Chapter 32

How to Utilize a Classroom Network to Support Teacher Feedback in Statistics Education

Jos Tolboom & Wilmad Kuiper

SLO • Netherlands institute for curriculum development

Tolboom, J., & Kuiper, W. (2013). How to utilize a classroom network to support teacher feedback in statistics education. In T. Plomp, & N. Nieveen (Eds.), *Educational design research – Part B: Illustrative cases* (pp. 665-692). Enschede, the Netherlands: SLO.

slo

Contents

32. How to utilize a classroom network to support teacher feedback in statistics education

Credits

32. How to utilize a classroom network to support teacher feedback in statistics education?

Jos Tolboom & Wilmad Kuiper

Abstract

By decreasing the amount of contact time during the past fifteen years in mathematics education in senior secondary education in the Netherlands the issue of organizing this contact time more efficiently has become increasingly relevant. Research shows that feedback is one of the most powerful tools for enhancing learning achievements. At the same time educational technology emerged in the field. Especially mobile technology seems to be promising. The main research question to be answered in this chapter is: "What are the possibilities of a wireless classroom network to support upper secondary statistics teachers in giving feedback?" An intervention was designed and evaluated that tries to utilize a classroom network of graphing calculators for the sake of feedback. The intervention was (re)designed and evaluated in four empirical educational design research cycles. The study was carried out in the domain of descriptive statistics. The study explicates the conditions to be fulfilled for the intervention to be successful.

1. Introduction to the problem

What is the problem in context?

The inducement of this study was a 1998 curriculum reform in senior secondary education in the Netherlands. This reform encompassed new learning goals and contents, reallocation of the aimed study load as well as pedagogical changes. The suggested pedagogical change aimed at students working self-regulated, deliberating in small groups and/or working individually on problems, with the teachers walking amongst them, available for consultation. This represents a change in the role of the teacher from 'a sage at the stage' to 'a guide by the side'.

In order to create time for teachers' coaching activities, school managers often decided to reduce time for the traditional face-to-face teaching practices. Teachers and students experienced an increase in workload (Van Streun, 2001). But there are more reasons than just reallocation of teaching time to embark on this study. Mathematics, the overarching learning domain of this research, is perceived by students to be difficult (Berch & Mazzocco, 2007; Geary, 2010; Hembree, 1990; Küchemann, 1981; Rosnick & Clement, 1980). There is a variety of problems that students may encounter when learning mathematics, such as: *dyscalculia* (McCloskey, Caramazza, & Basili, 1985), *innumeracy* (Paulos, 1988) and *mathematics learning disability* (or: *deficit*) (MLD) (Ginsburg, 1997). Especially poor performing students in mathematics consider statistics, the specific learning domain of this study, to be difficult (Hong & Karstensson, 2002). One interpretation of ‘difficult’ can be ‘difficult to study without any help’.

The question then rises: What can we add to mathematics education to support students in understanding this difficult subject? This question prompted the study reported in this chapter.

Why is EDR an appropriate research approach?

The aim of this study was to find out *if*, and if so, *how,* the supply of feedback in mathematics education can be improved by using a wireless classroom network (CN) of graphing calculators (GC). Several authors (Kelly, 2003; Van den Akker, Gravemeijer, McKenney, & Nieveen, 2006; Van den Akker & Kuiper, 2008) argue that educational design research (EDR) is appropriate for investigating this kind of improvement processes. Our main research question starts with *what*, implicating a *how*, that is: exploring a mode. After having underpinned this mode, we eventually tried to find out *why* our intervention succeeded or failed with respect to its goal. EDR is supposed to be a logical paradigm for this type of research. Kelly (2007) states that design research is most appropriate for *open wicked* problems (Rittel & Webber, 1973). Our problem is *wicked*, because feedback, although a classical theme in learning science, is still not completely understood (Cohen, 1985; Shute, 2008) and not very well structurally implemented in classroom practice. Our problem is *open* since it is highly unlikely that there is just one way to just one right answer. What is more, the technology we explored and its use were quite new, so we presumed we needed several iterations in order to create a teaching-learning setting specific enough to yield data that could possibly lead to an answer to our research question.

What kind of EDR is applied in this study?

Nieveen, McKenney and Van den Akker (2006) distinguish the nature of EDR in two, more or less complementary, approaches:

1. validation studies, aiming at developing, elaborating, and validating theories about both the process of learning and the resulting implications for the design of learning environments;
2. development studies, aiming at the derivation of design principles for use in practice.

The study we describe falls into the second category. How did we apply the approach of EDR, eclectic in its nature (Gravemeijer, 1994), in this research? After having sharpened our focus with an initial study, we conducted a literature study with respect to feedback, statistics education and information and communication technology (ICT). Next, we formulated principles that guided the design of a prototype of the intervention. We conducted a content analysis of two predominantly used mathematics textbooks with respect to whether these textbooks focussed strong enough on 'data literacy'. With the design principles, and the conclusions from the content analysis, we developed a prototype including corresponding so-called 'hypothetical teaching trajectories' (HTT). The design of the prototype was reviewed by field experts focussed on the question 'Does this prototype facilitate the teaching and learning of 'data literacy' with an emphasis on feedback?' We adjusted the prototype to their findings. We evaluated the prototype in classroom practice using student questionnaires and interviews with

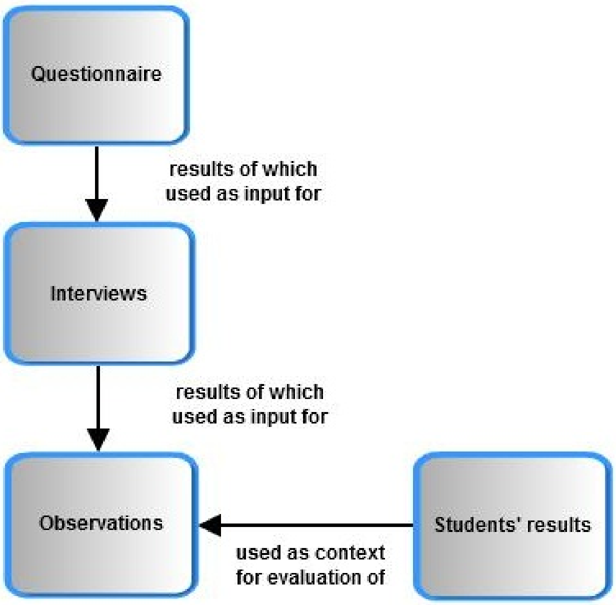

Figure 1: Data sources for this study

teachers and students. Analysis of these data offered input for a revision of the prototype. Again, the prototype was reviewed by experts and again adapted. Following this process the prototype was piloted in practice using the same evaluation techniques: observation, questionnaires and interviews.

The data sources are depicted in Figure 1. The key data source was observation of the classroom discourse. All lessons were observed and videotaped. We used a student questionnaire and a teacher questionnaire. Questionnaires data was used as input for student and teacher interviews. Interview data was, together with the students' results on exercises and a summative test, used to interpret observations data.

Statistics education

Shaughnessy (2010) argues that, from the quantitative techniques represented by the secondary school curriculum, statistics may be the most common element of mathematics used after completion of secondary school. In tertiary education, students use statistics during multiple quantitative courses. Professionals use statistics in their working and even in their private lives. Whether Shaughnessy exaggerated a little while stating that statistics is more often used than, for instance, elementary algebra, or not, we agree with him that statistics is for many students very useful. This usefulness is perhaps the most important argument for the firm place statistics nowadays has in the curriculum. The usefulness, together with the applied character of statistics, attributed to the decision to include statistics in the curriculum for mathematics A (the specific variant of secondary mathematics for those students who aim to study in higher economics education or social sciences). In this study we restricted ourselves to the sub-domain of *descriptive* statistics. And as far as learning activities are concerned the focus was on reasoning and sense making, to be called *data literacy* (DL), and on *algorithmic statistical skills* (ASS) (see review of literature below). In the section Review of literature, we will underpin the relevance of this distinction.

Mobile classroom networks

Abrahamson, one of the first researchers in the field of classroom networks, stated that the networked classroom can support real-time formative assessment for teachers (1999). Formative assessment offers a central place for feedback. Roschelle (2003, p. 263), in an overview of research, noted that: “*these systems* (referring to what he earlier calls ‘classroom response systems’) *enable students to receive much more feedback than normal*”, thus considering it an enhancement of feedback possibilities. He continues (2003, p. 268) with “*Students can perceive receiving much more individualised assessment feedback*…”. These remarks, made after having observed early practices around classroom networks, suggest that formative assessment, feedback and classroom networks are a quite logical combination. These authors apparently see advantages in the use of a classroom network in order to establish formative assessment for the sake of feedback.

Different terminology is used in reported research to describe more or less the same infrastructure from the perspective of functionality: *classroom response systems* (Fies & Marshall, 2006), *student response systems* (Penuel, Boscardin, Masyn, & Crawford, 2007) or *classroom networks* (Roschelle, Penuel, & Abrahamson, 2004). All of these share the same basic idea of using a wireless network of handheld devices. In this study we utilized graphing calculators as handheld devices, because of the mathematical functionality offered. The network allows teachers to provide mathematical content to the students, to gather students' work on this content, to analyse this work and to provide feedback on it.

The classroom network in practice

In order to get a concrete idea of classroom activity, we first describe in terms of ICT-facilities what we see in the classroom when using a classroom network (CN). This is illustrated in Figure 2. The functional interaction takes place between the teacher's computer and the students' graphing calculator (GC). All of the students have a GC at hand as is usual in Dutch upper secondary education. Technically, the GCs are in groups of four, connected to a hub. This hub communicates with an access point that in turn communicates with the teacher's computer. On the teacher's computer there is software that is able to analyse the students' work and to represent this analysis in a slide show. Using the data projector and the projector screen, the teacher can make this analysis available for the students.

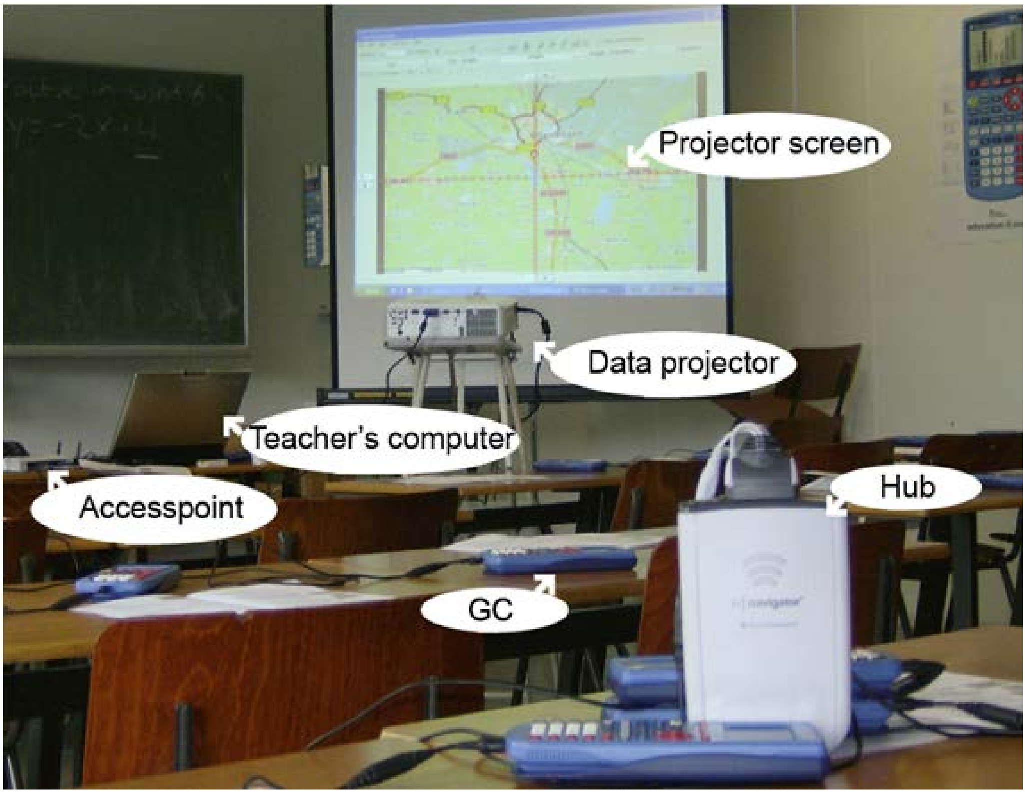

Figure 2: The components of a classroom network in practice

In the last adaption of the network the hubs were replaced by a Wi-Fi cradle for each GC, making the network really wireless.

Research questions

The main research question was as follows:

What is the potential of a wireless classroom network in supporting upper secondary statistics teachers with providing feedback?

Four sub-questions were derived as follows. With respect to the *feasibility* of the chosen approach, we asked ourselves whether the chosen technological configuration (see Figure 2) was adequate for supporting feedback and whether a mathematics teacher was able to utilize this support in order to provide feedback as intended.

Thus, two sub-questions arose from the main research question:

1. *Is technological support by means of the classroom network adequate for the intended feedback in the lessons?* (Conditional question)
2. *Is it possible for a mathematics teacher to implement the prototype in accordance with its intentions?* (Existence question)

With sub-question 2 also the quality of the feedback is investigated, by comparing the realised feedback with the intended feedback. After having distinguished the statistical learning goals, we asked ourselves whether the realised feedback was both on algorithmic statistical skills (ASS) and on data literacy (DL) activities. Therefore, we specified a third sub-question with respect to the general research question:

3. *Is the feedback support of the classroom network equal for ASS and DL?* (Didactical question)

With sub-question 3 we investigate the feedback provided by the network and by the teacher to the students as well as the feedback provided by the network and by the students to the teacher. While realising that teacher behaviour is the first focus of this study, we would like to identify the situations in which feedback was realised as intended and in which situations it was not. Thus, a fourth sub-question emerged:

4. *Which teacher characteristics promoted/hindered the implementation of the classroom network as intended?* (Identification question)

This fourth sub-question is rooted in our sense of the complexity of the intended intervention. A teacher will have to be able to manage a lot of actions simultaneously: the statistics itself and the learning goals, the students' input as collected and rudimentary analysed by the CN, the ICT environment, and the flow of the classroom discourse as initiated by the teacher feedback. A close observation of teacher behaviour should result in data with which this sub-question can be answered.

How has EDR been applied?

In Figure 3 we depict our study in terms of the well-known ADDIE components (Gustafson & Branch, 2002; Molenda, 2003; Molenda, Pershing, & Reigeluth, 1996): Analysis, Design, Development, Implementation, and Evaluation. Reflection *in-action* and *on action* (Van den Akker & Kuiper, 2008) with respect to the quality of both the intervention and the EDR process were part of each phase.

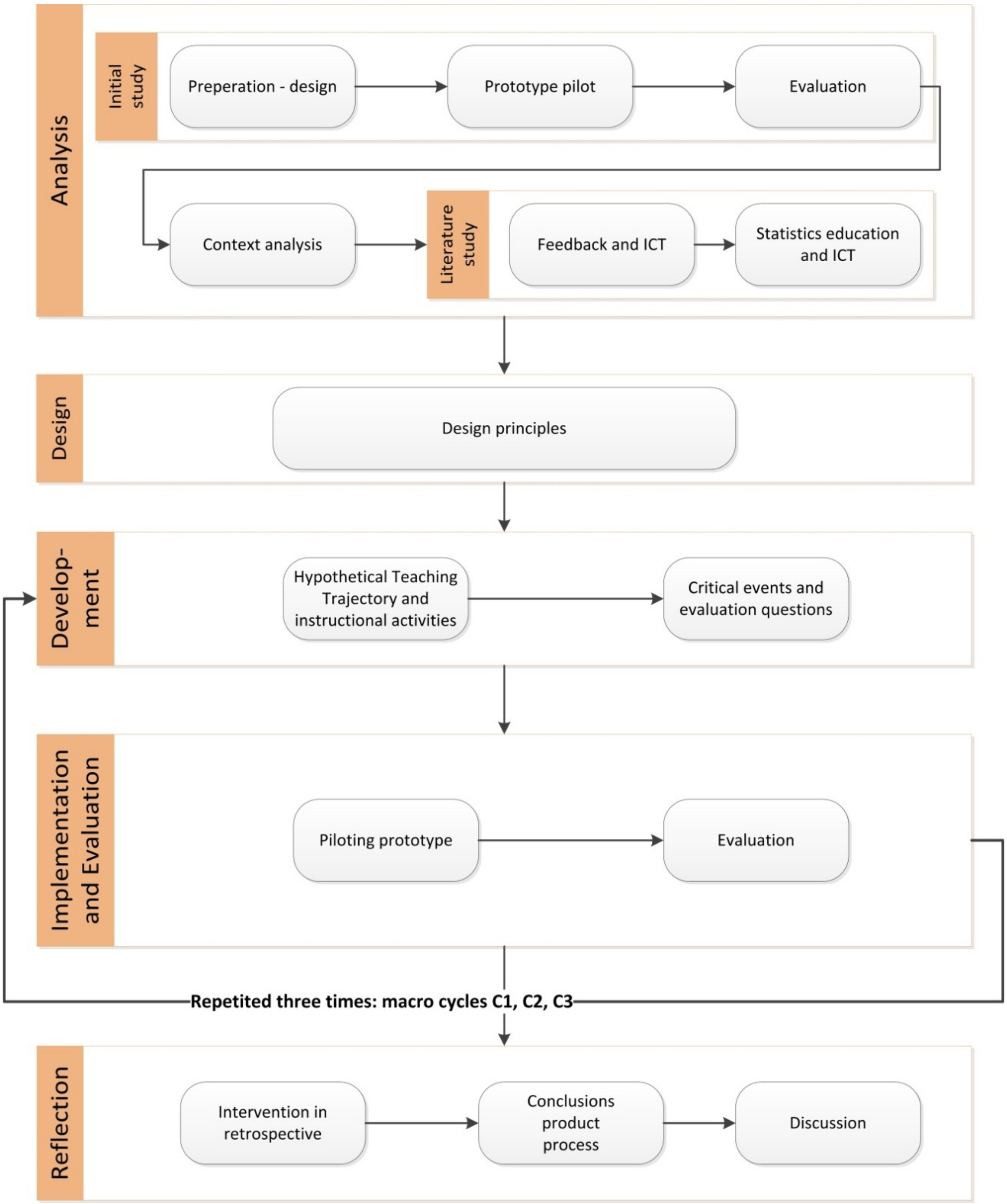

Figure 3: Research components of and stages in this study according to the ADDIE framework

The macro cycles (or: iterations) C1, C2 and C3 were the stages in which the three prototypes of the intervention were piloted. In the next sections, we will describe each of the ADDIE stages this study in more detail. This chapter, however, will just describe our findings and conclusions from the third and last iteration.

2. Analysis

Initial study

An initial small-scale intervention study (one teacher, one class, and a prototype different from the one described in this chapter), was conducted with the purpose to identify what can be observed when a wireless classroom network in mathematics education is utilised. For details on this study, we refer to Tolboom (2005). It turned out that teacher feedback on the students' work seemed the pivot for the intensified classroom discourse. The teacher and students indicated that the use of the classroom network was the facilitator of teacher feedback. The results of this initial study motivated us to set up a study into the potentials of a classroom network to support the supply of feedback in statistics education.

Review of literature

Feedback

Meta-analyses show that, from the perspective of students’ achievement, feedback has a large positive effect (Hattie, 1999; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Marzano, Pickering, & Pollock, 2001). In his synthesis of over 800 meta analyses on learning achievement, Hattie (2009, p. 12), states that “*…the most powerful single influence enhancing achievement is feedback*”. Sadler (1989) underpins the importance of feedback by stating that the acquirement of skills needs practicing, and that practicing needs feedback. This has been supported by subsequent research on the role of feedback for improving knowledge and skill acquisition (Azevedo & Bernard, 1995; Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Moreno, 2004; Pridemore & Klein, 1995). This is important, because statistics education inevitably requires skills. Lovett (2001) concluded that feedback was useful to help students improve their ability to select appropriate data analyses, which is specifically interesting for this study, because in the domain of descriptive statistics analysis of data is the main goal. There seems to be a broad consensus that when wanting to improve student achievement, feedback is a major variable to focus on.

Shute (2008) emphasises the interwovenness of feedback and formative assessment and explicitly focuses on the student as a recipient of feedback, in which she differs from Black and Wiliam (1998). The main goal of feedback is to diminish the gap between the learners’ factual level of knowledge, skills and beliefs, as shown in learners’ work, and the level which is aimed at by the teacher or, more generally, the curriculum. This ‘closing the knowledge gap’ (Ramaprasad, 1983) is an important example of ‘improving learning’ as posed by Shute (2008). Following Shute (2008) and Black and Wiliam (1998), the *goal of the feedback* is improving the learning process and result of *the student*.

What about the difference between Shute’s (2008) use of ‘feedback’ and its use by Black and Wiliam (1998)? Who is giving and who is receiving feedback? Shute’s definition of formative feedback (2008) does not point out which actor is responsible for giving feedback. Hattie and Timperley (2007) use an ‘agent’ for the supply of feedback. As possible agents they suggest: teacher, peer, book, parent, self, and experience. Which of these are the most important for this study? When realizing that giving feedback is part of an interaction (Rapp & Goldrick, 2004), feedback can be based on Moore’s (1989) identification of three types of instructional interactions: between the *learner* and the *instructor*, *between learners* and between *learners* and *the content* they are trying to master.

While focusing on the feedback aspect of interaction, all of these interactions contain interesting aspects for our study. For the use in this study, and based on Moore (1989), the following feedback types were chosen:

1. *teacher* feedback: given by the teacher on the work of the students;
2. *peer* feedback: from student to student;
3. *curricular* feedback: given by (parts of) the (computerised) curriculum materials to the students.

Curricular feedback was restricted to feedback provided through the classroom network. The students' work and behaviour are *at the same time* a source of feedback for the teacher in order to modify instruction. The teacher is in this case also the *receiver* of feedback.

Statistics education

With regard to the goals of statistics education, a distinction can be made between learning activities with a focus on 'data and concepts' and learning activities concentrating on 'recipes'. The former are to be seen as the most important goals of statistics education. They offer the possibility to reason about statistics itself (concepts) and its application (data) in various situations (concepts are transferable and sustainable). On the other hand, there are 'recipes', algorithmic in character, needed to apply statistics in practice. Cobb (1991) advises a focus on 'data and concepts' at the expense of 'fewer theory and recipes'. Nevertheless, working with these 'recipes' (thus: algorithmic activities) is considered to be indispensable for students in order to acquire a sense of where and why 'data and concepts' are needed. Cobb's advice particularly counts for introductory courses in statistics, where intuition should be stressed in order to make the students more receptive to more formal statistical techniques, a possible domain for their subsequent study. Cobb indicates the basic choice when it comes to determining the goals of statistics education. Thus, the learning activities in statistics education should reflect these two sides of statistical activities:

1. reasoning and sense making (Martin, et al., 2009) with and about data (Cobb's 'data and concepts'), to be called *data literacy* (DL);
2. *algorithmic statistical skills* (ASS) (Cobb's recipes).

When comparing both types of knowledge to the characterization by Van Streun (2001), we project ASS on *procedural* knowledge (know *how*; with a little declarative knowledge, know *what*) and DL on *conceptual* knowledge (know *why*).

The category 'Skills' is a rather obvious and traditionally well stressed goal when it comes to statistics education. Gal and Garfield (1997) summarized their own research and that of others on the goals of statistics education. They listed eight goals for statistical education. With respect to 'data literacy', we concentrate in this study on four of the eight goals: 'Understand the purpose and logic of statistical investigations'; 'Develop interpretive skills and statistical literacy'; 'Develop the ability to communicate statistically'; and 'Develop useful statistical dispositions'. The first two are typical for data literacy, the third deserves special attention because the design intervention aims to improve the classroom discourse in statistics classes through teacher feedback on students' work. This is essentially 'communicating statistically'. These dispositions are to be a somewhat implicit part of the intervention, to be made more explicit in the classroom discourse aimed to be induced by teacher feedback. 'Data literacy' is considered to be somewhat less technical than 'statistical literacy'. We consider 'reflection' and 'interpretation' as important elements of 'data literacy'. An emphasis on data and concepts is beneficial for the learning results, as reported by Smith (1998). Moore (1997) also stresses the importance of teaching the statistical concepts. By putting emphasis on statistical concepts, even these students will be able to build up statistical intuition.

This has also been stressed by Watson and colleagues (J.M. Watson, Collis, Callingham, & Moritz, 1995; J. M. Watson, Gal, & Garfield, 1997).

Feedback, statistics education and ICT

One of the historical applications of ICT in education, Skinner's teaching machine (Skinner, 1958), was focussed, amongst others, on feedback. This promise of ICT never left the educational research radar. There are several aspects of feedback, as distinguished for example by Shute (2008), for which the use of ICT offers advantages. The *specificity of feedback* is interpreted as the level of information presented in feedback messages (Goodman, Wood, & Hendrickx, 2004). Goodman and Wood (2004) found that increasing the specificity of feedback during practice, using a computer for feedback delivery, increasingly guided performers to correct responses, thus profiting from the increased level of specificity. They conclude with respect to ICT that the use of ICT has obvious advantages for providing specific feedback on a frequent basis. In our view, this has to do with the general advantages of ICT (speed and scalability). Davis et al. (2005) reported that ICT-supported feedback can distinguish the level of specificity and that high specificity was especially beneficial for students low in learning goal orientation. It is nowadays more easy to present *complex* feedback in an organised way. When using a hypertext structure, like, for example, on the WWW, it is possible to provide hints in a couple of phases. Nevertheless, the advice to be cautious with feedback complexity still remains. The fact that complex feedback can be easily realised with the support of ICT may never be the first reason to actually design and use complex feedback. For the *addressing* of feedback (to the task level, to the meta-task level, to the personal level) ICT offers no specific advantages. Its general advantages, of velocity or scalability, of course, remain. For the *timing* of feedback, ICT offers obvious advantages when it comes to immediate feedback. The immediateness of *personalised* feedback for a whole classroom cannot be realised without the use of ICT. So when wanting to profit from the benefits of immediate feedback for each individual student (Azevedo & Bernard, 1995; Kulik & Kulik, 1988), it seems inevitable to use ICT for the feedback delivery. Numerous studies have reported learning gains in working this way (Hannafin, Philips, Rieber, & Garhart, 1987; Phillips, Hannafin, & Tripp, 1988).

For the *presentation* of feedback, ICT offers specific possibilities. Think of simulations, animations or serious games. There is nevertheless no research reported in which ICT plays a leading role for the presentation of feedback. Even in the field of serious games there has been no serious research conducted on the possibilities of feedback, although these possibilities may seem promising (Wong, et al., 2007). It appears that researchers are not that interested in the possibilities of ICT for the presentation of feedback. For the *personalisation* of feedback, ICT offers possibilities. The ICT-based learning approach that specifically focuses on personalised feedback is called Intelligent Tutoring Systems (ITS) (Sleeman & Brown, 1982). Despite progress in this area, human teachers have proven so far to be unbeatable in personalising their feedback one-to-one to individual students (Sarrafzadeh, Alexander, Dadgostar, Fan, & Bigdeli, 2008; Streibel, 1986). Anderson et al. (1995, p. 168), as pioneer developers of ITS, gave up competing with human tutors. Lots of subtle attributes of human interaction, like intonation of the voice and body language, are very difficult if not impossible by means of the ICT of today to interpret, let alone to use actively when the feedback situation demands to. Besides that, personalisation of feedback by human means is an intensive and thus expensive process. Personalisation can best be reached in individual interaction, and human tutors are hard to split up in order to be used by more than one student at the same time. This means that for personalisation of feedback, a human tutor needs a lot of time. Computers, on the other hand, are very good at sharing their attention to more than one user. We can conclude that ICT is not as good in personalising feedback as human teachers. However, ICT can do it very fast and more or less independent from the number of students. Therefore it has a great potential. In our intervention we chose to use both ICT feedback as well as human feedback.

Summarising, ICT has been reported to add value to the timing and personalisation of feedback. It could have a yet unexplored potential for the presentation of feedback. It could also offer its general features of speed and scalability for realising specificity, complexity, and addressing of feedback. In statistics education, there is a very important practical consideration that can be used to advocate the use of ICT. When using sets of real data, as is broadly advocated (Martin, et al., 2009; NCTM, 2005), directing to the importance of data (Cobb, 1991) and the importance of attractiveness for the target group (Garfield & Ahlgren, 1988), students usually have to handle a considerable amount of data (Mills, 2002). This is almost impossible without the support of ICT. Moore (1990) stated that computers and calculators have made more complex analyses on larger data sets possible. Thus, in a way, the use of ICT in statistics education, with goals as are enclosed in our study, is inevitable.

3. Design and development

Design principles with respect to feedback

The format used for the formulation of design principles has been inspired by the one proposed by Van den Akker (1999) and is formulated as:
If you want to <*goal formulation*> you are best advised to <*procedure formulation*> because of <*argument*>.

With respect to feedback seven design principles were formulated, based on the literature review and practical reasoning, including the following two (for a description of all seven principles, see Tolboom, 2012):

F1. If you want to provide feedback to students on a task that mainly concerns *procedural and/or declarative* knowledge, you are best advised to provide *immediate* feedback supplied by a *computer*, because this improves the efficacy of the feedback (Azevedo & Bernard, 1995; Corbett & Anderson, 2001; Kulik & Kulik, 1988).

F2. If you want to provide feedback to students on a task that mainly concerns *conceptual* knowledge, you are best advised to provide *delayed* feedback by *the teacher* after one or some more days because this improves the efficacy of the feedback (Butler, Karpicke, & Roediger III, 2007; Schroth, 1992).

Design principles with respect to statistics education

With respect to statistics education nine principles guided the design of the intervention, including the following two (see Tolboom, 2012):

SE1. If you want to improve the students' conceptual knowledge of statistics, you are best advised to include exercises that aim at developing 'data literacy' by systematically focussing on the students' understanding of the purpose and logic of statistical investigations and their development of reflective and interpretive skills, because this approach appeals more to students having a non-statistical focus (Cobb, 1991; Roback, 2003; Wiberg, 2009).

SE2. If you want to improve the students' declarative and procedural knowledge of statistics, you are best advised to include exercises that aim at developing 'algorithmic statistical skills' by systematically focussing on the students' skills with respect to practical, well defined statistical procedures, often with a quantitative goal or nature because this improves the efficacy of the feedback (Azevedo & Bernard, 1995; Corbett & Anderson, 2001; Kulik & Kulik, 1988).

Feedback matrix for statistics education

When ‘immediate feedback’ versus ‘delayed feedback’ are crossed with ‘data literacy’ versus ‘algorithmic statistical skills’, the result is a feedback matrix for statistics education as depicted in Table 1.

Table 1: Four feedback types in the feedback matrix for statistics education

Timing of feedback / **Learning goal**	**Immediate**	**Delayed**
Data literacy	Type **I**	Type **II**
Algorithmic statistical skills	Type **III**	Type **IV**

For a full description of exercises of types **I-IV** and examples of each of these feedback types we refer to Tolboom (2012).

Procedural specification

To illustrate the relationship between the intervention to be implemented, the hypothetical teaching trajectory (HTT) and the realised concrete instruction, we use the movies industry as a metaphor. The prototype can be interpreted as the screenplay upon which the movie is based. The movie itself represents the actual realised instruction in practice. We regard the HTT as the complete script for the movie: it is a director’s preview of how she sees what the actors have to say, have to do, where on stage, and how this should look and sound. In this study, the researcher is the director and the teacher is the actor; as in producing a movie, the director uses more than the complete script in order to prepare the actors. Director and actor constantly discuss about how the script is to be translated into a performance before the cameras. We regard this discussion as a vital part of the *procedural specification* (Doyle & Ponder, 1977; McIntyre & Brown, 1979; Van den Akker, 1988).

4. Implementation and evaluation during third cycle

After having deployed and revised earlier versions of the prototype of the intervention during the stages C1 (1 teacher, 31 grade 10 students) and C2 (one teacher, 25 grade 10 students), it was implemented in a third stage C3 (6 teachers, 6 grade 10 classes of in total 128 students; each teacher/class was a case; cases were designated as: S1, 25 students; S2a, 23, students; S2b, 17 students; S3, 32 students; S4a, 15 students; S4b, 16 students). S2a and S2b were groups on the same school, with a different teacher. S4a and S4b were groups on the same school, with the same teacher. The intervention in the classroom during each of the stages C1, C2 and C3 took about three weeks, depending on the exact timetable of the school.

In this section the results of the evaluation of the *third prototype* are described. Compared with the evaluation of C1 and C2 this yielded the most specific data. As part of this evaluation a teacher questionnaire was used. In addition, all the teachers were interviewed individually directly after each case study and simultaneously during a group interview after the last case study. From the student perspective, also a questionnaire was administered. After analysis of the results of this questionnaire for each case study, we interviewed three students of each participating group. We concentrated in these evaluations on: how well was the feedback established compared to the intentions (teachers) and how well was feedback received (students)? In this evaluation, we explicitly identified the role of the CN.

Teacher questionnaire and supporting interviews

This section contains the results of the teacher questionnaire, completed with remarks made during the subsequent teacher interview. Although the questionnaire and interview also investigated experiences with the teaching materials, support and other important contextual issues, we restrict ourselves now to report on the core issue: feedback. The main findings were the following. There was a considerable agreement among the teachers with respect to the potential of a classroom network in order to support feedback in statistics education. This may be remarkable, as it turned out that there were big differences with respect to the resulting classroom discourse, as will be shown in Table 3. The teachers considered the prototype to be suitable for developing DL. They felt better informed about the work of their students. In general, they gave more and more specific feedback. This counts both for ASS as for DL activities. The learning yields were perhaps bigger than in the case of more traditional instruction; on this the teachers were not that sure. They considered about 50% of the lesson time spent on feedback activities supported by the network as optimal.

Student questionnaire

All 128 C3 students completed the questionnaire. The questionnaire consisted of 20 statements about statistics education and feedback supplied by a classroom network. The five statements that were selected the most, were chosen for a statistical analysis. Those five statements were the following:

1. Because the teacher gives feedback on my work in this way, I get *more* feedback.
2. Because the teacher gives feedback on my work in this way, I get *better* feedback.
3. Because the teacher gives feedback on my work in this way, I manage to master *algorithmic statistical skills* better.
4. Because the teacher gives feedback on my work in this way, I manage to master *reasoning about data and calculations* better.
5. I would prefer to contribute to the classroom discussion on the projection screen *anonymously.*

The statements were formulated on a four point Likert scale, referring to a difference with respect to education without feedback through a classroom network (1: I strongly disagree, 2: I disagree, 3: I agree, 4: I strongly agree). Significance was considered at $\alpha = 0.05$ with $\alpha^B = \frac{0.05}{5} = 0.01$, α^B being the Bonferroni corrected level of significance. Significant results of this analysis, per case study and with all the students simultaneously (Total), are shown in Table 2.

Table 2: Results of statistical analysis students' questionnaire

Case	Cronbach's α	Item	Response	p
S1	0.68	I would prefer to contribute anonymously to the projected students' answers	Disagree	0.008
S2A	0.52	None		
S3	0.25	Because the teacher gives feedback on my work in this way, I get *more* feedback.	Agree	0.007
S4A	0.57	Because the teacher gives feedback on my work in this way, I get *more* feedback.	Agree	0.002
S4B	0.78	None		
S2B	0.60	Because the teacher gives feedback on my work in this way, I get *better* feedback.	Disagree	0.007
Total	0.57	Because the teacher gives feedback on my work in this way, I get *more* feedback.	Agree	0.006
		I would prefer to contribute anonymously to the projected students' answers	Disagree	0.002

The conclusion is that, in general, the students perceived more feedback on their work. Differentiation between ASS and DL was not significant. Discussing their answers in public was not threatening. The results differ slightly per case study, as does the consistency of the students' responses.

Student interviews

From each participating class, three students were selected for an interview: a good (1), an average (2) and a weak (3) student with respect to their mathematical competence, as appointed by their teacher. During these individual interviews, the results of the students' questionnaire in the specific case study were used as input. Students in the six participating grade 10 classes were positive about the improvement of the feedback. Usually, they perceived greater, and more specific, feedback, both on ASS as on DL activities, when compared with more traditional lessons. There was not much difference in appreciation of the feedback between the groups, except for the fact that the students in case study S2b were negative about the lessons themselves. Neither a structural difference could be notice in the experiences of the students as seen from their capabilities in mathematics. The students stressed the contribution of collaborative work to their learning. They did not experience this way of working as more threatening when compared to lessons without a classroom network, except for one interviewed student (out of eighteen) with weak mathematics skills.

Realized feedback

The realised feedback, classified by the feedback coding scheme as described by Tolboom (2012), was expressed in correspondence scores (Tolboom, 2012), based on observable behaviour. This score can vary from 0 to 7, indicating the degree of similarity between the intended curriculum (hypothetical teaching trajectory, HTT) and the realised curriculum (as shown in the video-taped lessons). In Table 3 we present an overview of the mean correspondence scores (with standard deviation and percentage of missing feedback sessions) of the intended and implemented feedback in the six case studies.

Table 3: Correspondence score characteristics of the cases in C3

Case	Mean	SD	Mean_DL	Mean_ASS	M_DL-M_ASS	%-Missing
S1	5.14	1.70	4.89	5.60	-0.71	68.09
S2A	4.40	2.29	3.66	5.71	-2.05	14.89
S3	5.38	1.85	5.34	5.46	-0.12	0.00
S4A	3.60	2.13	3.65	3.44	0.21	31.91
S4B	3.89	2.09	4.11	3.33	0.78	34.04
S2B	2.04	1.70	1.95	2.25	-0.30	59.57

Legend: The columns respectively present: case study label, mean total correspondence score, standard deviation of total correspondence, mean correspondence with respect to data literacy exercises, mean correspondence with respect to algorithmic statistical skill exercises, subtraction of these means and percentage of missing feedback sessions as compared to the HTT.

The very high percentages of missing feedback sessions should be pointed at, as compared to the HTT in the first (S1) and the last (S2b) case study. In the other four case studies this percentage was 34% or lower, so one on three missing, or less. These high percentages occurred as a result of two completely different reasons: in case S1 we encountered severe technical problems, but when the technology was up and running, the teacher managed to implement the curriculum rather close to the intentions. In case S2b the relationship between the group and teacher was problematic. This was discouraging to the teacher to the extent that she could not really implement the intended curriculum.

Results per sub-question

This intervention study was carried out because mathematics teachers experienced a curricular overload. The main idea was to improve the quality of contact time through an improvement of teacher feedback. We therefore utilised the possibilities of a classroom network. Teachers and students participating in this study agreed on the fact that during the intervention more feedback was established than during the usual educational setting.

The main research question was operationalized in four sub-questions. Based on the findings presented in Table 3 as well as questionnaire and interview data, those sub-questions can be answered as follows:

1. *Was the technological support by means of the CN adequate for the intended feedback in the lessons?* (Conditional question)

After case S1, the technology was stable, both on the handheld and on the network side. The percentage missing was between 0 and 36 (when not including S2b). Case S3 proved that it was possible to conduct every intended feedback session in the classroom setting.
In cases S4a and S4b, with a percentage missing of 35%, it was mainly time management by the teacher that obstructed utilising the CN more frequently. We nevertheless consider this percentage of 35% as reasonable for classroom practice as roughly two out of three sessions have then been realised. In case S2b, with a percentage missing of 60%, we were faced with such a low motivated group of students that the teacher felt that establishing more feedback sessions was not useful.

2. *Has it been possible for a mathematics teacher to implement the prototype in accordance with the intentions?* (Existence question)

We consider case S3 to be a convincing implementation of the intended intervention. This had a high mean correspondence score, both with respect to ASS (5.46) and DL (5.34). Students and the teacher were equally enthusiastic about the improvement of feedback. All of the feedback sessions were carried out as intended, demonstrating that the technology served the intervention very well. Besides a convincing case study being a *proof of existence* for the goals of the intervention it is noteworthy that in every single case study there were feedback sessions with a convincing correspondence score. This means that every teacher, under the right circumstances, has been able to conduct a feedback session as planned. This is even true for case S2b, with a poor correspondence score (1.95 on DL and 2.25 on ASS), low learning gains and low student commitment, during which there were feedback sessions with a satisfactory correspondence score (5 and 4.5). The interviewed students in this group voiced a preference for an average use of the CN in order to start feedback sessions for 65% of the teaching time. Apparently even in this case, the essential power of feedback supported by a CN emerged. We consider these as *'micro proofs of existence*': the teacher succeeded in conducting at least one feedback session sufficiently according to the intentions while the students were convinced of the feedback potential. As this relates to case S2b, it suggests that for the other cases the evidence is much stronger.

3. *Was the feedback support of the CN equal for ASS and DL?* (Didactical question)

The support for ASS proved to be better on average, but, with a highly specified HTT, we managed to support the teachers in giving feedback on DL in a satisfactory way. The slightly better support for ASS is shown by the fact that the mean difference in correspondence score between DL and ASS was 0.37 (in the advantage of ASS). We consider this gap, with respect to a variable on a scale from 0 to 7, to be quite small. The 'built in' support of the CN for developing students' DL has to be completed by specific teaching methods and by more directing teacher preparation. For this preparation we used one-to-one instruction before, and a similar evaluation after, each lesson.

The big difference in mean correspondence between feedback on DL and ASS in case study S2a during C3 (2.05) deserves some attention. We noted that the teacher in this case study did not have a strong *functional extraversion*. That is, he was not very focussed on leading the students' learning input during the classroom discourse. This hindered him more considerably in the feedback sessions on DL than on ASS. We presume that this could be because, for a mathematics teacher, discussing 'hard' procedures is easier than 'soft' processes. Discussing DL could be perceived as more vulnerable ('Why nagging when having an answer?') and therefore would take more *functional extraversion* than it takes to discuss ASS. In mathematics education, there is a stronger tradition of focussing on procedures than on concepts and ideas.

4. *Which teacher characteristics promoted/hindered the implementation of the CN as intended?* (Identification question)

The data collected during C3 showed that improving feedback in statistics education by the use of a classroom network was possible. But what was needed to make the step from a successful implementation during two case studies (out of six, like we did in C3) to a successful implementation in further case studies? First, we state that we did not reach 'successive approximation' (Van den Akker, 1999) of our 'intended use of the intervention': case S3 (chronologically the third case) came closer to the intentions than the fourth, fifth and sixth case, despite the fact that we continuously used our experiences in order to prepare the teachers in a better way.

Apparently, there was a teacher influence, a group influence, or an interaction between the teacher and group that was bigger than the influence on the intervention. The little difference in corresponding scores between S4a and S4b (with the same teacher for different groups) suggests that correspondence is more teacher dependent than group dependent. With respect to the teacher, this brings up an interesting question related to our main research question: what are the strong teacher influences that cause this variation in correspondence score?

5. Reflection and discussion

Four conditions for feedback potential

Using the results of C3, we can conclude that there are at least *four conditions* that have to be met before a teacher, trained and supported as we did during C3, in a learning environment that is technically stable, can fully utilise the feedback potential of the classroom network in statistics education.

First of all, we observed during case S2B of C3 as well as in C1 there should be a relationship between teacher and group that is based on *sufficient mutual trust*. If this trust is lacking, all education is to fail, however well-resourced the learning environment potential may be. Good education is an intimate process. Feedback and classroom discussion are perhaps the most vulnerable parts of it. Mutual trust is indispensable for making these succeed.

Secondly, the teacher has to have deep *conversational skills*, including the attitude (or is it even 'personality'?) to apply them as productively as possible in the classroom discourse. This means that she or he has to be a 'conductor' (Drijvers, Doorman, Boon, Reed, & Gravemeijer, 2010) of the classroom discourse, which in this context should be interpreted as 'the spider in the web of the educational process'. In this educational process classroom discourse takes a prominent place and leading it means being able to take up responsibility, especially in a communicative way. A sufficient level of *functional extraversion* is needed in order to be able to take up this responsibility. As a supplier of constructive critical feedback on students' work and when acting as a conductor of classroom discourse, the teacher has a very prominent role in the “classroom theatre”. This role he not only has to deserve, he has to demand it. We would not go as far as stating that this concerns the immutable level of the teacher's personality. It is about functional behaviour, which can be acquired. But this is a fairly severe requirement which can make it hard for a considerable part of the population of mathematics teachers to utilise the full potential of a classroom network. Very important was, as we observed during all the case studies, that the teacher was using students' names in order to spread the discourse among the whole group. Using students' names is more confronting, because it is more difficult for students to hide. The teacher therefore has to compensate by creating a safe environment, for example by showing some things of his own, without undermining his position as a leader. He has, as a real conductor, to make his musicians excel in their own way, without losing the collective goal. We especially point at the teacher's timing as a conversational skill. As when conducting musicians playing together, timing is essential for the proper performance of a piece of music, timing is also essential for a teacher in order to optimally implement an intended curriculum (in this case: HTT). In general the teacher's repertoire on formative assessment (Black, Harrison, Lee, Marshall, & Wiliam, 2003; Black & Wiliam, 2009) and dialogic teaching (Alexander, 2008) has to be at a sufficient level.

Thirdly, besides these conversational skills, the teacher should have competence in *quickly interpreting students' answers* as he has a greater number of these to handle than without the use of a classroom network. He should be able to make ‘statistical sense’ of much more student input than before. Due to teacher feedback being needed in the case of new ASS or DL student activity, this input will very often have the form of open answers. In this case, making a rough 'feedback scheme' based on students' answers, as we observed in C3 during case S3, can be very useful in giving feedback the right direction, simultaneously doing justice to the students' input. This capability in interpretation of students' input requires sufficient *subject matter (mathematical) knowledge and pedagogical content knowledge* (PCK) (Marks, 1990). The teacher should have a sufficient level of both knowledge types. This may sound trivial or perhaps even offensive. However, in our view there is no simple mathematics. It takes a deep understanding of concepts, ideas, procedures, strategies and links between different mathematical subdomains to be able to effectively process all of the students' responses on mathematical questions (Even, 1993) in real-time. Having much more information to process, like has been reported in this study, makes this job harder.

Fourthly, the teacher should have *skills with respect to ICT*. Using technology, both on the handheld side as well as on the network side, should only result in a low cognitive load so that the teacher is able to concentrate on giving feedback and directing the classroom discourse towards meaningful interaction with respect to statistics. If the technology requires too much attention to be handled successfully, this may lower the flow in a discussion. This technical condition may seem somewhat trivial, but in general the teacher acquisition of ICT skills is not to be underestimated (Hakkarainen, et al., 2001) and the integration of ICT skills for pedagogical use is especially difficult (Hughes, 2005). In recent research this aspect of teacher skills is more and more stressed (Mishra & Koehler, 2006). It takes a sustainable effort to maintain these skills in order to be able to smoothly switch to new tools or to new versions of familiar tools. PCK is nowadays supplemented by technological pedagogical content knowledge (TPCK) (Koehler & Mishra, 2009).

During C3 teachers as well as students considered that the supply of feedback using a classroom network, as perceived during this intervention, has such a high potential that they advise dedicating on average half of the lesson time to this teaching activity. It is noteworthy that even in the case study which was far from a success (S2B) the interviewed students mentioned percentages between 50 and 100. Students experienced more feedback than during education without a classroom network. Teachers reported improved feedback possibilities, with no difference between feedback on algorithmic statistical skills and on data literacy. Teachers reported that there was more discussion on data literacy than usual and that students in general were more engaged in the classroom discourse which in turn was more focused on mathematics. Teachers are a little cautious in reporting positive learning gains. In three cases it is reported, in the other three cases the teacher said they did not know.

Generalizability

We now discuss the generalizability of our studied intervention in terms of 'analytical generalizability' (Yin, 2003). With respect to *external validity*, designated as *transferability* by Guba (1981) in qualitative research like this study, this is mostly a concern of the generalizability of the research findings. We have secured this by precisely describing the dependency of our conclusions with respect to the research context (Barab, Baek, Schatz, Scheckler, & Moore, 2008). Attributes of teachers, students, disciplinary content, and technology were taken into account when interpreting our data.

These can be re-used when extrapolating our findings to other educational contexts (Barab & Kirshner, 2001). Further, by conducting this research in 'everyday classroom practice' we made its results more transferable to other research settings.
We presume, for example, that the main findings of this study are applicable to domains of learning other than just mathematics. In none of the used procedures is there an intrinsic domain-bounded step. Of course, the distinction between data literacy and algorithmic statistical skills is a typical statistical phenomenon. However, we used this distinction as an instance of the more abstract distinction between conceptual and procedural/declarative knowledge. This abstract distinction could be used in order to develop a comparable feedback matrix for another learning domain. The restriction to be made here is that the distinction of conceptual versus declarative-procedural has to be meaningful in this learning domain. We cannot oversee other learning domains, but those rather similar to mathematics, for instance physics, should be able to apply the same methodology in order to get comparable results, as the work of Mazur (2009) seems to indicate.

Future research

We recommend an expansion of this intervention study using the materials and research design we developed. This would preferably be for a longer period, for instance nine weeks. In such a period three chapters of a textbook could be taught and learned. With an expansion like that the 'start bias' that comes with working in a new learning environment could be reduced to an acceptable level. From a research perspective, we especially hope for a further outlining of the coding scheme and the correspondence metric, which we see as the main methodological contributions of this study.

We recommend to conduct this research among substantially more than six teachers, for instance between ten and twenty. We expect that a sample of that size should generate sufficient 'variability' amongst teachers with respect to the teacher dependent *success conditions* we formulated in this study. A study like this should yield enough data for the fine tuning of the prototype. When the intervention is adapted according to the findings of a study as recommended, it would then be appropriate to conduct a quantitative effect study. In the training of the participating teachers, the videos collected and analysed in this study could be very useful, as we can conclude from the high appreciation of the videos shared with the teachers in our study using YouTube. Specific behaviour and critical events in the classroom discourse can be highlighted by using our materials.

It would be very interesting after the first intervention to repeat it three years later, meanwhile letting the teachers optimise their teaching according to the principles we formulated. This should show significantly better results (Adey, 2006).

Key sources

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practices, 5*(1), 7-74.

Tolboom, J.L.J. (2005). Draadloos netwerk in de wiskunde klas. [Wireless network in the mathematics classroom]. *Euclides, 81*(3), 108-112.

Tolboom, J.L.J. (2012). *The potential of a classroom network to support teacher feedback; A study in statistics education.* PhD, University of Groningen, Groningen. Retrieved from http://irs.ub.rug.nl/ppn/343084066.

References

Abrahamson, A.L. (1999). Teaching with classroom communication system - What it involves and why it works. Retrieved December 29, 2011, from http://www.bedu.com/Publications/PueblaFinal2.html.

Adey, P. (2006). A model for the professional development of teachers of thinking. *Thinking Skills and Creativity, 1*(1), 49-56. doi: 10.1016/j.tsc.2005.07.002.

Alexander, R. (2008). *Towards dialogic thinking: Rethinking classroom talk* (4 ed.). York, UK: Dialogos.

Anderson, J.R., Corbett, A.T., Koedinger, K.R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*(2), 167-207.

Azevedo, R., & Bernard, R.M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research, 13*(2), 111-127.

Bangert-Drowns, R.L., Kulik, C.L., Kulik, J.A., & Morgan, M.T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213-238.

Barab, S.A., Baek, E.-O., Schatz, S., Scheckler, R., & Moore, J. (2008). Illuminating the braids of change in a web-supported community; A design experiment by another name. In A. E. Kelly, R.A. Lesh, & J.Y. Baek (Eds.), *Handbook of design research methods in education* (pp. 320-352). New York, NY: Routledge.

Barab, S.A., & Kirshner, D. (2001). Rethinking methodology in the learning sciences. *Journal of the Learning Sciences, 10*(1&2), 5-15.

Berch, D.B., & Mazzocco, M.M.M. (Eds.). (2007). *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities.* Baltimore, MD: Paul H. Brookes Publishing Co.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: putting it into practice.* Maidenhead, Berkshire, UK: Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*: *Principles, Policy & Practices, 5*(1), 7-74.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5-31.

Butler, A.C., Karpicke, J.D., & Roediger III, H.L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied, 13*(4), 273-281.

Cobb, G.W. (1991). Teaching statistics: More data, less lecturing. *Amstat News*, 182, 1-4.

Cohen, V.B. (1985). A reexamination of feedback in computer-based instruction: Implications for instructional design. *Educational Technology*, *25*(1), 33-37.

Corbett, A.T., & Anderson, J.R. (2001). *Locus of feedback control in computer-based tutoring: impact on learning rate, achievement and attitudes.* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems, Seattle, Washington, United States.

Davis, W.D., Carson, C.M., Ammeter, A.P., & Treadway, D.C. (2005). The interactive effects of goal orientation and feedback specificity on task performance. *Human Performance, 18*(4), 409-426.

Doyle, W., & Ponder, G. (1977). The practical ethic and teacher decision-making. *Interchange, 8*(3), 1-12.

Drijvers, P., Doorman, M., Boon, P., Reed, H., & Gravemeijer, K. (2010). The teacher and the tool: Instrumental orchestrations in the technology-rich mathematics classroom. *Educational Studies in Mathematics, 75*(2), 213-234. doi: 10.1007/s10649-010-9254-5.

Even, R. (1993). Subject-matter knowledge and pedagogical content knowledge: Prospective secondary teachers and the function concept. *Journal for Research in Mathematics Education, 24*(2), 94-116.

Fies, C., & Marshall, J. (2006). Classroom response systems: A review of the literature. *Journal of Science Education and Technology, 15*(1), 101-109.

Gal, I., & Garfield, J.B. (1997). Curricular goals and assessment challenges in statistics education. In I. Gal & J.B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 1-13). Amsterdam: IOS.

Garfield, J.B., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education, 19*(1), 44-63.

Geary, D.C. (2010). Missouri longitudinal study of mathematical development and disability. *British Journal of Educational Psychology Monograph Series II, Number 7 - Understanding number development and difficulties, 1*(1), 31-49.

Ginsburg, H.P. (1997). Mathematics learning disabilities: A view from developmental psychology. *Journal of Learning Disabilities, 30*, 20-33.

Goodman, J.S., & Wood, R.E. (2004). Feedback specificity, learning opportunities, and learning. *Journal of Applied Psychology, 89*(5), 809-821.

Gravemeijer, K.P.E. (1994). *Developing realistic mathematics education*. Utrecht: CD-ß Press / Freudenthal Institute.

Guba, E. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational technology research and development, 29*(2), 75-91. doi: 10.1007/bf02766777.
Gustafson, K.L., & Branch, R.M. (2002). *Survey of instructional development models* (4th ed.). Syracuse, NY: ERIC Clearinghouse on Information and Technology.

Hakkarainen, K. A.I., Muukonen, H., Lipponen, L., Ilomaki, L., Rahikainen, M., & Lehtinen, E. (2001). Teachers' information and communication technology (ICT) skills and practices of using ICT. *Journal of Technology and Teacher Education, 9*(2), 181-197.

Hannafin, M., Philips, T., Rieber, T., & Garhart, C. (1987). The effects of orienting activities and cognitive processing time on factual and inferential learning. *Educational Communications and Technology Journal, 35*(2), 75-84.

Hattie, J.A. (1999). Influences on student learning. Retrieved December 22nd, 2011, from http://www.education.auckland.ac.nz/uoa/fms/default/education/staff/Prof.%20John%20Hattie/Documents/Presentations/influences/Influences_on_student_learning.pdf.

Hattie, J.A. (2009). *Visible learning: A synthesis of meta-analyses in education*. London: Routledge.

Hattie, J.A., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education, 21*(1), 33-46.

Hong, E., & Karstensson, L. (2002). Antecedents of state test anxiety. *Contemporary Educational Psychology, 27*(2), 348-367.

Hughes, J. (2005). The role of teacher knowledge and learning experiences in forming technology-integrated pedagogy. *Journal of Technology and Teacher Education, 13*(2), 277-302.

Kelly, A.E. (2003). Theme issue: The role of design in educational research. *Educational Researcher, 32*(1), 3.

Kelly, A.E. (2007). When is design research appropriate. In T. Plomp & N. Nieveen (Eds.), *An introduction to educational design research* (pp. 73-88). Enschede, the Netherlands: SLO.

Kluger, A.N., & DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284.

Koehler, M.J., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education, 9*(1).

Küchemann, D E. (1981). Algebra. In K. M. Hart (Ed.), *Children's understanding of mathematics: 11-16* (pp. 102-119). London: John Murray.

Kulik, J.A., & Kulik, C.L. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*(1), 79-97.

Lovett, M.C. (2001). A collaborative convergence on studying reasoning processes: a case study in statistics. In S. Carver & D. Klahr (Eds.), *Cognition and instruction: twenty-five years of progress* (pp. 347-384). Mahwah, NJ: Erlbaum.

Marks, R. (1990). Pedagogical content knowledge: From a mathematical case to a modified conception. *Journal of teacher education, 41*(3), 3-11.

Martin, G.W., Carter, J., Forster, S., Howe , R., Kader, G., Kepner, H., et al. (2009). *Focus in high school mathematics: reasoning and sense making*. Reston, VA: NCTM.

Marzano, R.J., Pickering, D.J., & Pollock, J.E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement.* Alexandria, VA: Association for Supervision and Curriculum Development.

Mazur, E. (2009). Farewell, lecture? *Science, 323*(5910), 50-51. doi: 10.1126/science.1168927.

McCloskey, M., Caramazza, A., & Basili, A. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition, 4*, 171-196.

McIntyre, D., & Brown, S. (1979). Science teachers' implementation of two intended innovations. *Scottish Educational Review, 11*(1), 42-57.

Mills, J.D. (2002). Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education, 10*(1).

Mishra, P., & Koehler, M.J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017-1054.

Molenda, M. (2003). In search of the elusive ADDIE model. *Performance Improvement, 42*(5), 34-36.

Molenda, M., Pershing, J. A., & Reigeluth, C.M. (1996). Designing instructional systems. In R.L. Craig (Ed.), *The ASTD training and development handbook* (4th ed., pp. 266-293). New York: McGraw-Hill.

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.

Moore, D.S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review, 65*(2), 123-137.

Moore, M.G. (1989). Three types of interaction. *American Journal of Distance Education, 3*(2), 1-6.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science, 32*, 99-113.
NCTM. (2005). *Computation, calculators, and common sense.* Position statement May 2005.

Paulos, J.A. (1988). *Innumeracy, mathematical illiteracy and its consequences*. London: Penguin Books.

Penuel, W.R., Boscardin, C.K., Masyn, K., & Crawford, V.M. (2007). Teaching with student response systems in elementary and secondary education settings: A survey study. *Educational Technology Research and Development, 55*(4), 315-336.

Phillips, T., Hannafin, M., & Tripp, S. (1988). The effects of practice and orienting activities on learning from interactive video. *Educational Communication and Technology Journal, 36*, 93-102.

Pridemore, D.R., & Klein, J.D. (1995). Control of practice and level of feedback in computer-based instruction. *Contemporary Educational Psychology, 20*(4), 444-450.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28*, 4-13.

Rapp, B., & Goldrick, M. (2004). Feedback by any other name is still interactivity: a reply to Roelofs (2004). *Psychological Review, 111*(2), 573-578.

Rittel, H.W.J., & Webber, M.M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*(2), 155-169.

Roback, P.J. (2003). Teaching an advanced methods course to a mixed audience. *Journal of Statistics Education, 11*(2).

Roschelle, J. (2003). Unlocking the learning value of wireless mobile devices. *Journal of Computer Assisted Learning, 19*(3), 260-272.

Roschelle, J., Penuel, W.R., & Abrahamson, L. (2004). The networked classroom. *Educational Leadership, 61*(5), 50-54.

Rosnick, P., & Clement, J. (1980). Learning without understanding: the effect of tutorial strategies on algebra misconceptions. *Journal of Mathematical Behaviour, 3*(1), 3-27.

Sadler, D.R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119.

Sarrafzadeh, A., Alexander, S., Dadgostar, F., Fan, C., & Bigdeli, A. (2008). "How do you know that I don't understand?" A look at the future of intelligent tutoring systems. *Computers in Human Behavior, 24*(4), 1342-1363.

Schroth, M.L. (1992). The effects of delay of feedback on a delayed concept formation transfer task. *Contemporary Educational Psychology, 17*(1), 78-82.

Shaughnessy, J M. (2010). Statistics for all - the flip side of quantitative reasoning. *President's Corner.* Retrieved December 29th, 2011, from http://www.nctm.org/about/content.aspx?id=26327.

Shute, V.J. (2008). The focus on formative feedback. *Review of Educational Research, 78*(1), 153-189.

Skinner, B.F. (1958). Teaching machines. *Science, 128*, 969-977.

Sleeman, D., & Brown, J. (1982). *Intelligent tutoring systems*: Academic Press.

Smith, G. (1998). Learning statistics by doing statistics. *Journal of Statistics Education, 6*(3).

Streibel, M.J. (1986). A critical analysis of the use of computers in education. *Educational Technology Research and Development, 34*(3), 137-161.

Tolboom, J.L.J. (2005). Draadloos netwerk in de wiskundeklas. [Wireless network in the mathematics classroom]. *Euclides, 81*(3), 108-112.

Tolboom, J.L.J. (2012). *The potential of a classroom network to support teacher feedback; A study in statistics education.* Doctoral dissertation. Groningen, the Netherlands: University of Groningen.

Van den Akker, J.J.H. (1988). *Ontwerp en implementatie van natuuronderwijs. [Design and implementation of science education].* Amsterdam/Lisse, the Netherlands: Swets & Zeitlinger.

Van den Akker, J.J.H. (1999). Principles and methods of development research. In J.J.H. van den Akker, R.M. Branch, K.L. Gustafson, N. Nieveen, & T. Plomp (Eds.), *Design approaches and tools in education and training* (pp. 1-14). Dordrecht: Kluwer Academic Publishers.

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

Van den Akker, J.J.H., & Kuiper, W. (2008). Research on models for instructional design. In J. M. Spector, M.D.M.D. Merrill, J. van Merriënboer & M.P. Driscoll (Eds.), *Handbook of research on educational communication and technology* (Third ed., pp. 739-748). New York: Lawrence Erlbaum Associates.

Van Streun, A. (2001). *Het denken bevorderen. [Promoting thinking].* Groningen, the Netherlands: RUG.

Watson, J.M., Collis, K.F., Callingham, R.A., & Moritz, J.B. (1995). A model for assessing higher order thinking in statistics. *Educational Research and Evaluation, 1*(3), 247-275.

Watson, J.M., Gal, I., & Garfield, J.B. (1997). Assessing statistical thinking using the media. In J.B. Garfield & I. Gal (Eds.), *The assessment challenge in statistics education*. Amsterdam, the Netherlands: IOS Press and International Statistical Institute.

Wiberg, M. (2009). Teaching statistics in integration with psychology. *Journal of Statistics Education, 17*(1).

Wong, W.L., Shen, C., Nocera, L., Carriazo, E., Tang, F., Bugga, S., et al. (2007). *Serious video game effectiveness.* Paper presented at the International conference on advances in computer entertainment technology, Salzburg, Austria.

Yin, R.K. (2003). *Case study research. Design and methods* (4th ed.). Thousand Oaks, CA: Sage.

Jos Tolboom (1966) is a curriculum developer mathematics for upper secondary education at SLO [Netherlands Institute for Curriculum Development]. His work concentrates on coordinating projects involved with development of the Dutch arithmetic and mathematics curriculum. He holds a MSc in econometrics, mathematics education and computer science education and a PhD in mathematics education. From 1991-2001 he was a teacher of mathematics and computer science in upper secondary education, from 1997-2010 he worked as a lecturer in mathematics and mathematics education at the University of Groningen. His dissertation, in 2012, called ‘The potential of a classroom network to support teacher feedback; A case study in statistics education’, describes the design, development, implementation and evaluation of an intervention that utilizes a network of graphing calculators in upper secondary statistics education for the sake of feedback. This dissertation is the base for the underlying chapter of this book

Email: j.tolboom@slo.nl

Wilmad Kuiper is a professor at the Freudenthal Institute for Mathematics and Science Education at Utrecht University in the Netherlands. He is also head of the research department of SLO, the Netherlands Institute for Curriculum Development. His dissertation (1993) is about the implementation of science subject curriculum renewals in junior secondary education in the Netherlands. His research interests include curriculum evaluation, in particular as regards mathematics and science education, and the consistency between (science) curriculum renewal and assessment.

Email: w.kuiper@slo.nl