_{t=1}^{T} d_t \right)^2}{T} \right)} \right]$$

$$=\sigma^2\left[1+\frac{1}{T}+\frac{\left(d_t-\overline{d}\right)^2}{\sum_{t=1}^{T}d_t^2-\frac{\left(\sum_{t=1}^{T}d_t\right)^2}{T}}\right]$$

$$=\sigma^2\left[1+\frac{1}{T}+\frac{\left(d_t-\overline{d}\right)^2}{\sum_{t=1}^{T}d_t^2-T\left(\frac{\sum_{t=1}^{T}d_t}{T}\right)^2}\right]$$

$$=\sigma^2\left[1+\frac{1}{T}+\frac{\left(d_t-\overline{d}\right)^2}{\sum_{t=1}^{T}d_t^2-T\left(\overline{d}\right)^2}\right]$$

$$=\sigma^2\left[1+\frac{1}{T}+\frac{\left(d_t-\overline{d}\right)^2}{\sum_{t=1}^{T}d_t^2-2T\left(\overline{d}\right)^2+T\left(\overline{d}\right)^2}\right]$$

$$=\sigma^2\left[1+\frac{1}{T}+\frac{\left(d_t-\overline{d}\right)^2}{\sum_{t=1}^{T}d_t^2-2T\overline{d}\,\frac{\sum_{t=1}^{T}d_t}{T}+T\left(\overline{d}\right)^2}\right]$$

$$=\sigma^2\left[1+\frac{1}{T}+\frac{\left(d_t-\overline{d}\right)^2}{\sum_{t=1}^{T}d_t^2-2\overline{d}\sum_{t=1}^{T}d_t+T\left(\overline{d}\right)^2}\right]$$

$$=\sigma^2\left[1+\frac{1}{T}+\frac{\left(d_t-\overline{d}\right)^2}{\sum_{t=1}^{T}\left(d_t^2-2\overline{d}\cdot d_t+\left(\overline{d}\right)^2\right)}\right]$$

$$=\sigma^2\left[1+\frac{1}{T}+\frac{\left(d_t-\overline{d}\right)^2}{\sum_{t=1}^{T}\left(d_t-\overline{d}\right)^2}\right] \qquad \square$$

Remember $T = 44$ and $\overline{d} = \frac{\sum_{t=1}^{44} d_t}{44}$. So, d_t in the numerator is the only changeable value. The greater $d_t - \overline{d}$, the greater the forecast error. Hence, σ_f is hyperbolic (Figure 1) with asymptote $H_t = \left(d_t - \overline{d}\right)\sqrt{\frac{\sigma^2}{\sum_{t=1}^{T}\left(d_t - \overline{d}\right)^2}}$.

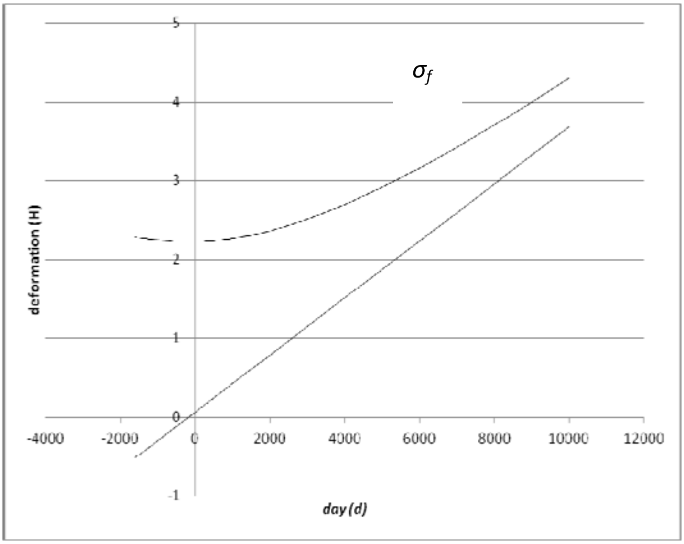

Figure 1. Hyperbolic σ_f of deformation during time in days.

This implies a nonlinear safety margin for the dyke deformations (Figure 2).

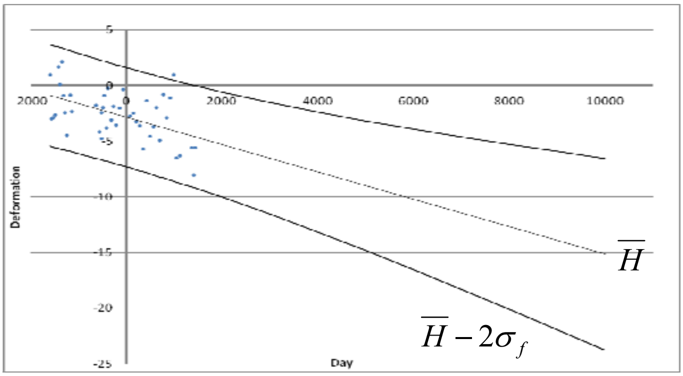

Figure 2. Nonlinear safety margin for the dyke deformations.

Appendix B

TABLE OF CONTENTS OF THE INSTRUCTIONAL UNIT STATISTICS AS BRIDGE BETWEEN MATHEMATICS AND SCIENCE

Instructional unit: Statistics as Bridge between Mathematics and Science – Table of contents

1 The sports physiologist and statistics

1.1 Measuring condition (pp. 5-10)

Introduction of the authentic professional practice of a sports physiologist, who advises clients on how to improve their physical condition. Students learn to measure their fellow students' heart rates and to model their data to find a measure for their physical condition.

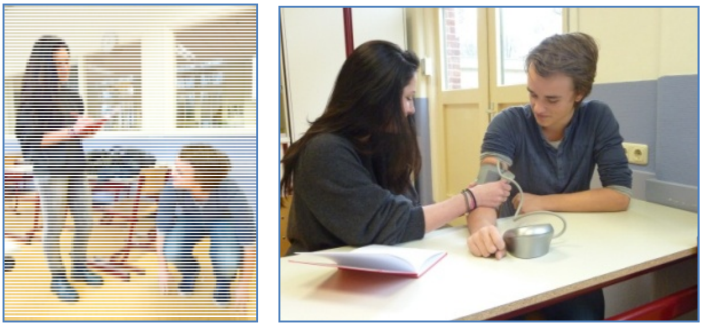

Figure 1. Students measuring physical condition using a sphygmomanometer (right figure).

1.1.1 Scatterplots

Introduction of scatterplots to visualize the relation between two variables.

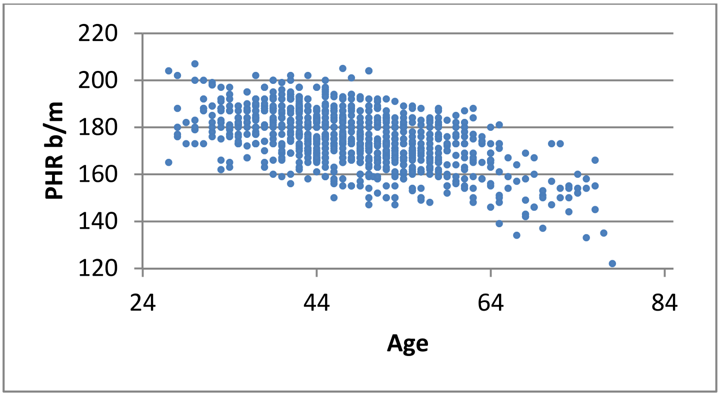

Figure 2. Scatterplot to visualize the relation between Age and the peak heart rate.

1.2 The role of statistics in improving condition (pp. 6-18)

1.2.1 Heart rate and condition

Physiological background information on the relation between heart rates and physical condition.

1.2.2 The threshold point

Background on the threshold point as the heart rate someone can maintain over a longer period of time without the muscles reaching the metabolic threshold (between aerobic and anaerobic metabolism). Also background on training schemes based on the threshold point.

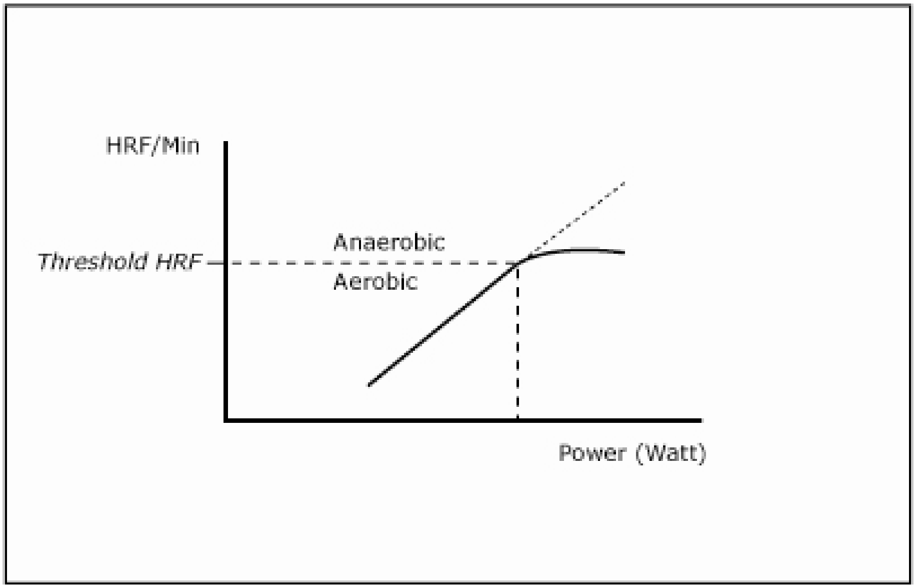

Figure 3. Model to predict the threshold point.

1.2.3 General equations and variation

Introduction of some common formulas for the peak heart rate depending on age). Students compare these with their own modelling results; an informal introduction to variability.

1.3 Statistics and individual improvement of condition (pp. 19-23)

Tasks for students to reason about how to determine their own threshold points in a gym.

1.3.1 Model of the heart rate

Introduction of the model of the relationship between the heart rate and power output when using a treadmill. Tasks to understand the model.

1.3.2 The Conconi Test to determine the threshold point

Activity for students with the Conconi Test to determine the threshold point of a fellow student using linear regression. A regression line is drawn by eye-balling.

1.4 Scatterplot (pp. 24-35)

Introductory tasks about variability in relation to scatterplots.

1.4.1 Data variation

Tasks to investigate variables other than the heart rate in relation to physical condition. Students practise with scatterplots and have to reason about variability.

1.4.2 Correlation and regression

Correlation (positive, negative and absent) is introduced and students practise with it informally.

In the last two research cycles we included at this point a sampling task to confront students with variability and "shuttling" between the contextual and statistical spheres.

2 Statistics and water management

The professional practice of monitoring dyke heights is introduced. Students see a video with accompanying questions and get information on dykes constructed in order to prevent flooding (pp. 36-38).

2.1 Monitoring heights of dykes (pp. 39-42)

2.1.1 How do you decide when to heighten a dyke?

Introduction of scatterplots as a means to decide when to heighten a dyke. Students are asked how such scatterplots can help and to reason about the variability around the regression line.

2.1.2 Levee Patrollers

Students learn how to recognize and explain problems with dykes. When in practice data analysis predicts problems at a dyke position, a Levee Patroller is sent to investigate the situation. Students play an interactive game which is used to train real Levee Patrollers. This game involves a virtual dyke with a lot of problems and heavy weather. The intention is to detect, report and explain the problems.

2.2 Regression lines (pp. 43-50)

2.2.1 Recognizing a trend

Students get real data of dyke heights and are asked to find a trend.
Again students have to reason about regression lines and variability around regression lines in relation to this professional practice to draw inferences about when to heighten a dyke.

2.2.2 Regression lines and Excel

Students learn to use Excel for plotting scatterplots, and learn to produce relevant basic calculations with Excel, such as sum, mean etcetera.

2.2.3 Regression lines and residuals

Students learn to use Excel to calculate residuals.

2.2.4 A measurement for variation

Students learn to calculate the standard deviation using Excel.

2.2.5 Variation and safety margin

Students reason about safety margins and how to include them in their Excel drawings and calculations.

2.3 Regression coefficients (pp. 51-70)

2.3.1 The central point

Student learn to use Excel to determine the central point $(\overline{d}, \overline{H})$ of correlated data with d as number of days and H as deformation of heights.

2.3.2 Calculations using sigma-notation for summation

Introduction of the sigma notation ($\sum$). Student practise with calculations with sigmas.

2.3.3 The least square method

Student use special software (TI-Nspire) to draw a scatterplot and have to draw the best possible line. Using the sum of squares option of the software they improve their regression line before using the option of showing the software's regression line. Students have to understand and reason about the least square method.

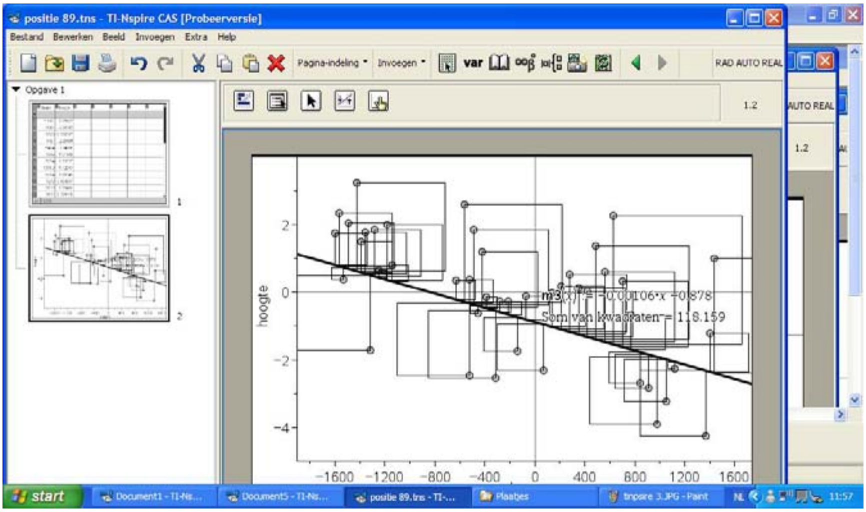

Figure 4. Screenshot TI-Nspire

2.3.4 *Regression coefficients*

Students practise with derivations. Students learn how to find a system of normal equations by calculating the derivatives of the sum of squares. By solving the system the regression coefficients are determined.

2.3.5 *Excel and regression coefficients*

Students learn how to use Excel to produce the regression coefficients.

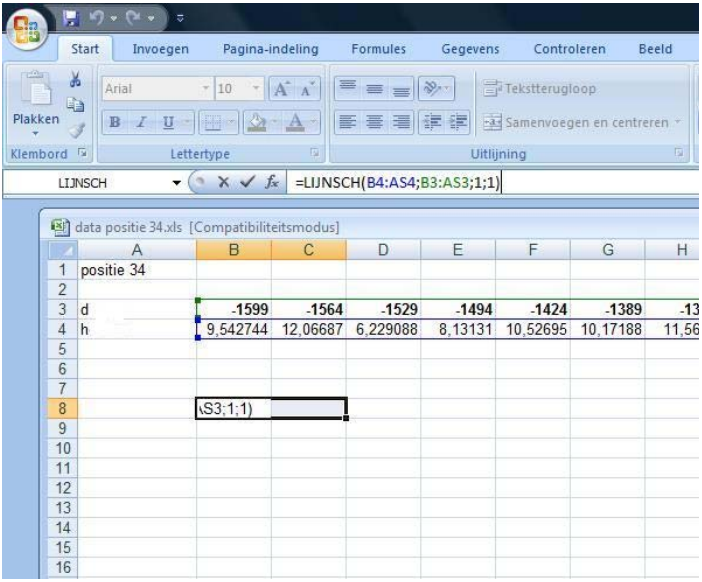

Figure 5. Screenshot Excel to instruct the students how to calculate the regression coefficients with the Linest function (Dutch: Lijnsch).

2.4 Correlation coefficient (pp. 71-76)

2.4.1 *A measure for a relation*

Mathematical background on correlation and how to use Excel to calculate the correlation. In the last task of this section the students get data of three dyke positions. They are asked to decide which position has to be heightened first (there is only money to heighten one position). When students draw scatterplots it is obvious that one position could wait. To decide which of the two other positions has to be heightened they first have to use correlation and regression.

In the last two research cycles we included at this point another sampling task to confront students with variability and "shuttling" between the contextual and statistical spheres.

2.5 Reflection (p. 77)

In this section we asked the students to make connections between the first two chapters. We asked explicitly how the techniques learned in chapter two could be used in chapter 1 where they constructed regression lines by eye-balling.

3 The role of correlation and regression in laboratories

3.1 Calibration (pp. 78-87)

Introduction of the professional practice of calibrating thermometers. Students practise with "precision" (small variability) and "correctness" (correct average in repeated measures). Students learn when it is possible to calibrate an instrument and the role of correlation and regression. In this chapter there are no examples of how to use correlation and regression. Students can show whether they are able to apply their knowledge in this new situation.

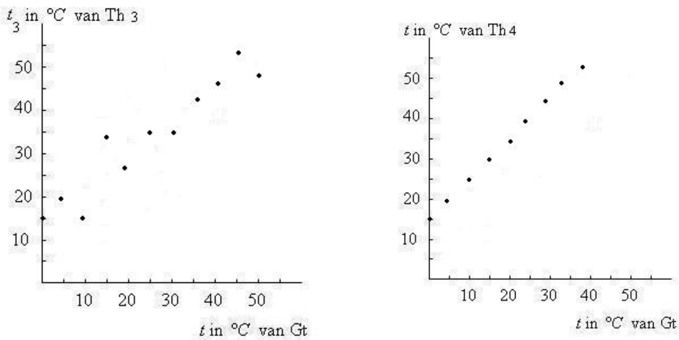

Figure 6. Pictures to show the difference between precise and correct.

3.2 Extra Tasks (pp. 88-93)

Some tasks to use correlation and regression in activities from other professional practices in laboratories.