# EDUCATIONAL VERSIONS OF AUTHENTIC PRACTICES AS CONTEXTS TO TEACH STATISTICAL MODELING

Adri Dierdorp, Arthur Bakker, Jan van Maanen and Harrie Eijkelhof
Freudenthal Institute for Science and Mathematics Education,
Utrecht University, Utrecht, the Netherlands
a.dierdorp@fi.uu.nl

*For the teaching of modeling correlation and regression an 11th-grade unit was designed on the basis of authentic practices in which practitioners use linear regression models. These practices include identifying suitable sports training programmes, dike monitoring and calibration of measurement instruments. The question we intend to answer is how educational versions of such practices can support students in learning to make data-based predictions in authentic situations. For example, the authentic practice of dike monitoring is used to motivate students to study scatter plots of dike height over time and learn how to model a regression line for finding a possible trend to predict when the dike needs to be heightened. The analysis of three teaching experiments shows that most students worked eagerly and with good results, but many continue to struggle with coordinating statistical and contextual knowledge.*

## INTRODUCTION

In the research reported here we take seriously Moore's (1992, p. 15) observation that data are "numbers with a context." We have analyzed how practitioners in relevant authentic (out-of-school) contexts make use of statistical techniques such as correlation and regression to make an informed design of an instructional unit on correlation and regression. This approach of using authentic practices as the source of inspiration for instructional units has been used in science education with the effect that students see the relevance of their learning activities and experience a 'need to know' (e.g., Westbroek et al., in press), in such a way that they properly link scientific concepts to contexts. The purpose of this paper is to report how our unit, which was based on authentic practices, with an accompanying theoretical rationale for the teacher, supported 11th-grade students to make data-based predictions with help of linear regression. In particular, this paper addresses the question of how the unit and teacher supported students in making predictions about an authentic problem, specifically when to strengthen or heighten a dike.

## AUTHENTIC PRACTICES

Over the past decades various approaches to context-based science and mathematics education have been proposed and tested. Gilbert (2006) classified four canonical models of context and argues on theoretical grounds that the model of presenting context as an authentic practice is the most promising one. Several researchers (Edelson, 1998; Prins et al., 2008; Westra, 2008) investigated how elements from authentic practices could be converted to elements of a learning strand in which students experience the concepts learned and the learning itself as relevant. Elements of an authentic practice are motives to perform certain actions, as well as procedures and knowledge used as tools to achieve particular goals. Another goal of the aforementioned researchers was to link students' knowledge of context to the concepts which they have to learn because this is a persistent problem in education.

Westbroek et al. (in press), for example, used the practice of water quality testing to engage students in the question of how drinking and swimming water are tested in practice. The unit relied on a procedure similar to the authentic procedure. This helped students to predict the next rational step in testing water quality. The students worked with measurement tools like those used in actual practice, though simpler. In this way, Westbroek et al. designed an educational version of an authentic practice that gave students a clear goal and made chemical knowledge to be learned meaningful to them. Prins et al. (2008) 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 involves students in modeling meaningful processes. The emphasis in Prins et al.'s teaching unit is on chemistry, and the statistical knowledge of regression required to model water treatment processes was touched upon only marginally. Our study

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php 

deliberatively focused on statistical modeling at the service of data-based conclusions and predictions.

## METHODS

Because there was neither a teaching unit that fit our criteria, nor an instructional theory on how to support the intended learning, we had to design a unit and accompanying rationale ourselves. This required an iterative design process common in educational design research (Van den Akker et al., 2006). To test the suitability of the unit, designed to support student reasoning, we have so far carried out three design cycles (N=12, 6, and 18). All students were from Grade 11 of the pre-university track. To monitor the students' ability to 'see the goal,' their understanding of the contextual problem, statistical skills, and the ability to make inferences, their work in class was observed and recorded (audio and video). Data collection further included written student work, field notes, interviews (at the site and afterwards), questionnaires, and tests.

We constructed a hypothetical teaching and learning trajectory (HTLT, cf. Bakker, 2004), which is used as a link between educational theory and teaching practice. The results of comparing this HTLT with students' actual teaching and learning trajectory (ATLT) give arguments to support or revise the unit and accompanying rationale (HTLT).

The first chapter of the unit is based on the practice of sport instructors and physiotherapists who have to originate the best training program for clients. Apart from convincing students of the need of statistics in such an authentic practice, this chapter also serves to let students get acquainted with collecting data themselves and to make them familiar with scatter plots, variability, and correlation (in a qualitative sense). The key instructional activity in Chapter 2 is to make a prediction about the moment a dike should be heightened. To achieve this, we designed a four phase progression in the teaching materials. In the first phase students speculate about possible criteria for producing a regression line. The second phase deals with the theoretical background of regression (including rules for summation of ∑-signs). In the third phase the students learn how to make scatter plots, how to calculate the formula of a regression line, and how to calculate correlation coefficients using a spreadsheet. In the fourth phase the students practiced the skills they had developed, initially in the context of dike monitoring, later in that of calibration so we could study if they could transfer what they had learned to other contexts.

More general HTLT predictions are that students would 1) know why they have to do certain activities, 2) understand the contextual problem, 3) understand the existence and role of variability, 4) be able to recognize a linear trend and use a spreadsheet to compute regression coefficients and correlation, 5) learn about the mathematical background of regression, and 6) make an inference beyond the data. In the analysis for this paper we were especially interested in how students learned to make a data-based prediction, linking statistical techniques to the contextual problem. A final assessment was used to gauge how much they had learned about correlation and regression in authentic problems, with one item being a dike monitoring problem.

## RESULTS

From the qualitative analysis of the second design cycle we chose three sections to illustrate the support needed by students when they made a data-based prediction to solve a contextual problem. For each section we compare the hypothetical and actual teaching-learning processes, and reflect on the question of which support students needed to make informal statistical inferences about a contextual problem.

We jump to the stage in Chapter 2 at which students were able to produce scatter plots in Excel (see Figure 1). This authentic data set, stemming from Delft University of Technology (Dentz et al., 2006) shows the deviation of the height of a certain dike at a particular location over a few years (here from day -1599 until day 1435).

In line with the HTLT, the teacher asked the students to give reasons for the variability of the deviation of the height of the dikes, and thus treat the data as Moore's "numbers with a context." Because the students had to deal with such issues in other subjects (e.g., geography, chemistry, and biology) we predicted that they would give answers about ground movements and vegetation on the dikes, or perhaps measurement errors. The ATLT showed that this was not the case, but at least the students interpreted the data in contextual terms. The analysis shows that students were engaged and it confirms that students linked the data back to the context, but the teacher had to spend considerable time on explanations such as ground movements and vegetation on dikes. We considered such knowledge relevant, because we wanted students to understand the variability in the data when they had to draw their own conclusions at a later stage (cf. Wild & Pfannkuch, 1999).

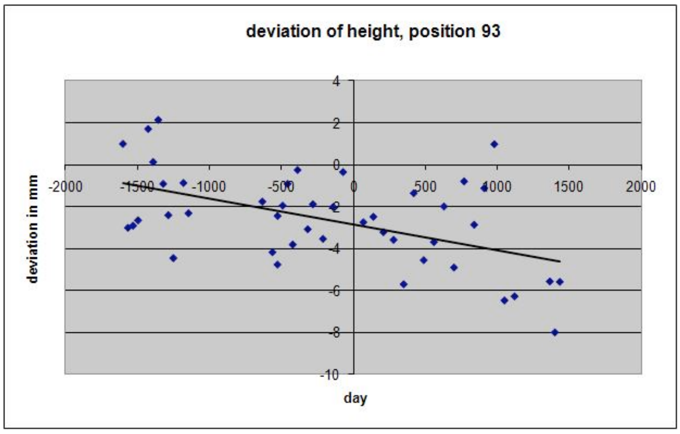


Figure 1. Deviation of the heights of a dike at position 93 during 3034 days

Despite the problems reported in the literature on seeing trends in scatter plots (e.g., Hoyles et al., 2007; Noss et al., 1999), we assumed that these students would see a trend. We also expected that students would see that variability affects the prediction about the moment to heighten the dike. In other words: we predicted that students would take into account that the lower values calculated by the regression line minus a safety margin were the relevant ones in this context, and that a straightforward linear regression would not fit the bill.

To illustrate that we had been too optimistic, we mention a student who just observed that not all the points were close to the regression line. The teacher heard this remark and used it to lead the conversation to variability. Despite some students' awareness of the variability, none were ready at this stage to make the kind of prediction we were after. This required additional discussion. In line with the HTLT all students, except Ida, one of the girls in class, saw a trend. She claimed that there was no real trend, and that a spread of 20 mm is of no importance within this context. Another girl, Joan acknowledged a trend, but Ida remained skeptical. She seemed to find the variability too big for a reliable prediction and said: "If you look at for instance -1500 and zero they are more or less at the same spot, but the line there is also drawn downward. I think that is not precise." We conclude that although most students perceived a linear downward trend, the noise of the variability was too big for Ida to believe there was a signal of a real change in dike height. She might have needed a learning activity similar to the one used by Noss et al. (1999) in which box plots of vertical slices of the scatter plot were compared. This convinced the nurses in Noss et al.'s study that there was a trend where they initially had not seen one.

The classroom discussion proved a useful starting point for questions about the slope and position of the regression line, because the students only knew that the spreadsheet program can draw regression lines, but not how it does so. Carla asked: "Isn't the line supposed to touch upon the highest number of dots?" The teacher used this question to ask how regression lines should be determined. First, some students said the regression line is a kind of average. But when the teacher asked them what they meant by that, they could not further articulate their answer. That is why the teacher asked Ida to draw a line where she expected the regression line to be on the smart board. She drew the line in the upper region of the scatter plot (see Figure 2), after which the teacher made one closer to the center of the scatter plot. Jan did not agree and drew a third regression line under the two other lines (see Figure 2 for a reconstruction).

The teacher used these three lines to develop collaboratively some rules to draw regression lines in general; this was before the students learned how the spreadsheet computes such lines. In this way, the teacher (in line with the HTLT) intended to have the students see the point of learning how regression lines have to be computed irrespective of a particular context. Despite his attempt to discuss this in general, Ida brought in the context again:

T: We are figuring out why and how you need to draw a line like this. Jan, I criticize your line, and I think I do that in a correct way. Do you agree, Ida?
Ida: Well, just like your line, I do criticize that too.
T: Oh, tell me!
Ida: You draw it in an extreme way. Actually, you're trying to frighten people a bit. If it concerned that dike, with this subsidence, then you would draw the utmost of the utmost as the most.

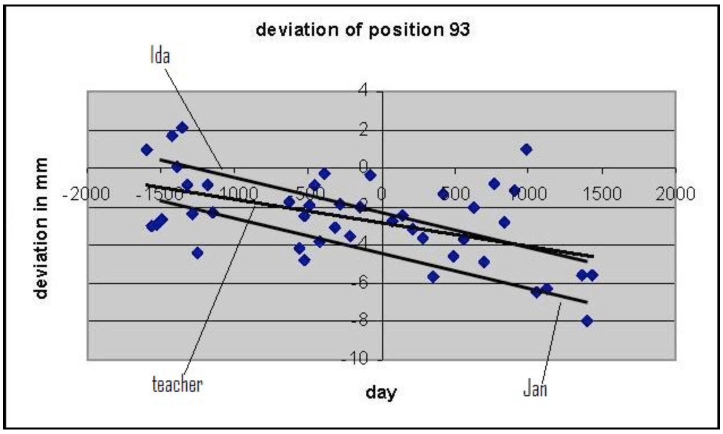


Figure 2. Deviation of the heights of a dike at position 93 during 3034 days, two students and the teacher draw a line to predict the graph from the regression line

Despite the effort of the teacher to support the students to reason about the regression line, the students linked their ideas to the contexts, for example in Ida's last intervention. Ida thought that if you draw the line too low, then people become frightened and there is no reason for it, because 20 mm is not a big deal. Joan took a more forward looking view. She recognized a trend and told Ida that 20 mm now, becomes meters in the future. Ida looked at the data and had difficulties in making generalizations, but linked explicitly the data and the context. Joan also linked the data and the context but was able to make a prediction beyond the data. She did foresee the necessity of making a context sensible prediction to prevent flooding problems in the future because she recognized a trend. All students, except Ida, agreed with Joan.

The more general point is that it seems hard for the students to switch between a contextual problem (once they engage with it) and a more abstract discussion about the technique. The role of the teacher is crucial in these cases; s/he has to focus and lead the classroom discussions in such a way that students come to see the importance of context-independent rules.

At the end of the unit the students had to do a test. For one of the tasks, concerning another dike, students had to predict when the dike needed to be heightened. In the HTLT we expected students to find the formula for the regression line and to take variability into account in predicting the day to heighten the dike. We also expected that students would give contextual reasons for taking the variability into account, and that the risk of flooding would require a safety margin of more than 1 standard deviation (SD).

The students in the second design cycle had practiced questions like this, but there had been no time for the teacher to thoroughly discuss this question with the students. All six students found a correct formula for the regression line and used this formula for their prediction, but none of them took variability into account; they did not subtract a specific margin to minimize the risk of flooding. We therefore decided to spend more time on the contextual consequences of variability in the third design cycle. Within this cycle we used a whole extra 50 minute lesson to let the students reason about the phenomenon of variability. The next example comes from this lesson in the third cycle. Using a spreadsheet the students had drawn a scatter plot with the regression line.

The teacher then asked the students to draw a zone around the regression line by adding and subtracting (initially) one standard deviation to the regression value. The students had to determine the moment when the dike becomes too low. The example below shows that students used the regression line and did not take the margin into consideration. One student argued that he and another student used the lowest line (regression line – 1 SD) because of the safety margin. When the teacher asked whether 1 SD was enough as a safety margin, some students referred to the rules of normal distribution (about 68% within a 1 SD margin and about 95% within a 2 SD margin). We present one example, where students drew conclusions about a particular data set, taking that variability into account. In this case the students were told that -15 mm was a critical deviation.

S3: We used the bottom line because you must subtract 1 SD.
T: Why?
S3: You have to take a margin of safety. You have to take the lowest value.
T: Is this safe enough?
S4: No the chance is only 68%.
T: A chance of 68% that there will be no problems?
S5: No, 100 - 68% = 34%. (S5 made a small calculation error)
S6: Above the line is also no problem, so it is 16%.
T: What do you take into account?
S6: Variability.

Although we had hoped that students would take a larger safety margin of at least 3 SD, the third design cycle was more successful in this respect than the second design cycle. We expected that the test results would be better, because we explicitly claimed more time in this third design cycle to let the students reason about variability and its effect on making a prediction. When the students made the same test as the students in the second design cycle, six out of eighteen took variability into account. This was a small improvement compared to the results of the second cycle where none of the students took variability into account when making an inference about the moment to heighten the dike. However, this improvement cannot be attributed only to the increased support during the third design cycle: The third group also had a better technical background than the second group.

From these last findings we conclude that we had underestimated the difficulties students would have with risk (see Eijkelhof, 1996; Pratt et al., 2009). Drawing a valid inference—a prediction of when to heighten a dike—not only requires coordinating knowledge about regression lines and about the context of the data, but also of risk: There is a small probability that dikes are relatively low and water levels high, and such a situation can have a disastrous effect; hence dike heights should stay well above the critical value.

## CONCLUSION

This paper addressed the question of how the unit and teacher supported students in making predictions about an authentic problem, specifically when to strengthen or heighten a dike. We found evidence that the unit based on authentic practices, with the teacher following the accompanying rationale (HTLT), provided students with opportunities to reason beyond correlated data and provoked them to make predictions. In the second chapter of the unit students inferred before which date a dike needed to be heightened. To achieve this, they had to make a scatter plot and had to arrive at the formula for the regression line. In order to stimulate students' linking such a formula with a real-world problem, it proved necessary to stimulate discussions beyond straightforward observations. The teacher had to ask many 'how and why' questions to stimulate the students to reason about contextual problems. This took more time than expected, but also provided students with deeper understanding. Questions such as predicting the time of action to heighten the dike stimulate students to coordinate their contextual knowledge and what they learned about correlation and regression, and the students were able to make the desired calculations for the regression formula and the correlation coefficient. However, the actual teaching and learning trajectories showed that student awareness of the goal of these activities needs more explicit attention, especially in classroom cultures in which word problems and 'bare' mathematical exercises are the norm. The test in the last design cycle showed that a number of students took the variability into account, but there were only six correct scores out of eighteen.

Taken together, the results of this study suggest that it is possible, under favorable conditions, to teach the modeling of correlation and regression with educational versions of authentic practices, in such a way that students make sensible predictions. A second major finding was that most students appreciated the way the unit was based on authentic practices and that this led them to understand why they had to learn correlation and regression. The answers on the student questionnaires, to be reported in a future paper, tend to confirm this. The third result of this investigation shows that although we would expect that students take the context into account when they draw conclusions from a regression analysis, observations show that this does not happen without specific attention in the lessons. Most students are able to make the link, but some stay within the realm of formal procedures, apparently a persistent problem; perhaps we were too ambitious? The context about monitoring the heights of dikes is perhaps too complex, but this holds for many contexts, especially those that involve risk. These findings emphasize the importance of designing exercises when explicitly asking students to reason beyond formal procedures. The findings also show that the use of an authentic practice offers opportunities to find meaningful exercises for students. Consequently, the teacher must be well focused and prepared to structure the discussions in class. Sufficient time and careful classroom management are crucial, next to an appealing carefully designed educational unit.

REFERENCES

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

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

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*(3/4), 273-278.

Gilbert, J. (2006). On the nature of "context" in chemical education. *International Journal of Science Education*, 28(9), 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*(4), 331-360.

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: The Mathematical Association of America.

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

Pratt, D., Ainley, J., Kent, P., Levinson, R., Yogui, C., & Kapadia, R. (2009). *Informal inferential reasoning as a contextualised risk/based enterprise*. Paper presented at the Statistical Reasoning Thinking & Literacy, Brisbane.

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

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

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

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

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223.