FIsme Scientific Library

Christian Bokhove

Use of ICT for acquiring, practicing and assessing algebraic expertise

Faculteit Bètawetenschappen **FIsme**

Use of ICT for acquiring, practicing and assessing algebraic expertise

Bokhove, Christian

Use of ICT for acquiring, practicing and assessing algebraic expertise/ C. Bokhove - Utrecht: Freudenthal Institute for Science and Mathematics Education, Faculty of Science, Utrecht University / FIsme Scientific Library (formerly published as CD-β Scientific Library), no 68, 2011.

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

ISBN 978-90-70786-09-0
Key words: algebraic expertise / algebraic skills / assessment / feedback / ICT / online environment / symbol sense

Cover design: Vormgeving Faculteit Bètawetenschappen
Printed by: Drukkerij Wilco

USE OF ICT FOR ACQUIRING, PRACTICING AND ASSESSING ALGEBRAIC EXPERTISE

HET GEBRUIK VAN ICT BIJ HET VERWERVEN, OEFENEN EN TOETSEN VAN ALGEBRAÏSCHE EXPERTISE

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universtiteit Utrecht, op gezag van
de rector magnificus, prof. dr. G. J. van der Zwaan, ingevolge het besluit van
het college voor promoties in het openbaar te verdedigen
op maandag 12 december 2011 des middags te 12.45 uur

door
Christian Bokhove

geboren 16 september 1975
te Amsterdam, Nederland

Promotor:	Prof. dr. J.A. van Maanen
Co-promotor:	Dr. P.H.M. Drijvers

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

Table of Contents

Preface

I have been involved in the use of ICT for mathematics for many years now. As a mathematics teacher, my implementation of ICT in the curriculum was ‘quick and dirty’: I would think of using a certain computer tool, and make an accompanying lesson plan; the plan was then executed, and revised almost daily. Some experiments, which seemed successful, were added to the curriculum permanently; other experiments were disastrous, with the result that I would not look at certain tools for years. All this is the typical environment in which a teacher operates. By 2007 I was becoming more and more curious about the effects of all these wondrous tools I was deploying, especially in the field of algebra. I wanted this to be more than ‘just a hunch’ and applied for a position within the DUDOC Program to carry out educational research on a relevant mathematical topic in secondary school. This thesis documents the development and outcomes of this research that was carried out between 2007 and 2011 at the Freudenthal Institute for Science and Mathematics Education, Utrecht University. In addition to my scientific aims it was my intention to build bridges on several levels.

First and foremost I sought to bridge theory and practice. In theory, educational research has educational practice as its object of study. It can, however, be problematic to apply theoretical results of research to educational practice: theory should be informed more by practice. Equally, actors such as students and teachers in the field (practice) should try to reap the benefits of research that has taken place. I am sure that bridging the gap between theory and practice would be beneficial for both educational research and practice. For this reason one goal of the study was to disseminate findings from the study to the realm of mathematics teachers in the Netherlands (e.g. Bokhove, 2011).

Second, in using a mixed method approach qualitative and quantitative methods are combined not only because the accompanying research questions invite this approach, but also because the two methods complement each other. A qualitative approach examines the underlying factors: why does it work and how does it work? A quantitative approach often provides a basis for generalizations: does it work and what factors determine that it works? As many studies seem to rely on just one approach, this study shows the benefits of a mixed method by providing an integrative picture.

Finally, this study was undertaken in the firm belief that procedural skills and conceptual understanding are closely interlinked: understanding leads to improved skills, better skills can lead to deeper understanding. In this sense, the study would like to contribute to ending the ongoing debate between proponents of both ‘sides’, and bridge the gap.

I think that if we can bridge these gaps, step by step, we can all benefit: teachers from researchers, researchers from teachers, and thus in the end the whole of society.

To Miriam, Kieran, Skye, Fionn,
Bran and Alec

If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is. ~John von Neumann

Chapter 1

Introduction

Introduction

1 Research context: ICT and algebra

For several years the skill level of students leaving secondary education in the Netherlands has been discussed. Lecturers in higher education -for example- often complain of an apparent lack of algebraic skills. This problem seems to have grown larger during the last few years and is not restricted to the Netherlands. In 1995 already, the London Mathematical Society reported:

> (i) a serious lack of essential technical facility—the ability to undertake numerical and algebraic calculation with fluency and accuracy;
> (ii) a marked decline in analytical powers when faced with simple problems requiring more than one step;
> (iii) a changed perception of what mathematics is — in particular of the essential place within it of precision and proof.
>
> (London Mathematical Society, 1995, p. 3)

. The key point of the relationship between procedural skills and conceptual understanding has been widely debated. This relationship plays a central role in the 'Math wars' discussions (Schoenfeld, 2004). An important issue in this debate is how students best acquire algebraic expertise: by practicing algorithms, or by focusing on reasoning and strategic problem solving activities. The first approach sees computational skills as a prerequisite for understanding mathematical concepts (US Department of Education, 2007). In the latter approach, the focus is on conceptual understanding (ibid.). Even if the idea is shared that both procedural skills and conceptual understanding are important, there are disagreements on their relationship and the priorities between the two. What is certain is that the debate also gained a foothold in Dutch education and therefore many Dutch institutions for higher education established entry exams and bridging courses to assess and improve students' mathematical skills (Heck & Van Gastel, 2006). In 2006 a Dutch project called NKBW (National Knowledgebase Basic skills Mathematics) was started to address and scrutinize this gap in mathematical skills between secondary and higher education. The final report of the NKBW project (Consortium NKBW, 2007) reaffirmed that there is a problem with mathematical skills, algebraic skills in particular, but concluded that more research on the nature of the problem must be conducted. To tackle the issues of limited algebraic skills, this research focuses on two relevant issues in mathematics education at Dutch upper secondary level. On the one hand signals from higher and secondary education that students lack algebraic skills (Tempelaar, 2007; Tempelaar & Caspers, 2008; Vos, 2007); on the other hand the use of ICT in mathematics education. To stress the importance of both issues they are also addressed in the vision document entitled ‘Rich in meaning’ by the Dutch national committee for mathematics curriculum reform (cTWO, 2007). Amongst others, the report stresses the importance of numbers, formulas, functions, and change, and claims that ICT should be ‘used to learn’ and not ‘learned to use’. Other important points in the cTWO report are:

- A specific case is made for the transition of students from secondary education towards higher education. It is stressed that this transition needs more attention.
- The importance of assessment of algebraic skills is stressed.
- One statement mentions the pen-and-paper aspect of mathematics. In the end transfer should not only take place mentally, but also notation wise.

. We conclude that there is ample reason to address the issue at stake. As algebraic skills will have a more prominent position in the 2015 mathematics curricula, the research is closely related to the reform of science and mathematics education in the Netherlands. Having said this, in this modern age it also makes sense to use the potential of ICT for addressing algebraic skills. According to many available resources, the potential of computers for secondary education has been widely recognized (Voogt & Knezek, 2008). In line with recent research findings (e.g. Goos et al., 2009; Heid & Blume, 2008a, 2008b; Li & Ma, 2010; Pierce & Stacey, 2010), the U.S. National Council of Teachers of Mathematics (2008) acknowledges the potential of ICT for mathematics education in its position statement. The Freudenthal Institute for Science and Mathematics Education at Utrecht University also has a tradition in the topic of assessment for mathematics (e.g. Van den Heuvel-Panhuizen, 1996) and the use of digital tools with feedback (e.g. Kolovou, 2011). Because of this potential, the last decade has seen a rise of online environments for algebra. We should however keep in mind that learning algebra with technology can be different from learning with pen-and-paper. Research within the framework of instrumental and anthropological approaches shows that the use of ICT tools and conceptual understanding interact (Artigue, 2002). To exploit ICT's potential for the development of algebraic expertise, it is crucial that students can reconcile conventional pen-and-paper techniques and ICT techniques (Kieran & Drijvers, 2006). This is the main reason why the design of a digital tool or intervention is complex. Beeson (1998) sums up design principles that are important for digital algebra environments. Building on Buchberger’s (1990) whitebox/blackbox distinction, Beeson stresses the importance of transparency of the solution process for educational practice. Combining the development of algebraic expertise and educational use of ICT, the aim of this study is to design an online environment for the formative learning of both procedural skills and conceptual understanding in algebra, investigate the effects of the environment, and to identify decisive factors that influence the outcome.

2 Research questions

To address these issues, this study focuses on the following central research question: *in what way can the use of ICT support acquiring, practicing and assessing algebraic expertise?*

As a first step, this question is briefly analyzed word-by-word:

In what way: the premise of the research is that ICT can be used to support learning, testing and assessing mathematical skills. The question is how this should take place.

Use of ICT: the second premise is that the use of ICT should have an important role in the math curriculum, as it has the potential for developing skills any time, any place.

Acquiring is about developing actual understanding of the subject at hand.

Practicing refers to the development of skills through practice.

Assessing: not only final results, grades and scores are important, but also the ways in which mathematical concepts are learned and tested diagnostically. The premise is that (formative) assessment for learning should be central.

Algebraic expertise: when students leave secondary education they are expected to have learned certain skills. Here we focus on algebraic skills, with particular attention given to skills in relation to conceptual understanding. In the literature, this is addressed by the notion of algebraic expertise.

. The main question is elaborated in several sub-questions.

A first category of questions concerns the choice of a digital tool. Before we set out to use ICT tools it is important to know what characteristics such a tool should have:

1a) Which criteria are relevant for the evaluation of digital tools for algebra education?

Next, the identified characteristics should guide the choice of tool for the study:

1b) Which digital algebra tool best meets these criteria?

A next category of questions focuses on algebraic expertise. Does an intervention made with this tool actually enable students to show or not show algebraic expertise? And how can this be investigated? We wonder which theoretical concepts can help.

2a) Do the concepts of symbol sense, gestalt view and pattern and local visual salience, described in a pre-digital era, help us in understanding what students do in a digital environment?

As feedback plays an important role in the intervention it is important to see if it can be improved by making use of students' domain knowledge:

2b) Can the feedback design of a digital tool be improved with students?

And if it can be improved, how could this be done best:

2c) What methodology can be used to elaborate appropriate feedback for students?

The third category of sub-questions focuses on feedback design. The question to answer is in what way feedback can be implemented and used in the design of the intervention: in what way can feedback be used in the design of an intervention and what are its effects? Three principles that already came up in answering question 2a are investigated: timing and fading, crises and variation:
3a) Timing and fading: are there indications that formative scenarios improve the acquisition of algebraic expertise?
3b) Crises: do crises in algebraic tasks improve the acquisition of algebraic expertise?
3c) Variation: does variation in feedback influence scores and student behavior?

Finally, the overall effects of the intervention have to be studied, and in particular the factors that predict algebraic performance.
4a) What is the effect of an intervention on the development of algebraic expertise, including both procedural skills and symbol sense?
4b) What factors predict the resulting algebraic performance, taking into account the interplay between multilevel factors, and compositional effects on both class and student level?

The hypothesis is that the use of ICT tools, if carefully designed and integrated in an intervention, can increase algebraic skill performance in general and algebraic expertise in particular.

The research aims at results including:

- Characteristics and evaluation criteria for algebra tools;
- Design and evaluation of a digital module for acquiring, practicing and assessing algebraic expertise;
- A methodology for implementing student-informed feedback into a digital algebra tool;
- A local instruction theory with design principles for using an ICT tool for acquiring, practicing and assessing algebraic skills;
- More insight in the use of an algebra tools, its effects and what factors influence these effects.

Four key topics emerge from these research questions: ICT tool use, algebraic expertise, assessment and feedback. The latter two are closely related in the theoretical lens of formative assessment, and therefore we will treat these as one within this framework. In this introductory chapter we now globally describe the notions that underpin the study; further elaborations can be found in the following chapters.

3 Theoretical framework

The context of this research combines three key perspectives: ICT tool use, algebraic expertise, assessment and feedback.

ICT tool use

Technology has an impact on mathematics education. Research with calculators (Ellington, 2003) has shown that the pedagogical role of tool use should not be underestimated. The use of tools seems to strengthen a positive attitude towards education, showing that there is more to learning than just practicing and testing. Guin and Trouche (1999) notice that students have different 'styles' of coping with problems: random, mechanical, rational, resourceful and theoretical. An important distinction in type of activity is between exploratory and expressive tools and activities (Bliss & Ogborn, 1989). They reside on a continuum. When a procedure is described it is considered to be exploratory; choosing one's own procedure is expressive (in an albeit limited manner). Initial play with a technological tool is often beneficial: it stimulates expression but also builds a purposeful relationship with the tool , and thus instrumental genesis (Guin & Trouche, 1999) can take place. However, structured guidance is often necessary, as to avoid the 'play paradox' (Noss & Hoyles, 1992). This means that 'playing' with a tool sometimes enables students to accomplish an activity without learning the intended concepts. For this study we contend that some form of guidance is necessary. Tool use is an integrated part of human behavior. Vygotsky (1978) sees a tool as a mediator, a "new intermediary element between the object and the psychic operation, directed at it". Verillon and Rabardel (1995) distinguish artifact and instrument. The artifact is just the tool. The instrument is a psychological notion: the relationship between a person and the artifact. Only after this relationship is established one can call it an instrument. The mental processes that come with this are called schemes. In short: instrument = artifact + instrumentation scheme. Trouche (2004) distinguishes instrumentation (how the tool shapes the tool-use) and instrumentalization (the way the user shapes a tool). Instrumental genesis is the process of an artifact becoming an instrument. In this process both conceptual and technical knowledge play a role ('use to learn' and 'learn to use'). To overcome the contrast between pen-and-paper and ICT based learning, techniques for using an ICT environment have to correspond with traditional techniques (Kieran & Drijvers, 2006). According to several studies (Artigue, 2002; Guin, Ruthven, & Trouche, 2005), instrumental genesis in the case of computer algebra systems is a time-consuming and lengthy process. When focusing on particular aspects of instrumental genesis, for example instrumentation, instrumentalization and technique (Guin & Trouche, 1999), it becomes clear how students can use tools more effectively and what obstacles hinder conceptual and technical understanding. The instrumental approach in this study played a role (i) when choosing an appropriate tool, and (ii) when thinking of ways to improve the transfer from ICT towards pen-and-paper. This lens will return in chapter 2.

Algebraic expertise

The notion of understanding of mathematical concepts has been given different names. Hoch and Dreyfus called this ‘structure sense’ (Hoch & Dreyfus, 2004). Arcavi (1994) coined the term ‘symbol sense’. Symbol sense is an intuitive feel for when to call on symbols in the process of solving a problem, and conversely, when to abandon a symbolic treatment for better tools. Arcavi (1994) describes several ‘behaviors’ with one being ‘flexible manipulation skills’. The behavior of flexible manipulation skills requires a certain ‘Gestalt’ skill: one has to recognize certain features of algebraic expressions. Wenger (1987) and Gravemeijer (1990) also studied this aspect of these global characteristics of expressions. Here, ‘Gestalt’ plays a role: does he or she recognize similar parts of an equation? Not recognizing patterns often leads to ‘circularity’: rewriting an expression to finally end up where you began. In line with this work, Drijvers, Goddijn and Kindt (2010) see algebraic expertise as a dimension ranging from basic skills to symbol sense, as indicated in Figure 1.

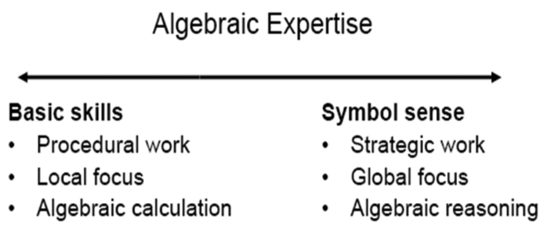


Fig.1: Algebraic expertise as a dimension (Drijvers, Goddijn & Kindt, 2010)

Basic skills involve procedural work with a local focus and emphasis on algebraic calculation, while symbol sense involves strategic work with a global focus and emphasis on algebraic reasoning. Or as Zorn (2002) puts it: “By symbol sense I mean a very general ability to extract mathematical meaning and structure from symbols, to encode meaning efficiently in symbols, and to manipulate symbols effectively to discover new mathematical meaning and structure” (p. 4). The lens of algebraic expertise will be used for evaluating the items we should use: series of tasks should not only be based on repetition and skills, but also on the possibility to discover patterns and symbols. Chapter 3 elaborates on the lens of algebraic expertise.

Assessment and feedback

Black and Wiliam (2004) distinguish three functions for assessment:

- supporting learning (formative)
- certifying the achievements or potential of individuals (summative)
- evaluating the quality of educational programs or institutions (evaluative)

Summative assessment is also characterized as assessment *of* learning and is contrasted with formative assessment, which is assessment *for* learning. In most curricula summative assessment is used. Summative assessment is mostly aimed at grading and scoring. Some researchers argue that, instead of providing a certain grade which seems to say what level of knowledge a student has, formative tests give the student an insight into the nature of -for example- their misconceptions. Black and William (2004) have made a case for more formative assessment. In their article from 1998 they state that "improving formative assessment raises standards" (Black & Wiliam, 1998). Actively involving the student, implementing formative assessment as an essential part of the curriculum and motivating students through self-assessment are key benefits of formative assessment. Means to do this are feedback, self-assessment, reflection and interaction. This makes a good case for tools that aid these factors. Black and Wiliam (1998) define assessment as being formative when the feedback from learning activities is actually used to modify teaching to meet the learner's needs. Using pre-emptive formative assessment means using feedback as a central element in the learning process (Hattie, Biggs, & Purdie, 1996). Instead of serving up feedback too little too late, feedback is used pre-emptively, to make sure that the student is on the right track. From all of this it is clear that feedback plays a pivotal role in the process of formative assessment.

According to an overview by Mason and Bruning (2001) determining factors for effectiveness of feedback are: elaboration, student achievement levels, depth of understanding, attitude toward feedback, learner control, response certitude, and timing. To give an example, Morrisson et al. (1995) found that delayed and knowledge-of-correct-response feedback may be more beneficial than answer-until-correct or no-feedback for lower level learning, while Clariana (1990) concluded that answer-until-correct feedback may be more effective for higher order learning than for lower level processing. While lower ability students may benefit from more immediate, specific forms of feedback, higher ability students may gain more knowledge from feedback that allows for active processing by the student. Hattie and Timperley (2007) conducted a meta-review of the effectiveness of different types of feedback. The feedback effects of cues and corrective feedback are deemed best. Feedback question works at four levels (focus of the feedback): the task level: how well tasks are understood and performed (FT), the process level: the main process needed to understand/perform tasks (FP), the self-regulation level: self-monitoring, directing and regulating of actions (FR), and the self-level: personal evaluations and affect (usually positive) about the learner (FS). Hattie and Timperley (2007) also provide some statements on the effectiveness of (combinations of) feedback types, including that FS feedback is least effective, simple FT feedback is more effective than complex FT feedback, FT and FS do not mix well ("Well done, that is correct" is worse than "Correct" only), and that FT is more powerful when it's about faulty interpretations. Furthermore they state that we should be attentive to the varying importance of the feedback information during study of the task. A review by Vasilyeva et al. (2007) also provides recommendations for the effective use of feedback in web-based application. In this study we focus on the use of formative assessment, embedded in a didactical scenario, so-called formative scenarios, for algebra. As feedback plays a pivotal role in these notions, the design of feedback should be informed by these principles. Chapters 4 and 5 elaborate on the lenses of formative assessment and feedback.

4 Methods

As we aim to design an intervention in several iterations, the research is based on the principles of design research (Van den Akker, 2006). The development throughout the cycles can be characterized by a shift in focus: from more qualitative formative towards quantitative summative. This involves upscaling from a small target audience towards a larger target audience, as indicated in Figure 2 (Tessmer, 1993). This pragmatic approach requires the methodology to be mixed: at first more qualitative, and more quantitative later on, using a more quasi-experimental approach with pre- and post-tests. Research will take place in one preparatory cycle and three subsequent cycles. Table 1 shows the study in its different cycles, with foci on the different research sub-questions.

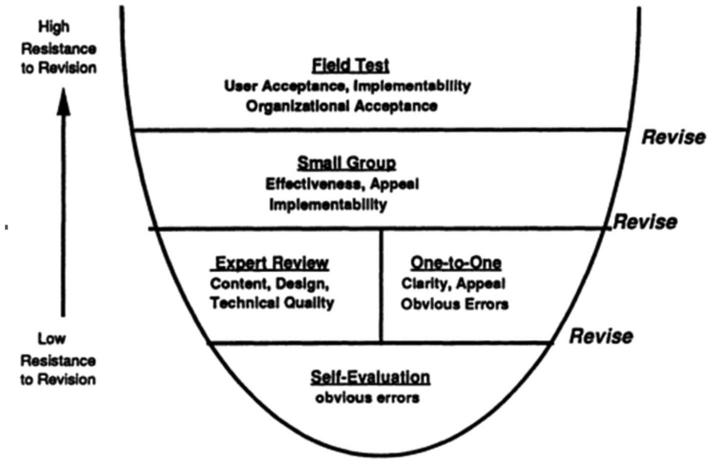


Fig. 2: Layers of formative evaluation (Tessmer, 1993)

Table 1: Cyclic setup of the study

When	Cycle	Description	Sub-questions(s)	N	Ch.
2008	Preliminary cycle	In this cycle we set out to formulate criteria and characteristics for algebra tools, and evaluated a large number of tools. This allowed us to choose the tool that best fit our research goals.	1a & 1b		2
Dec 2008	Cycle one	In this cycle we wanted to see how a prototypical design of a digital intervention would work in one-to-one sessions with students. This informed us about the potential of using students to improve feedback within the tool, and the way in which this could be done successfully.	2a, 2b & 2c	5	3+4
Jan/Feb 2010	Cycle two	In two cycles we implemented three main design principles in the initial prototype and studied the effects of the digital intervention, as well as the factors that influenced this outcome.	3a, 3b & 3c	31	5
End 2010	Cycle three		4a & 4b	324	6

5 Structure of the thesis

The PhD thesis comprises a series of articles each addressing a different perspective of this study on the use of ICT for acquiring, practicing and assessing algebraic expertise. These articles have been submitted to different journals. Figure 3 illustrates the structure of the thesis and shows which research question is answered in each chapter. This thesis consists of one introductory chapter, five chapters that reflect journal articles, and finally a concluding chapter. Chapter 2 focuses on criteria for and evaluation of algebra tools. It aims to give a clear picture of what characteristics are deemed important by experts (N=27) and as such provides a framework for studying the plethora of tools that are available. The results of this chapter inform the tool choice: why choose this tool and what characteristics of the tool were paramount. With this tool we designed a prototypical online intervention for algebraic expertise.

Chapter 3 reports on one-to-one think-aloud sessions with five 12th grade students. In these sessions, the prototype of the digital intervention was tested. In particular, it was investigated whether using the online activities would invite both procedural skills and symbol sense. In chapter 4 we report on how the same qualitative experiment also informed us about ways in which we could design feedback for our tool by studying student behavior. Based on the findings from previous chapters, the intervention was redesigned. We deployed the intervention in a second cycle for two classes (N=31), after which we made the final improvements based on several design principles.

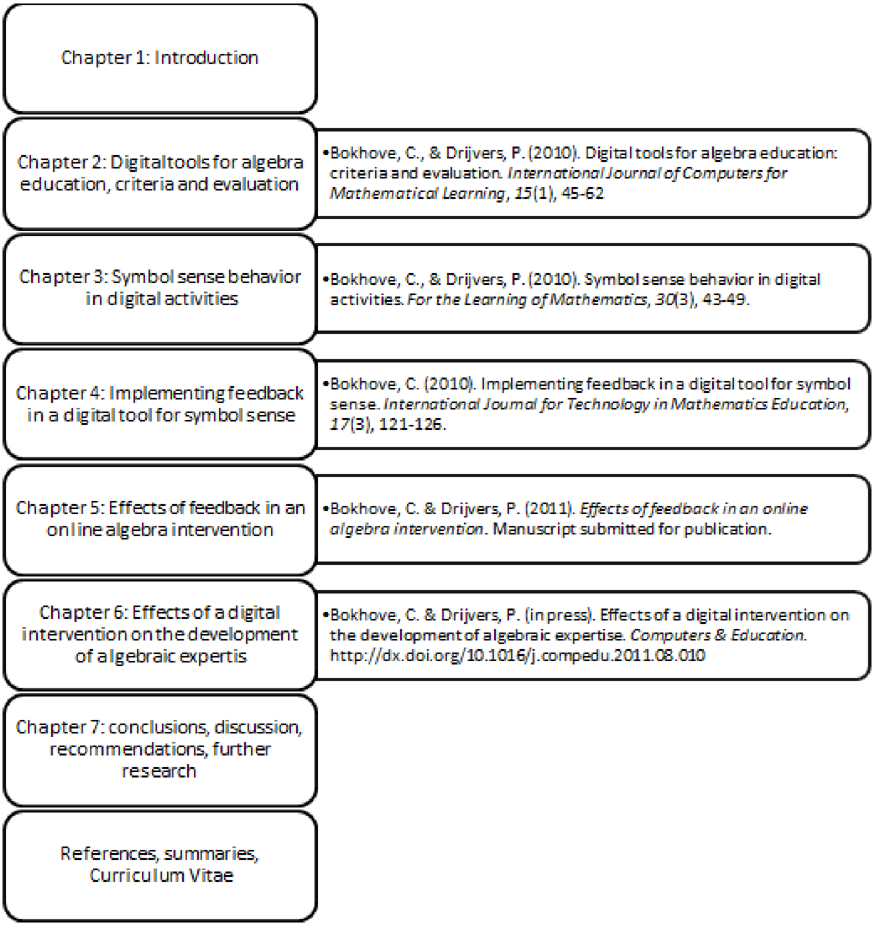


Fig. 3: Thesis outline

In chapter 5 we elaborate on the design principles we used and improved in an iterative approach, and the way these principles worked or didn't work in the third and last cycle in nine different schools (N=324), with students again from 12th grade. In chapter 6 we describe the overall effects of the online intervention in the third and last cycle (N=324). For the data analysis multilevel modeling was used.

Chapter 7 synthesizes the gains from the study and discusses the lessons to be drawn from its findings. Furthermore, it contains recommendations, discussion points and possible further research.

Chapter 2

Digital tools for algebra education: criteria and evaluation

Bokhove, C., & Drijvers, P. (2010). Digital tools for algebra education: criteria and evaluation. *International Journal of Computers for Mathematical Learning, 15*(1), 45-62.

Digital tools for algebra education: criteria and evaluation

1 Introduction

Currently, algebra education is the subject of worldwide discussions. Different opinions on goals, approaches, and achievements are at the heart of 'math war' debates (Klein, 2007; Schoenfeld, 2004). Crucial in these debates is the relationship between procedural skills and conceptual understanding in the teaching and learning of algebra. On the one hand, computational skills are seen as a prerequisite for conceptual understanding (National Mathematics Advisory Panel, 2007). Complaints from tertiary education focus on the lack of such procedural skills, and in several countries higher education makes a plea for entrance tests on basic algebraic skills (Engineering Council, 2000). On the other hand, some see the core of algebra education to be the development of strategic problem-solving and reasoning skills, symbol sense and flexibility, rather than procedural fluency (National Mathematics Advisory Panel, 2007). Future societal and professional needs lie in flexible analytical reasoning skills rather than in procedural skills, according to this point of view. As a consequence, algebra education should change its goals, focusing on new epistemologies and aiming at new types of understanding. This position is expressed in the Discussion Document of the 12th ICMI study on algebra education:

> An algebra curriculum that serves its students well in the coming century may look very different from an ideal curriculum from some years ago. The increased availability of computers and calculators will change what mathematics is useful as well as changing how mathematics is done. At the same time as challenging the content of what is taught, the technological revolution is also providing rich prospects for teaching and is offering students new paths to understanding. (Stacey & Chick, 2000, p. 216)

The above quote raises the issue of technology in algebra education. Educational technology plays a two-fold role in the discussion on the teaching and learning of algebra. First, the availability of technology challenges the goals of algebra education. How much procedural fluency is needed if computer tools can carry out the work for us? What types of skills are needed, or become increasingly important through the availability of technological tools? Second, technology offers opportunities for algebra education and in that sense is not only part of the problem, but might also be part of its solution. How can technological tools become integrated in algebra education so that they support the development of both meaning and procedural skills? To what new epistemologies and reconceptualizations can and should the integration of ICT lead, and what learning formats become feasible for teaching as well as for formative and summative assessment?

If the teaching and learning of algebra might benefit from the integration of technology, the subsequent question must be what type of technology to use, and what criteria determine this choice. Many different types of software tools are available, each providing opportunities and constraints for different activity structures and even different types of knowledge to emerge. It is not a straightforward issue to foresee these effects and to decide adequately on which tools to involve in the learning process and why. What is adequate, of course, depends on the goals of and views on algebra education, on knowledge acquisitions and learning, as well as on situational factors.

In the Netherlands as well, algebra education, and the relationship between skills and conceptual understanding in particular, is an important issue (Drijvers, 2006; Heck & Van Gastel, 2006; Tempelaar, 2007). Digital technologies offer opportunities to change epistemologies and activity structures and, as a consequence, to improve students in their process of meaning making and skill acquisition. In order to investigate these opportunities, we faced the challenge of deciding what tools to use. In our quest for an appropriate tool, clear identification of relevant tool properties and measurable criteria were needed, as well as making explicit our own goals and expectations. This led to the development of an instrument for the evaluation of digital tools for algebra education, which embodies the ideas on how digital technologies may enhance algebra education. The evaluation instrument consists of a set of criteria for such digital tools. The process of choosing and evaluating tools often remains implicit, including the even more implicit criteria that researchers or designers have while doing so. So the proposed evaluation instrument helps us to better and more consciously carry out the process of choosing tools, in a way that informs our research.The design process and use of this evaluation instrument is the topic of this chapter.

2 Conceptual framework and research questions

In order to design an instrument for the evaluation of technological tools for algebra education, a clear view on *the teaching and learning of algebra* is a first prerequisite. In particular, what does the looked-for algebraic expertise include? The distinction between procedural skills and conceptual understanding is helpful to frame the ideas on algebra education in this study. The book *Adding it up* (Kilpatrick, Swafford, & Findell, 2001) synthesized the research on this issue. The central concept is mathematical proficiency, which consists of five strands: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning and productive disposition. Here, conceptual understanding is defined as “comprehension of mathematical concepts, operations, and relations” (p. 116), and procedural fluency as “skill in carrying out procedures flexibly, accurately, efficiently, and appropriately” (ibid.). Furthermore, “the five strands are interwoven and interdependent in the development of proficiency in mathematics” (ibid.).

Algebraic expertise thus includes both procedural skills and conceptual understanding. To capture the latter, the notion of symbol sense is powerful (Arcavi, 1994). Arcavi (1994) defined symbol sense as “an intuitive feel for when to call on symbols in the process of solving a problem, and conversely, when to abandon a symbolic treatment for better tools” (p. 25), or, in the words from Zorn (2002), “a very general ability to extract mathematical meaning and structure from symbols, to encode meaning efficiently in symbols, and to manipulate symbols effectively to discover new mathematical meaning and structure” (p. 4). This is developed further in the concepts of structure sense and manipulation skills (Hoch & Dreyfus, 2004). Procedural skills and symbol sense are intimately related: understanding of concepts makes procedural skills understandable, and procedural skills can reinforce conceptual understanding (Arcavi, 2005). In this study, therefore, we focus on an integrated approach of algebraic expertise, and on the co-development of procedural skills and symbol sense.

As well as having a view on algebra education, a study on the *role of technology* in algebra education should also be clear about its view on technology. What is the role of technological tools in the teaching and learning of algebra? Speaking in general, technology is considered as a potentially important tool for learning mathematics. The National Council of Teachers of Mathematics states:

> Technology is an essential tool for learning mathematics in the 21st century, and all schools must ensure that all their students have access to technology. Effective teachers maximize the potential of technology to develop students’ understanding, stimulate their interest, and increase their proficiency in mathematics. When technology is used strategically, it can provide access to mathematics for all students. (National Council of Teachers of Mathematics, 2008)

In line with this, the general hypothesis underpinning this study is that the use of ICT tools, if carefully integrated, can increase both algebraic skill performance and symbol sense in the classroom. The use of technological tools in mathematics education is a specific case of tool use in general, which is an integrated part of human behaviour. Vygotsky (1978) saw a tool as a new intermediary element between the object and the psychic operation directed at it. Verillon and Rabardel (1995) distinguished artefact and instrument. Instrumental genesis, then, is the process of an artefact becoming an instrument, including mental schemes for using the tool, containing both conceptual and technical knowledge. As an aside, we remark that the use of the expression ‘evaluation instrument’ in this chapter does not refer to the instrumental framework just described, but to its more general meaning in educational science of a ‘tool to measure’.

More specific for algebra education, several studies (Artigue, 2002; Guin, Ruthven, & Trouche, 2005) showed that instrumental genesis is a time-consuming and lengthy process. Adapted form of Chevallard’s (1999) framework, Kieran & Drijvers (2006) stressed the need for congruence between tool techniques and paper-and-pencil techniques and use a Task-Technique-Theory triad to capture the relationship between tool techniques and conceptual understanding. The congruence between tool techniques and paper-and-pencil techniques, therefore, is an important criterion for useful ICT application. Furthermore, activities with technological tools should not address procedural skills in isolation, but should offer means to relate procedural techniques and symbol sense insights. Activity structures that exploit these opportunities can affect students' epistemologies and knowledge acquisition in a positive manner.

Of particular interest for its value for algebra education is the tool's *potential for providing feedback*, an essential condition for supporting student learning and improving chances of success (Gibbs & Simpson, 2004; Hattie & Timperley, 2007; Nicol & MacFarlane-Dick, 2006). The feedback can be direct (Bokhove, Koolstra, Heck, & Boon, 2006), but also implicit, for example by providing the possibility of combining multiple representations (Van Streun, 2000). Feedback is crucial if we want technology to act as a learning environment in which students can engage in a process of practice or meaning making without the help of the teacher. If feedback on learning activities is actually used to modify teaching to meet the learner's needs, one of the conditions for formative assessment is fulfilled (Black & Wiliam, 1998). Formative assessment is assessment *for* learning rather than summative assessment *of* learning. Black and Wiliam (ibid) made a case for more room for formative assessment and claim that improving formative assessment raises standards. Formative assessment is an essential part of the curriculum and has the key benefit of motivating students through self-assessment. Therefore, for the purpose of designing an evaluation instrument, particular attention is paid to formative assessment and feedback characteristics of a tool.

The question now is how this conceptual framework, with its three key aspects from algebra didactics (symbol sense), theories on tool use (instrumental genesis) and assessment (feedback and formative assessment), can guide the choice of a digital tool for algebra that offers good opportunities for developing new epistemologies and improved symbol sense to the students. More precisely, the research questions that we address in this chapter are as follows:
(1) Which criteria are relevant for the evaluation of digital tools for algebra education?
(2) Which digital algebra tool best meets these criteria?

3 Methods

For the *design of the evaluation instrument* a modified Delphi process was used (Hearnshaw, Harker, Cheater, Baker, & Grimshaw, 2001). As a first step in the design process, the research team drew up an initial set of criteria for digital tools for algebra education. This set was informed by the conceptual framework described above, which resulted in the criteria being grouped into three main categories algebra, tool and assessment. For example, principles of good feedback practice yielded criteria for the use of feedback, placed in the assessment category. Criteria of a more general and often practical nature (e.g. cost of the software) were put in a fourth, general category.

Items for this initial set were selected from literature sources by the researchers. Some of these sources concerned mathematics, or algebra in particular: cognitive fidelity (Bliss & Ogborn, 1989), mathematical fidelity (Dick, 2007) and expressive/exploratory environments (Beeson, 1989). Others were based on design choices reported by the designers of the software Aplusix (Nicaud, Bouhineau, & Chaachoua, 2004; Nicaud, Bouhineau, Chaachoua, & Trgalova, 2006). General criteria for educational applets were found in Underwood et al. (2005). Principles on authoring facilities for teachers, addressing the needs of the 'neglected learners', were addressed by Sangwin and Grove (2006) and are in line with Dick's opinion that the possibility for teachers to author content themselves could bring tool and pedagogical content together (Dick, 2007). Finally, several sources concerned the third component of the conceptual framework, namely assessment. Amongst others, the seven principles of good feedback practice (Nicol & MacFarlane-Dick, 2006) and the eleven conditions under which assessment supports students' learning defined by Gibbs and Simpson (2004) were considered. Assessment also involved the types of feedback distinguished by the University of Waterloo (2000). In addition to these literature sources, the researchers used their experience from past projects on the use of technology for algebra. This process resulted in an initial set of criteria, grouped into four categories: algebra criteria, tool criteria, assessment criteria and general criteria. The first three criteria were linked to our conceptual framework, the last category contained general characteristics.

A second step in the design of the evaluation instrument involved the external validation of the initial set, including a check for completeness and redundancy. As not all criteria are supposed to be equally important, we also wanted weights to be attributed to each of the items, reflecting their relative importance. Therefore, the evaluation instrument was sent to 47 national and international experts in the fields of mathematics and algebra education, educational use of technology, and/or assessment. The experts were identified through their contributions to research literature in this field. Out of the 47, 33 experts responded, six of whom qualified themselves as not knowledgeable enough, or not willing to comment. The remaining 27 experts rated the importance of every criterion on a Likert scale from 1 to 5, denoting 'unimportant', 'slightly unimportant', 'neutral', 'slightly important', 'important' and the option to give no answer. These scores were used to determine the relative weights of the individual criteria. Other approaches than using Likert scales, like ranking and constant sum methods, were rejected as they would not answer our main question: is the list of criteria complete? In order to address completeness, experts were asked to comment on the thoroughness of the list and to add criteria that they found to be missing. This information enabled us to deduct which criteria should be included into the evaluation instrument and provided insight into the relative importance of these criteria.

A next step was to *use the evaluation instrument* in the process of selecting a technological tool for the research study on the learning of algebra. In order to find out which ICT tool best met the criteria according to the evaluation instrument, a 'longlist' of such tools was compiled. The research team set up this list by consulting different sources, such as the work of the Special Interest Group Mathematics on assessment standards (Bedaux & Boldy, 2007), a research study on digital assessment (Jonker, 2007), the Freudenthal Institute's mathematics wiki[1] on digital assessment and math software, and Google searches. Also, experiences from previous research projects were included. As there are hundreds of tools, the research team needed to filter out some tools that were not appropriate for algebra education. For this, the tool's main functionality was first considered. For example, a geometry tool with very limited algebra support was excluded. This yielded a longlist of 34 ICT tools.

To reduce the longlist to a shortlist of 'nominations', the researchers chose four criteria from the evaluation instrument as a prerequisite for further investigation:

- Math support: formulas should be displayed correctly in conventional mathematical notation and algebraic operations should be supported. This enhances congruence between tool techniques and paper-and-pencil technique.
- Authoring capability, configurability: because we wanted to use the tool for our own purposes, teachers or researchers should be able to add or modify content. This also enhances fidelity (Dick, 2007).
- Data storage: it was considered essential that the tool could be used anytime, anyplace, and that student data was stored centrally, so that analysis could take place.
- Technical support: it was important that the tool was supported and that continuity was guaranteed.

Based on these four requirements, the longlist was reduced to a shortlist of seven ICT tools. To be on the shortlist, the tool should at least feature *all* of these four criteria.

Next, the seven 'nominations' at the shortlist were considered in more detail. After gathering more information and installing the tool, a first evaluation consisted of using the tool with already existing content. Quadratic equations were used as a test topic, as this is a subject that is addressed in most educational systems. Next, we used the tool for authoring the content we intended to use in our further research, while keeping logs through screenshots. Finally, we graded each of the tools on every criterion of the instrument in a qualitative way, i.e. on a 5-point scale ranging from 1 to 5. This resulted in separate descriptions for each of the seven tools, and a matrix, providing an overview of the tools' strong and weak points.

These results were validated through agreement analysis. A second coding was done by an external expert, who individually coded 2 out of 7 tools (28% of all items, PRE). Next, the researcher and the external expert discussed the ratings and eventually revised them (POST). Only obvious lacks of domain knowledge were corrected in the POST analysis. The level of agreement was calculated with Krippendorff's alpha (De Wever, Schel-

1. http://www.fi.uu.nl/wiki/index.php/Categorie:Ict

lens, Valcke, & Van Keer, 2006). This yielded a value of .65 for the PRE ratings and .86 for the POST ratings. The improvement of Krippendorff's alpha was due to original differences in understanding criteria. For example, one discrepancy in score was explained by the fact that the external expert rated one of the tools as a tool without assessment modes, whereas the researcher took into account the possibilities of Moodle's[2] quiz module, which formed an integral part of the tool. Another explanation could be a bias factor, which we will address in the discussion section.

4 Results

Design of the instrument: categories and weights

A first result of this enterprise is the evaluation instrument itself, which is organized around the three key elements of the conceptual framework: algebra, tool use, and assessment. Furthermore, a fourth category with general, factual criteria is included.

The criteria operationalize several aspects of the conceptual framework. For algebra, for example, the link between the instrument and the conceptual framework is manifest in criterion 6: *The tool is able to check a student's answer on equivalence through a computer algebra engine.* According to this criterion, the tool is able to recognize algebraic equivalence. This corresponds to our desire to be able to detect both symbol sense and procedural skills, and to identify different problem solving strategies with equivalent results. A criterion that exemplifies the relation between the conceptual framework and the tool criteria of the evaluation instrument, criterion 10 states: *The tool is easy to use for a student (e.g. equation editor, short learning curve, interface).* As congruence between tool techniques and paper-and-pencil techniques is an important theoretical notion, we want students to be able to use the same mathematical notations as on paper. Within the assessment category, criterion 18 – *The tool caters for several types of feedback (e.g. conceptual, procedural, corrective)* – reflects the relevance of feedback, as it was identified as an essential prerequisite for formative assessment.

Appendix A shows the complete instrument, including 27 criteria and their weights that resulted from the expert review. The individual weights resulted in the category weights presented in table I.

Table 1: Weights of the four categories

Category	Weight
Algebra criteria	4.34
Tool criteria	4.39
Assessment criteria	4.35
General criteria	4.17

2. Moodle is an open source Virtual Learning Environment, http://www.moodle.org

These results show that the experts valued the different categories as more or less equally important. Only the category of general criteria scored slightly lower on average. For considering the scores for individual items, table II shows the top five of important criteria with their accompanying weight means, as well as the bottom five.

Table 2: The top five of most important and bottom five of least important criteria

Rank	Description	Weight
1	The stability and performance of the tool	4.89
2	The tool is easy to use for a student (e.g. equation editor, short learning curve, interface)	4.85
3	The tool is able to display formulas correctly	4.81
4	The tool is mathematically sound and faithful to the underlying mathematical properties (e.g. conventional representations, sound operations)	4.74
5	The tool stores the answers given by a student	4.70
...		...
23	The tool has the ability to randomize algebra assignments	3.96
24	The cost of the tool	3.74
25	The tool makes use of standards (e.g. QTI, SCORM)	3.72
26	The tool enables the student to use a computer algebra system as a tool	3.63
27	The licensing of the tool (e.g. open, proprietary)	3.41

Overall, the expert review shows a large level of agreement on the criteria. All criteria have an ‘above neutral’ weight. The least important criterion still has an average weight of 3.41, qualifying it slightly more important than neutral. No extra criteria were suggested by the experts.

Application of the instrument

Now that the criteria and their weights are established, we use this instrument to categorize and evaluate digital tools for algebra education. The first, inventory round of this evaluation resulted in a longlist of 34 digital algebra tools. Applying the minimal requirements yielded a shortlist of seven tools. Some interesting tools, ticking almost all the boxes on the checklist, did not meet the minimal requirements. In the case of Aplusix (Nicaud et al., 2004), for example, the necessity to install the software locally on the computer implies no central data storage. Several web-based tools had company backing (thus continuity) and good support for mathematical formulae, but lacked the feature to author tasks. The shortlist of seven tools consisted of Wims, STACK, Maple TA, Digital Mathematics Environment (DME), Wiris, Activemath, and Webwork. Appendix B provides a data sheet with more information on each of these tools.

We rated the seven tools on each of the 27 instrument criteria. Table III gives the scores for each of the four instrument categories. These scores were calculated by adding up the weight from the evaluation instrument (see appendix A) multiplied by the score for each criterion. The scores were standardized by taking into account the number of criteria per category, to avoid criteria from a smaller category having a relatively smaller weight than a criterion from a larger category.

Table 3: Scores of the seven digital tools for assessing algebraic skills

Tool → **Item category**	**WIMS**	**STACK**	**MAPLE TA**	**DME**	**WIRIS**	**ACTIVE MATH**	**WEBWORK**
Algebra total	138.48	133.51	101.51	154.39	127.24	137.45	117.71
Tool total	105.77	106.16	140.02	145.69	92.53	99.62	118.62
Assessment total	111.52	112.40	125.27	119.11	34.80	97.26	125.27
General total	130.72	100.74	75.95	103.21	103.26	112.05	102.27
Total score	486.49	452.82	442.76	522.40	357.82	446.37	463.87
Rank	2	4	6	1	7	5	3

The results show that the digital algebra tool DME obtained the highest overall score. In the both the algebra and tool category, DME obtained the highest score. The assessment category yielded highest scores for both Maple TA and Webwork. Finally, Wims scored highest in the general category.

5 Criteria exemplified

This section aims to exemplify some characteristic criteria from the evaluation instrument described earlier, and illustrate the differences and similarities of the evaluated tools. We provide examples for the three main item categories (algebra, tool and assessment criteria) as well as for each of the seven tools on the shortlist. It is by no means a report on all the ratings of the criteria for all the tools. An overview of the ratings is available online[3].

Algebra criteria exemplified

The first evaluation criterion in the algebra category is stated as follows: the tool enables the student to apply his or her own paper-and-pencil reasoning steps and strategies. This criterion concerns how well a tool can be used ‘the same way as paper and pencil’. As discussed in the conceptual framework section, this criterion reflects the following underlying assumption: if we want students to acquire an integrated perception of algebraic skills,

3. http://www.fi.uu.nl/~christianb

the techniques for using the tool have to resemble the way students use algebra with paper and pencil. Offering options to apply stepwise strategies within the tool is rated as ‘better’ than not being able to apply these steps.

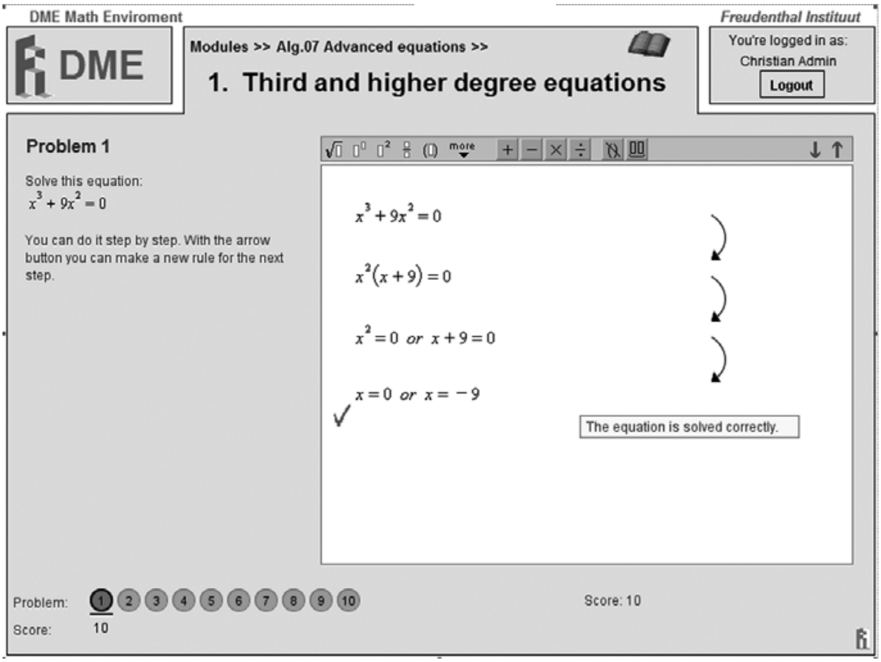


Fig.1: Stepwise strategy in the DME

This criterion is exemplified in figure 1 for the case of the DME tool. Figure 1 shows that the student can choose what step to apply next to the equation to be solved. The tool enables the student to use a stepwise problem solving approach to get to his/her answer.

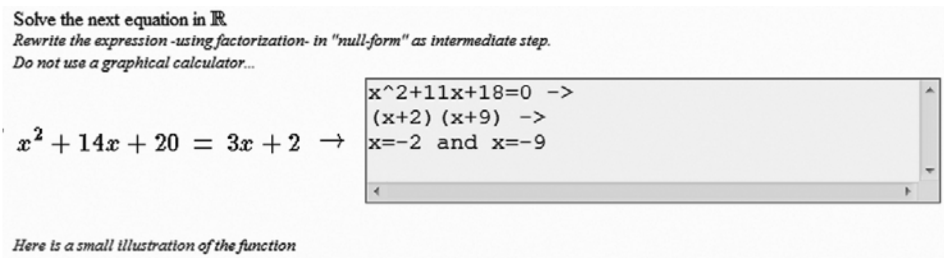


Fig.2: Steps in Wims

Every step is evaluated on correctness – which is a criterion in the assessment category. Many tools just enable the student to give one (final) answer. The way in which the tools support a stepwise approach differs, as is shown in the Wims screen displayed in figure 2. Here the student can enter more than one algebraic step, starting with the equation that has to be solved. The system evaluates the whole sequence of steps after submitting the solution.

Evaluation criterion 5 in the Algebra category is: The tool has the ability to combine questions into larger units to enable multi-component tasks. Many algebra assignments consist of several sub-items. Together these items form a more complex assignment. It is important that a tool can cater for such multi-component tasks, not only in the assignment text, but also in grading the answers to sub-questions and providing adequate feedback. Several examples of multi-component tasks are implemented in STACK. Figure 3 shows a task consisting of three parts. An incorrect answer to question 1 would lead to incorrect answers for 2 and 3 as well. By combining the three questions into one logical unit STACK is able to 'follow through' a mistake made in question 1.

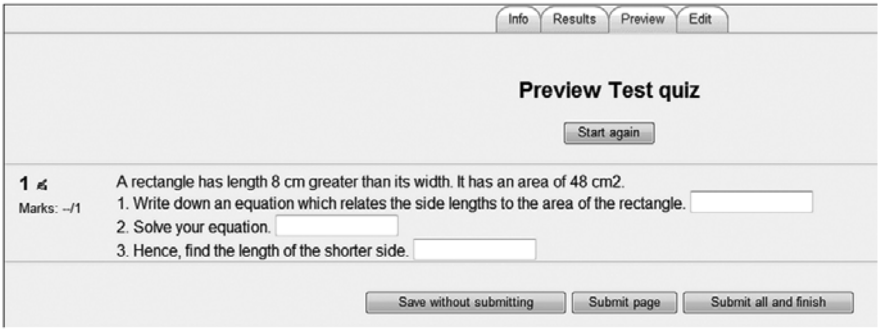


Fig.3: Example of a multi-component question that can be authored in STACK and Moodle.

Tool criteria exemplified

Evaluation criterion 10 concerns the tool. It is stated as follows: The tool is easy to use for a student (e.g. equation editor, short learning curve, interface). The use of a tool needs to be very intuitive, as using a tool should be a question of 'use to learn' instead of 'learn to use'. This criterion links to the congruence between tool techniques and paper-and-pencil technique, so that students are able to use the same mathematical notations and techniques in the technological environment as on paper (Kieran & Drijvers, 2006), as well as to the notion of instrumental genesis described in the conceptual framework. The example in figure 4 shows WIRIS providing an intuitive interface with notations that resemble conventional mathematics representations.

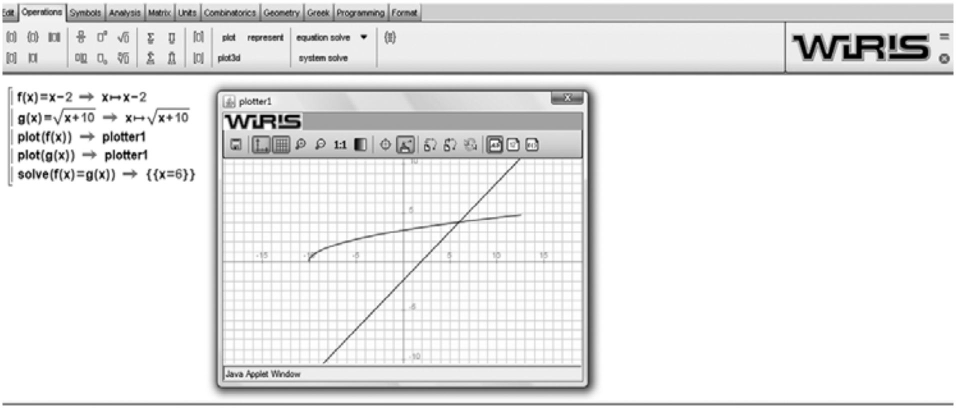


Fig.4: Wiris' graphical user interface

Evaluation criterion 14 also concerns the tool: the tool provides the author/teacher with question management facilities. Using a tool in an assessment setting means being able to add, copy and move items, perhaps from and into so-called item banks. When these facilities are lacking or are inadequate, constructing digital tests, be they formative or summative, will be painstaking and slow. We therefore contend that digital algebra tools need to provide easy-to-use question management facilities.

A relevant example is provided in figure 5. In Maple TA, a test is constructed by choosing questions from 'Question Banks'. These question banks can be exchanged between users of the program. This approach makes it possible to reuse questions.

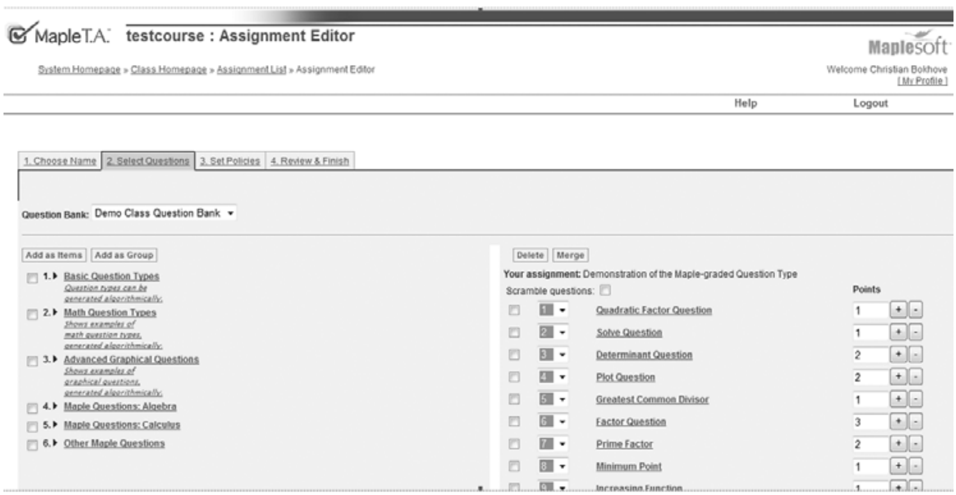


Fig.5: Question banks in Maple TA

Evaluation criterion 16 also concerns the tool, and says: The tool has readily available content. Not every teacher wants to make his or her own content. Using ready available content can be convenient for many teachers. For example, figure 6 shows that Webwork – in use at many universities in the United States – comes with a very large database of questions at university level.

Fig. 6: A massive amount of readily available content in Webwork

Assessment criteria exemplified

Evaluation criterion 17 concerns the assessment focus within the conceptual framework. It is stated as follows: The tool provides several assessment modes (e.g. practice, test). Providing more than just summative testing, scoring and grading is an important prerequisite for formative assessment, which is identified as important in the conceptual framework. Therefore, providing several 'modes' to offer questions and content is considered an important feature of software for formative assessment of algebraic expertise.

As an example, figure 7 shows that Wims provides several modes when using the tool: training, total control over the configuration, paper test (providing a printed version of the test), practice digital test, actual digital test, all deep HTML links on one page (for use in one's own Virtual Learning Environment).

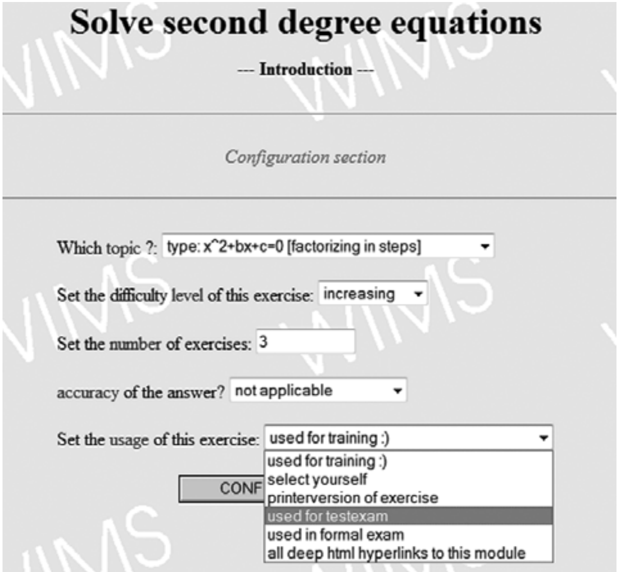


Fig. 7: Assessment modes in Wims

Evaluation criterion 19 also concerns assessment and says: The tool takes the student's profile and mastery account and serves up appropriate questions (adaptivity). Adaptivity is useful for providing user-dependent content. A student who does not know a topic well will be presented with more tasks and exercises on that subject. Students who display mastery of a subject will not be served remedial questions. Figure 8 shows how such adaptivity is implemented in a learner model within Activemath. The system 'knows' what the student does or does not know. It can also take into account several learning styles, and serves up appropriate content based on these variables.

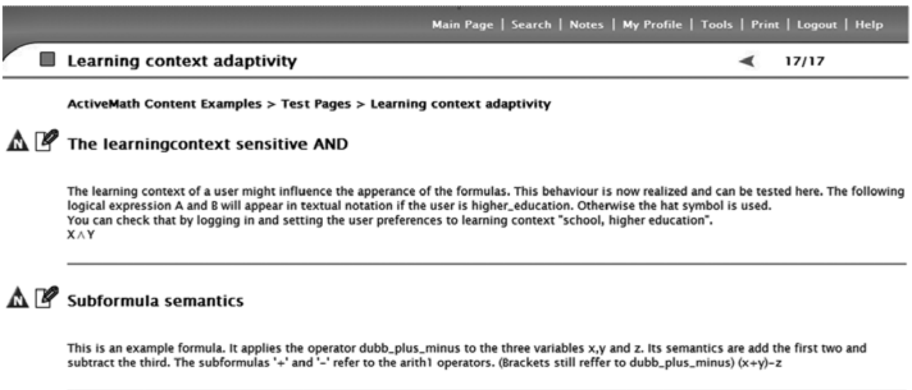


Fig. 8: Adaptive content in Activemath

6 Conclusions and discussion

In this chapter we set out to answer two research questions. The first one concerns the identification of criteria that are relevant for the evaluation of digital tools for algebra education. We constructed an evaluation instrument consisting of 27 criteria grouped in four categories. The categories were based on a conceptual framework that matched the goals and intentions of the study, and consisted of three key aspects from algebra didactics (symbol sense), theories on tool use (instrumental genesis) and assessment (feedback and formative assessment). A fourth category concerned general characteristics. The modified Delphi approach, conducted to validate the criteria, revealed a large agreement among external experts on these criteria. The weights of the criteria led to the identification of the most important ones: stability and performance, correct display of mathematical formulas, ease of use, mathematical soundness, and storage of the work. We conclude that the designed instrument provides a good evaluation instrument for describing characteristics of digital tools for algebra that we consider relevant for the purpose of our study. The instrument provides insight in the different features of a tool, as well as in our priorities in interest. It can also be very helpful for software development in mathematics education, especially the ones regarding algebra education.

The second question at stake is which digital algebra tool best meets these criteria. Using the evaluation instrument, we rated seven tools that met the minimal criteria and had our codes validated by an external expert scoring. We conclude that the Digital Mathematics Environment scores highest overall and thus is best suited for addressing the research goals on the co-emergence of procedural skill fluency and symbol sense expertise. A key feature from DME is that it enables stepwise problem solving strategies. It is easy to use, stores the solution process of the student, and is well suited for formative assessment, as it offers several strategy modes, feedback and self-review.

Reflecting on these conclusions, a first remark to be made is that the actual process of designing our evaluation instrument helped greatly in listing important characteristics for digital tools for algebra education. The process helped in transforming our conceptual framework into a set of concrete and applicable criteria, and made these criteria tangible by looking at a set of available tools. It also helped us to better and more consciously carry out the process of choosing tools, in a way that informed our research. These transformation and operationalization aspects were somewhat unexpected, as we initially just set out to 'choose a tool', but in retrospect we find them extremely valuable. The resulting instrument can now serve as a means to identify tool characteristics and help choosing the most suitable tool, depending on the educational or scientific context. While doing this, it remains important to take heed of the questions raised most frequently by the experts during the validation process of the evaluation instrument, such as 'what target audience is assessed?' and 'which algebraic skills are tested?' This shows that, even if the criteria for the instrument presented here can be applied in many contexts, the weights that are given to them greatly depend on this context and its educational goals and aims. In our case, this context is upper secondary education, and the goal is to integrate procedural skills and symbol sense expertise. These differences in contexts and goals make it difficult to really compare tools. The instrument and the description of how to design and validate, however, do provide a blueprint of criteria that might be considered and of a process that might be gone through when choosing a digital tool for algebra education.

In line with this, a second issue raised by the external experts is that formative assessment is never an isolated activity, but is rooted in a social and educational context. The benefits of using a digital tool for algebra also depend on classroom dynamics and factors such as gender distribution and culture. We think that the designed evaluation instrument should always be used with an awareness of the context in which the tool is going to be used. For example, if the research takes place in a context in which classroom teaching is the predominant paradigm, the 'anytime, anyplace' criterion can be considered as less important than in a context of distance learning.

Finally, as a methodological limitation we notice that rating the different digital tools requires a profound knowledge of and familiarity with each of the tools, which is difficult to acquire for one single researcher. This difficulty emerged in establishing the inter-rater reliability, with the expert reviewer being very familiar with one specific tool and less familiar with some other tools. This clearly complicates comparative studies of digital tools. Ideally, all coders should have an extended domain knowledge of all the tools that are available. Even if we tried to deal with this issue through detailed study of each of the tools, this is a methodological limitation.

Appendix A List of criteria, weights and rankings

Criteria	Average weight	Rank
1. Algebra category		
This category contains criteria related to mathematics and algebra in particular.		
The tool enables the student to apply his or her own paper-and-pencil reasoning steps and strategies.	4.26	17
The tool is mathematically sound and faithful to the underlying mathematical properties (e.g. conventional representations, sound operations).	4.74	4
The tool's openness enables the student to express mathematical ideas and strategies.	4.54	7
The tool has the ability to randomize algebra assignments.	3.96	23
The tool has the ability to combine questions into larger units to enable multi-component tasks.	4.32	15
The tool is able to check a student's answer on equivalence through a computer algebra engine.	4.42	12
The tool enables the student to use a computer algebra system as a tool.	3.63	26
The tool is able to display formulas correctly.	4.81	3
2. Tool category		
This category contains criteria related to tool use.		
The tool has an authoring function that enables teachers to add or modify content. (e.g. questions, texts, links, graphs, feedback).	4.33	14
The tool is easy to use for a student (e.g. equation editor, short learning curve, interface).	4.85	2
The tool is accessible anytime, anywhere.	4.15	21
The tool stores the answers given by a student.	4.70	5
The tool stores the solution process of the student.	4.63	6
The tool provides the author/teacher with question management facilities.	4.26	17
The tool makes use of standards (e.g. QTI, SCORM).	3.72	25
The tool has readily available content.	4.48	11
3. Assessment category		
This category contains criteria related to assessment.		
The tool provides several assessment modes (e.g. practice, test).	4.19	20
The tool caters for several types of feedback (e.g. conceptual, procedural, corrective).	4.52	8
The tool takes the mastery and profile of the student into account and serves up appropriate questions (adaptivity).	4.22	19

The tool has a review mode showing what the student has done wrong or right.	4.52	8
The tool allows for the use of several question types (e.g. multiple choice, open).	4.30	16
4. General category		
This category contains general criteria for a digital tool for education.		
The cost of the tool.	3.74	24
The licensing of the tool (e.g. open, proprietary).	3.41	27
The technical support of the tool.	4.48	10
The languages the tool supports.	4.11	22
The stability and performance of the tool.	4.89	1
The structure and attractiveness of the tool’s interface.	4.37	13

Appendix B Tool data sheet

Name	Digital Mathematical Environment (DME)
Date	12-01-2009
Version	Version 12-01-2009
URLs	Several tests were made in a special environment for secondary and higher education topics: http://www.fi.uu.nl/dwo/voho
Features, characteristics	Strong points are the performance, multi-steps within exercises plus feedback, authoring capability, SCORM export. Disadvantages: emphasis on algebra (not extendable, dependent on the programmer), source code not available.

Name	WIMS
Date	12-01-2009
Version	3.57 (3.64 released)
URLs	http://wims.unice.fr/ For Dutch content: http://wims.math.leidenuniv.nl/wims
Features, characteristics	Wims probably is the quickest and most complete one of the tools investigated, with features for geometry, algebra, etc. A fair amount of Dutch content is available. It is let down by the feedback and the fact only one answer can be entered. Of course this can be programmed, as it is a very powerful package, but here we see a steep learning curve.

Name	STACK
Date	12-01-2009
Version	2.0
URLs	http://www.stack.bham.ac.uk/
Features, characteristics	STACK has a good philosophy with ‘potential responses’ and multistep questions. Also, the integration with a VLE -unfortunately only moodle- is a plus. Installation, stability and performance is a negative (slow and cumbersome), as well as its looks. It is also very experimental, providing almost no continuity

Name	Maple TA
Date	12-01-2009
Version	4.0
URLs	http://www.maplesoft.com/ Dutch distributor: http://www.candiensten.nl/software/details.php?id=26 Dutch Mathmatch content: http://www.mathmatch.nl/onderwerpen.diag.php

Features, characteristics	Maple TA has many points that STACK has, but with better looks and no real support for 'potential responses' and ' multistep questions'. They can be programmed, but this means -like WIMS- coping with a steep learning curve. As it has its roots in assessment software question types are well provided for. One could say that Maple TA started as assessment software and his moving towards software for learning, while STACK started with learning and is moving towards assessment software

Name	Wiris
Date	12-01-2009
Version	2
URLs	http://www.mathsformore.com/ http://www.wirisonline.net/
Features, characteristics	Wiris is an attractive tool for standalone use within for example Moodle. However, the lack of assessment functions means it is not suitable enough for our research

Name	Activemath
Date	12-01-2009
Version	1.0
URLs	http://www.activemath.org/
Features, characteristics	Activemath is more of an Intelligent Tutoring System than a tool for assessment. The question module (ecstasy) is powerful, providing transition diagrams. However, authoring and technical aspects make it less suitable for the key aspects we want to observe

Name	WebWork
Date	12-01-2009
Version	2.4.1
URLs	http://webwork.maa.org/moodle/
Features, characteristics	Webwork has a high score for assessment, and average scores for the other items. This makes it an all-round tool. Webwork has a reasonable amount of readily available content.

Chapter 3

Symbol sense behavior in digital activities

Bokhove, C., & Drijvers, P. (2010). Symbol sense behavior in digital activities. *For the Learning of Mathematics, 30*(3), 43-49.

Symbol sense behavior in digital activities

1 Introduction

During the last twenty years the relationship between procedural skills and conceptual understanding has been widely debated. This relationship plays a central role in the ‘Math wars’ discussions (Schoenfeld, 2004). An important issue in this debate is how students best acquire algebraic expertise: by practicing algorithms, or by focusing on reasoning and strategic problem solving activities. The first approach sees computational skills as a prerequisite for understanding mathematical concepts (US Department of Education, 2007). In the latter approach, the focus is on conceptual understanding (ibid.). Even if the idea is shared that both procedural skills and conceptual understanding are important, there are disagreements on their relationship and the priorities between the two.

The last decades can also be characterized by the advent of the use of technology in mathematics education. In its position statement the National Council of Teachers of Mathematics (2008) acknowledges the potential of ICT for learning. The advance of technology may strength the relevance of ‘real understanding’ in mathematics (Zorn, 2002). Still, there is a firm tradition of educational use of ICT for rote skill training, often referred to as ‘drill and practice’; for symbol sense skills, such a tradition is lacking. The issue at stake, therefore, is twofold: how can the development of procedural skills and symbol sense skills be reconciled, and how can the potential of ICT be exploited for this ambitious goal?

2 Procedural skills ‘versus’ conceptual understanding

The distinction between procedural skills and conceptual understanding is a highly researched field of interest. The book *Adding it up* (Kilpatrick, Swafford, & Findell, 2001) synthesizes the research on this issue in the concept of *mathematical proficiency*, which comprises five strands: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning and productive disposition. Here, conceptual understanding is defined as “the comprehension of mathematical concepts, operations, and relations” (p. 116), and procedural fluency as the “skill in carrying out procedures flexibly, accurately, efficiently, and appropriately” (ibid.). Furthermore, “the five strands are interwoven and interdependent in the development of proficiency in mathematics” (ibid.).

Two papers in *For the learning of mathematics*, written by Arcavi, provided a breakthrough in the thinking on procedural skill and conceptual understanding in algebra (Arcavi, 1994, 2005). Arcavi (1994) introduces the notion of *symbol sense*, which includes “an intuitive feel for when to call on symbols in the process of solving a problem, and conversely, when to abandon a symbolic treatment for better tools" (p. 25). Using appealing examples, Arcavi describes eight behaviors in which symbol sense manifests itself. The examples show the intertwinement between procedural skills and conceptual understanding as complementary aspects of algebraic expertise. Both procedural skills and symbol sense need to be addressed in algebra education, as they are intimately related: understanding of concepts makes basic skills understandable, and basic skills can reinforce conceptual understanding (Arcavi, 2005).

In line with the work of Arcavi, Drijvers (2006) sees algebraic expertise as a dimension ranging from basic skills to symbol sense (see figure 1). Basic skills involve procedural work with a local focus and emphasis on algebraic calculation, while symbol sense involves strategic work with a global focus and emphasis on algebraic reasoning.

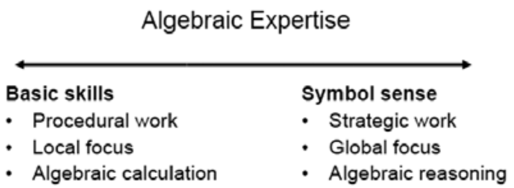


Fig. 1: Algebraic expertise as a dimension (Drijvers, Goddijn & Kindt, 2010)

One of the behaviors identified by Arcavi (1994) concerns flexible manipulation skills. It includes the versatile ability to manipulate expressions, not only technically but also with insight, so that the student is in control of the work and oversees the strategy. Two important, and interlinked, characteristics of flexible manipulations skill behavior are the gestalt view on algebraic expressions (Arcavi, 1994) and appropriate ways to deal with their visual salience (Wenger, 1987; Kirshner & Awtry, 2004).

A *gestalt view on algebraic expressions* involves the ability to consider an algebraic expression as a whole, to recognize its global characteristics, to 'read through' algebraic expressions and equations, and to foresee the effects of a manipulation strategy. Arcavi (1994) claims that having a gestalt view on specific expressions makes symbol handling more efficient, and emphasizes that 'reading through' expressions can make the results more reasonable. A gestalt view on algebraic expressions is a prerequisite for carrying out basic procedural skills and for deciding which type of manipulation to perform.

Flexible manipulation skills also involve dealing with visual cues of algebraic expressions and equations, their so-called *visual salience*. Kirshner and Awtry (2004) provide a definition of visual salience and tabulate several expressions with greater and lesser visual salience, respectively. They claim that "visually salient rules have a visual coherence that makes the left- and right-hand sides of the equation appear naturally related to one another" (p. 11). This coherence is strengthened by two properties of the equation under consideration: (i) repetition of elements across the equal sign, and (ii) a visual reparsing of elements across the equal sign (Awtry & Kirshner, 1994). Visual reparsing "manifests itself as a dynamic visual displacement of elements" (p. 11). Take for example:

(A) $\left(\frac{x}{y}\right)\left(\frac{w}{z}\right) = \frac{xw}{yz}$, and

(B) $(x - y) + (w - z) = (x + w) - (y + z)$

In identity A, the right hand side seems to follow immediately from the left hand side. In identity B this is not so much the case. However, the two identities are structurally similar: replacing multiplication and division signs in A by addition and subtraction, respectively, yields identity B. In spite of this shared structure, identity A is more visually salient than B. Awtry and Kirshner conclude that many errors in algebra are not the result of conceptual misunderstanding, but of an over-reliance on visual salience. The way Awtry and Kirshner perceive visual salience seems to be closely related to our perception of gestalt.

In line with Wenger (1987), who describes salient patterns and salient symbols, in this study we distinguish two different types of visual salience: pattern salience and local salience. *Pattern salience* (PS) concerns the recognition of patterns in expressions and equations, and as such is close to the ideas of Awtry and Kirschner described above. If a pattern is recognized by the student by means of a gestalt view, it may recall a standard procedure and invite its application. *Local salience* (LS) concerns the salience of visual attractors such as exponents, square root signs and fractions. Whether it is good or bad to resist the local visual salience depends on the situation. Using our extended definition of visual salience, developing a feeling for when to resist or succumb to both pattern and local visual salience is part of the acquisition of a gestalt view and thus of algebraic expertise. In short, a gestalt view includes both pattern salience, involving the recognition of visual patterns, and local salience, involving the attraction by local algebraic symbols. In both cases, a gestalt view is needed to decide whether to resist or succumb to the salience. A gestalt view, therefore, includes the learner's strategic *decision* of what to do next. This is graphically depicted in figure 2.

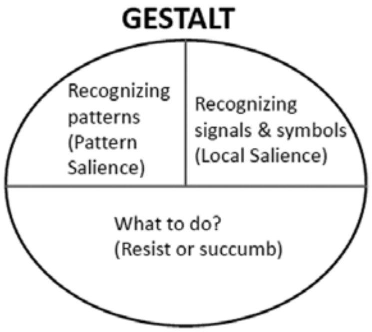


Fig. 2: Gestalt view: pattern salience, local salience and strategic decision

It should be noted that visual salience is not a matter of ‘yes’ or ‘no’: algebraic expressions may have different degrees of visually salience, that also depend on the context and on the knowledge of the student.

The *resistance to visual salience* refers to the ability to resist visually salient properties of expressions, and their implicit invitation to carry out specific operations. For example, students who perceive brackets may be tempted to expand the expression, whereas this does not necessarily bring them closer to the desired result. Another example is the sensibility to square root signs in an equation, that in the students’ eyes ‘beg to be squared’, even if this may complicate the equation. The opposite can be said for exponents on both hand sides of an equation: here taking roots can or can not be an efficient operation.

3 How might technology fit in?

Now how about the role for technology in the acquisition of algebraic expertise in the sense of both procedural skills and symbol sense, and with a focus on a gestalt view on, and the visual salience of, algebraic expressions? Educational use of ICT often consists of ‘drill-and-practice’ activities, and as such seems to focus on procedural skills rather than on conceptual understanding. However, research in the frame of instrumental and anthropological approaches shows that there is an interaction between the use of ICT tools and conceptual understanding (Artigue, 2002). This interaction is at the heart of instrumental genesis: the process of an artifact becoming an instrument. In this process both conceptual and technical knowledge play a role. To exploit ICT’s potential for the development of algebraic expertise, it is crucial that students can reconcile conventional pen-and-paper techniques and ICT techniques (Kieran & Drijvers, 2006). Important characteristics of ICT tools that can be used for addressing both procedural skills and conceptual understanding are options for the registration of the student’s solution process, and the possibility for the student to use different strategies through a stepwise approach. This enables the student to apply his or her own paper-and-pencil reasoning steps and strategies (Bokhove & Drijvers, 2010).

The opportunities that technology offers for the development of such algebraic expertise so far remain unexploited. Our goal, therefore, is to design and pilot digital activities that cater for the development of both procedural fluency and conceptual understanding. More specifically, we try to observe symbol sense behavior in digital activities. Do the concepts of symbol sense, gestalt view and pattern and local visual salience, described in a pre-digital era, help us in understanding what students do in a digital environment? This is the main topic of this chapter, In answering this question we will not focus on the characteristics of the digital tool (Bokhove, 2010; Bokhove & Drijvers, 2010); rather, we focus on the mathematical aspects.

4 Categories of items with symbol sense opportunities

To address the above issue, we first have to decide what we want to observe. We want to be able to see which strategic decisions students make while solving algebraic tasks in a digital environment. We want to know what salient characteristics, be they pattern salience or local salience, students resist or succumb to. This can only be done if the tasks offer symbol sense opportunities. For the task design, we used sources related to the transition from secondary to tertiary education, such as exit and entry examinations, remedial courses, text books and journals. Several suitable 'symbol sense type items' were identified and selected according to their focus on gestalt view and visual salience and supported by theoretical reflections from literature. The main criterion was that items would invite both procedural skills and symbol sense. This yielded a collection of thirty items, grouped into four categories, addressing both procedural skills and symbol sense, with an emphasis on the latter. We defined four categories of items, (1) on solving equations with common factors, (2) on covering up sub-expressions, (3) items asking for resisting visual salience in powers of sub-expressions, and (4) items that involve recognizing 'hidden' factors. Even if these categories may seem quite specific, they share the overall characteristic of an intertwinement between local and global, procedural and strategic focus.

Category 1: Solving equations with common factors

Items in this category are equations with a common factor on the left and right-hand side, such as:

Solve the equation:

$$(x^2 - 7x + 12) \cdot (8x - 11) = (x^2 - 7x + 12) \cdot (3x + 14)$$

A symbol sense approach involves recognizing the common pattern – in this case the common quadratic factor. This is considered as a sign of pattern salience, involving the pattern $AB = AC$. After recognizing the pattern, students have to decide whether or not to expand the brackets. The decision not to expand the brackets is seen as a sign of gestalt view and of resistance to the pattern salience of the pairs of brackets on both sides of the equation. After deciding not to expand, students could be tempted to just cancel out the quadratic terms on both hand sides of the equation, relying on the rule $AB = AC \Rightarrow B = C$ and thereby forgetting that $A = 0$ also yields solutions. This could be the result of a wrong rewrite rule applied to a recognized pattern. A non-symbol sense approach would involve expanding both sides of the equation, in this case yielding a third order equation that cannot be solved by the average student.

Category 2: Covering up sub-expressions

In this category, sub-expressions need to be considered as algebraic entities that can be covered up without caring for their content. A well-known example is:

Solve for v:
$$v \cdot \sqrt{u} = 1 + 2v \cdot \sqrt{1+u}$$

A symbol sense approach consists of noticing that the expressions under the square root signs are not important for the solution procedure (gestalt) and can be covered up. This requires a resistance to the local salient square root signs. In addition to this, a resistance is needed to the tendency to just isolate the v on the left hand side of the expression by dividing by the square root of u, which would leave v on the right hand side. Thus, resistance to pattern salience is required as well, and not doing so shows a limited gestalt view.

A non-symbol sense approach might focus on the visually attractive square roots and try to get rid of them by squaring both sides. This would be a strategic error, and does not bring the solution any closer. This equation is presented by Wenger (1987), who explained the issue as follows:

> If you can see your way past the morass of symbols and observe the equation #1 ($v \cdot \sqrt{u} = 1 + 2v \cdot \sqrt{1+u}$ which is to be solved for v) is linear in v, the problem is essentially solved: an equation of the form $av=b+cv$, has a solution of the form $v=b/(a-c)$, if $a \neq c$, no matter how complicated the expressions a, b and c may be. Yet students consistently have great difficulty with such problems. (p. 219)

Recognizing the salient pattern of a linear function $AV = B + CV$ and what to do with it is deemed a gestalt view, as defined in our conceptual framework. Gravemeijer (1990) elaborates on the same example and emphasizes the importance of recognizing global characteristics of functions and equations.

Category 3: Resisting visual salience in powers of sub-expressions

This category is about recognizing when to expand expressions and when not. It contains equations with sub-expressions that just beg to be expanded because they are raised to a power:

Solve the equation:
$$(x-3)^2 + 4 = 40$$

A symbol sense approach would include the recognition that after subtracting 4, both sides are squares, of $x - 3$ and $\sqrt{36}$, respectively. One should resist the temptation of expanding the left-hand side of the equation (resistance to pattern salience). Expanding the square to get rid of the brackets would be quite inefficient, and therefore rather is considered a non-symbol sense approach. Once the two squares of the pattern $A^2 = B^2$ are recognized, it is a sign of good gestalt view to succumb to the pattern salience by taking the square roots of both sides of the equation.

This item has several variants. For example, what if $(x - 3)$ is raised to the seventh power in the above example? The amount of work involved expanding this expression may stimulate students to look for alternative solutions.

Category 4: Recognizing ‘hidden’ factors

This category concerns the recognition of factors that are not immediately apparent (gestalt). An example is the following item adapted from Tempelaar (2007):

$$\text{Rewrite: } \frac{x^2 - x}{x^2 - 2x + 1}$$

A symbol sense approach would involve recognizing a common factor in both numerator and denominator and noticing that both numerator and denominator can be factored by $(x - 1)$. A pattern $\frac{A \cdot B}{A \cdot C}$ is then recognized. A further manifestation of what to do next, a gestalt view, facilitates further simplification and may lead to an equation resembling those of the first category. Not recognizing these factors results in complex rewriting. A non-symbol sense approach would come down to the manipulation of algebraic fractions without much result.

5 The design of a prototypical digital environment

The next step was to design a prototypical digital environment containing the items we defined. For this we carried out an inventory of digital tools for algebra and choose to use the Digital Mathematical Environment DME (Bokhove & Drijvers, 2010a). Key features of DME that led to its choice are that it enables students to use stepwise strategies and that it stores these stepwise solution processes. It also offers different levels of feedback, allows for item randomization and has proved to be stable.

For the design we used the DME’s authoring tool. Figure 3 shows some of its main features: the question text, the initial expression, the answer model, navigation, scoring and the possible use of randomized parameters. Figure 4 shows the implementation of an item from the first category. It is important to note that the algebraic steps are provided by the student, while the tool formats the steps, checks them algebraically and provides feedback.

An English version of the prototype can be found at http://www.fi.uu.nl/dwo/en/. For storage of the results, registration is required, but one can also enter as a guest user.

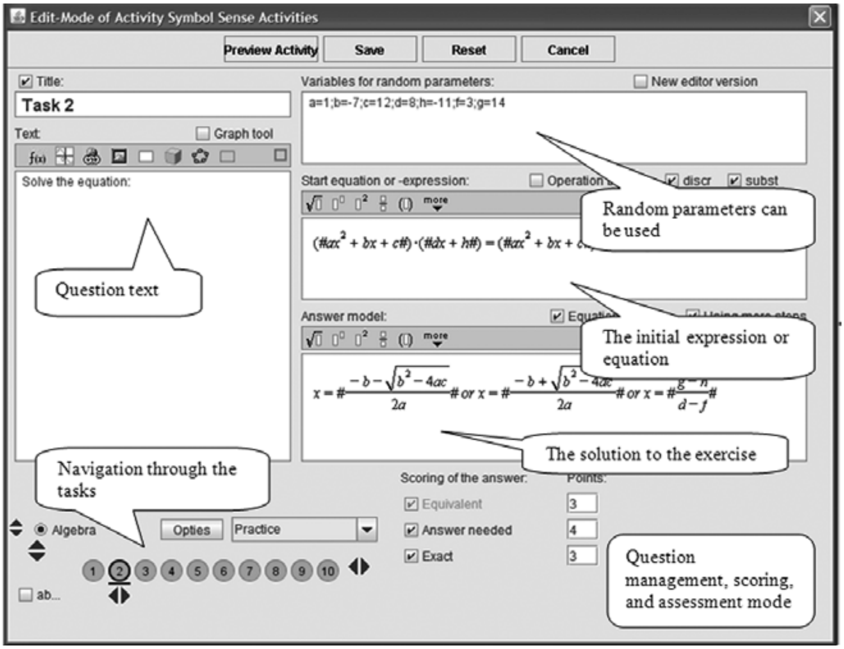


Fig. 3: Authoring an item on the equation
$(x^2 - 7x + 12) \,.\, (8x - 11) = (x^2 - 7x + 12) \,.\, (3x + 14)$

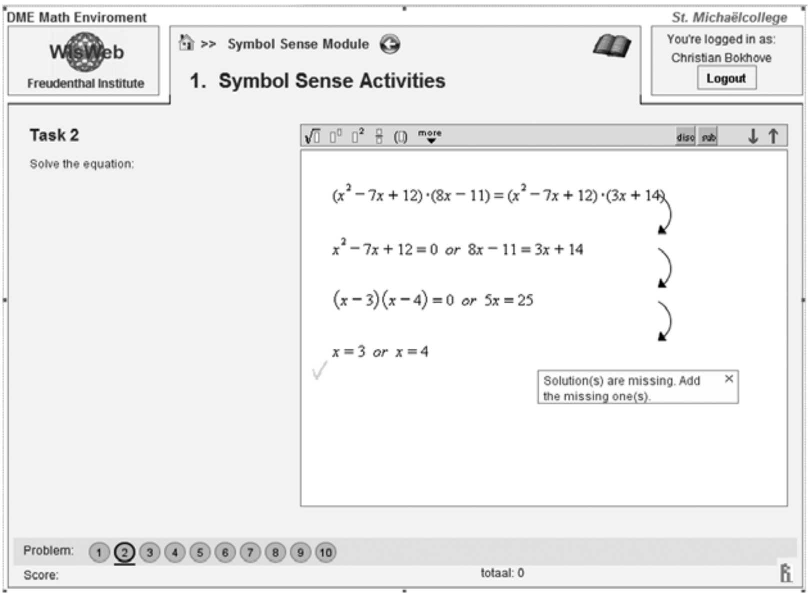


Fig. 4: Example of student steps and feedback provided by the tool

6 Piloting through one-to-one sessions

To find out whether the concepts of symbol sense, gestalt view and pattern and local visual salience help us in understanding what students do in a digital environment, five one-to-one sessions with pre-university grade 12 students (17 year olds) were held. The students all had C+ grades for mathematics.

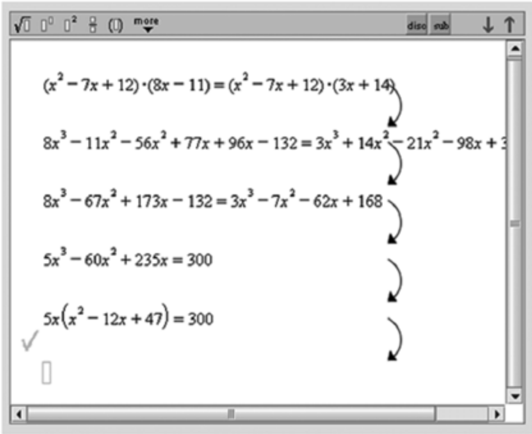
During the two-hour sessions, students worked through the digital activities. They were asked to think aloud while working. If a student was not able to complete (part of) a task, the observer asked what information would help in proceeding. On occasions where student used wrong strategies or made specific procedural choices, the observer asked the student what he or she was thinking. This informed possible feedback for a future revision of the prototype. After completing the session, the observer and the student went through the student's work and reflected on the solutions, discussing the student's arguments and alternative solution paths.

Data consists of audio and video registrations and computer screen recordings. Data analysis focused on the types of behaviors shown by students while working with the digital activities, and in particular on signs of (a lack of) symbol sense, and was carried out with software for qualitative data analysis. One first round of analysis concerned students' technical behavior when performing algebraic activities: factoring expressions, rewriting expressions, aggregating terms, expanding expressions and canceling terms. A second round of analysis concerned the identification of gestalt and/or visual salience features in the students' behaviors[1].

We now summarize the findings of the one-to-one sessions for each of the item categories. Per category one typical example of student behavior concerning gestalt and/or visual salience is given, as well as an overall description of the observed behaviors. We provide a rough time indication Δt in minutes per step, the technique used and comments on behavior related to gestalt view and visual salience.

1. Data is available through http://www.fi.uu.nl/~christianb

Student behavior on category 1 items



Δt	Technique	Interpretation
3	Expand	The student does not recognize the common factors and expands to remove the brackets
2	Aggregate common Terms	
3	Rewrite	
3	Factor	The student recognizes the common factor $5x$
		Student realizes he cannot solve a third degree equation, and gives up.

Fig. 5: Student solving an equation containing common factors

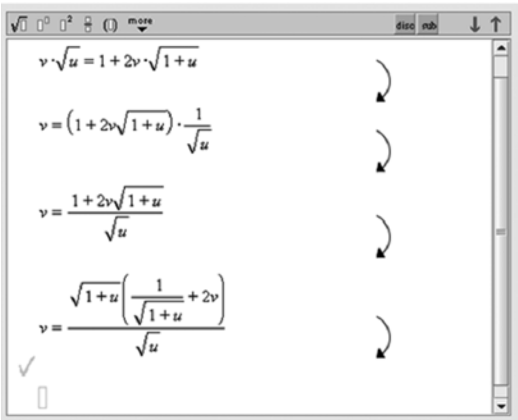
Figure 5 shows the work of one of the students, Martin. Martin did not recognize the common factors on the left- and right-hand side. In the first step he expanded the expressions on both sides, a strategy that he incorrectly described as “always works”. His inability to notice the common factor, and the pattern salience of the pairs of brackets, lead to his expanding strategy, a strategy which he used successfully – though not efficiently – in the previous task. In the second and third step Martin looked for terms that could be aggregated and rewrote the expression in the form ‘expression = number’. Next he tried to factor the left-hand side. Although he showed good rewriting skills, and was even able to factor the expression later on, he gave up eventually. Martin realized that he could not solve a third degree equation. His approach does not reflect a gestalt view on the initial equation.

In general, student behavior on this task and similar ones in this category showed that both too much routine and a lack of self-confidence play an important role in obstructing gestalt. For example, student Laura solved several equations correctly, but always worked towards the Quadratic Formula. She also solved one equation correctly with a symbol sense strategy, but when confronted with a similar equation with fractional terms, she was reluctant to solve it as she immediately stated she “was not skilled enough”. Only later did she recognize that, although the equation looked different, a similar technique could be used. Ideally a student would recognize the zero product theorem here. Another solution involved ‘just’ canceling out the common factors. As described in the category descriptions this indicates that on the one hand there is gestalt and resistance to pattern salience (“I’m not expanding both hand sides of the equation”).

On the other hand, however, students also succumb to an incorrect pattern salience, a buggy rule, of $AB = AC \Rightarrow B = C$ by just canceling out the common factors.

From this category we conclude that a gestalt view, and the observation of the salience of the common factor pattern in particular, is not evident for many students. Even skilled students show a lack of gestalt view on encountering this type of task in a digital environment.

Student behavior on category 2 items



Δt	Technique	Interpretation
0 3	Rewrite	The student isolates v by dividing by $\sqrt{u}$
4	Rewrite	
5	Factor	The student realizes $2v$ poses a problem and tries to lose this term by factoring
9		The student tries a next step, backtracks, tries again and finally gives up

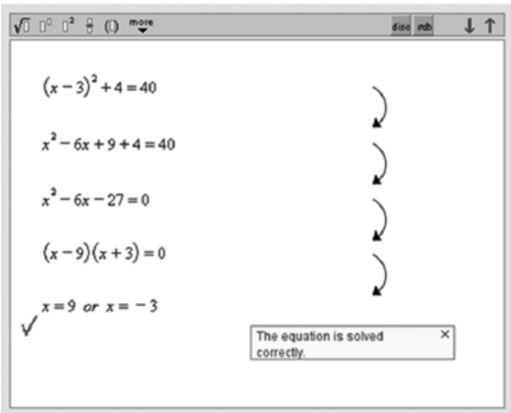
Fig. 6: Student behavior on the Wenger equation (Wenger, 1987)

Figure 6 shows Barbara's digital work on equation by Wenger (1987). Barbara was instantly alerted by the square root signs, knowing that squaring these would not bring a solution any closer. Thus, she resisted the local salience of the square root signs. She was triggered by the task to write the expression in the form $v = \ldots$ and first divided by the symbol in front of the *v*. This corresponds with Wenger's observed strategy: "Divide the equation by the coefficient of one of the occurrences of v in de given equation" (Wenger, 1987, p. 230). This can be seen as succumbing to pattern salience. Asking for an expression in the form $v =$ is directly transferred to the expression, and the quickest path to a solution in that form is dividing by $\sqrt{u}$. Barbara used one more step to rewrite the right-hand side as one fraction. She then took some time and circled numerous times round the term $2v$ exclaiming "I want to get rid of this term". She then started to rewrite the numerator, stating: "I want to simplify the numerator. I think this helps" and "I often do this to create a sort of hunch. That I look at the exercise in a different way, as to see what can be better or must be done". She used her procedural skill to rewrite terms hoping that this might provide insight into the correct solution path. After backtracking, she tried another approach, but again ended up with a term $2v$ on the right-hand side. She then gave up.

In general, the students started with similar steps as Barbara did, focusing only on the *v* on the left-hand side of the equation. Some gave up because of circularity: "… the process of symbolic manipulation which results in an obvious or tautological identity, which is uninformative and unproductive" (Arcavi, 1994, p. 29). Two of the students backtracked after unsuccessful attempts and seemed to have a better idea what to do, finally ending up completing the task correctly. This was facilitated by the fact that the tool provided feedback on the correct or incorrect nature of an answer. This can be seen in figure 4, where the system responds with the comment that solutions are missing. Other versions of this type of task, presented right after this one, but with different variables, were recognized by most students. Remarkably, the students with the higher marks for calculus saw some of these tasks as completely new ones. These students solved them correctly, but in a very inefficient way. Apparently, showing a high procedural skill mastery does not necessarily imply that a student sees the general in the particular.

From this category we conclude that in the digital environment students show the same specific behavior when covering up irrelevant sub-expressions as Wenger reported earlier: students show resistance to local salience but fall victim of pattern salience. The chosen actions by the student reveal a lack of gestalt view.

Student behavior on category 3 items



Δt	Technique	Interpretation
1	Expand	The student cannot resist the temptation of expanding the expression on the left hand side
1	Rewrite	
1	Factor	The student recognizes the expression can be factored
		The equation is solved correctly.

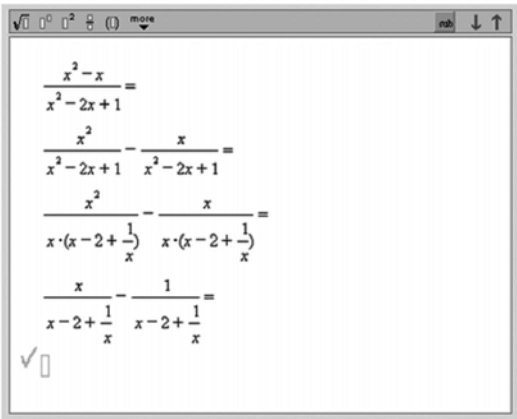
Fig. 7: Student not resisting visual salience

Figure 7 shows the work by Laura on a category 3 item. Laura did not recognize 36 as a square, and expanded the left-hand side. The diagonal juxtaposition, as described by Kirshner (1989), was too strong to withstand: the square must be eliminated and this was done by expanding the term (non-resistance to pattern salience). Laura preferred to use her standard procedure for quadratic equations: to first rewrite as a quadratic equation equaling zero. Then she factored the expression on the left-hand side. Laura stated that this was possible because of the “nice numbers”, which can be seen as a gestalt view. Otherwise she would have used the Quadratic Formula which “always works”. This process finally yielded the correct solutions. She could have reached this solution more efficiently if she had recognized 36 as a square, and then had noticed that both sides of the equations could be considered as squares, i.e. she would have observed the pattern $A^2 = B^2$. In this case her standard procedure obstructed any thoughts on alternate strategies.

In general, students did recognize both sides as squares. In contrast, the previous task involved the equation $x^2 - 6x + 9 = 36$, in fact the same equation with expanded left-hand side. It was remarkable that no student noticed these tasks were similar.

From this category we conclude that lack of gestalt view on the initial equation, and lack of resistance to pattern salience, obstructs students thinking about alternate strategies, as is the case in a pen-and-paper setting as well.

Student behavior on category 4 items



Δt	Technique	Interpretation
3	Rewrite as a sum of fractions.	The student does not recognize the common factor and applies knowledge of the sum of fractions
2	Factor the denominator	The student expects to be able to cancel out the factor x
1	Cancel	
1		The student reflects on the task and backtracks to the initial assignment

Fig. 8: Student not recognizing a hidden common factor

Figure 7 shows Barbara’s work on a category 4 task. Barbara instantly started rewriting, applying her knowledge of fractions. Instead of recognizing a common factor in both nominator and denominator - the pattern $\frac{A\cdot B}{A\cdot C}$ -, she started with what she did best: rewrite the expression as a sum of fractions - the pattern rule $\frac{A+B}{C}=\frac{A}{C}+\frac{B}{C}$. We see this as succumbing to a weak form of pattern salience and a lack of gestalt view. Next she factored the denominator and canceled out x in both terms. After several steps she noticed that the expression was becoming increasingly complex. The tick symbol denoted that the algebraic operations so far were correct. This, however, did not bring her to a more simplified expression. While carrying out these operations, Barbara became aware of the fact that $x-1$ played an important role in both numerator and denominator. She then backtracked, rewrote the initial expression with $x-1$ as factors and canceled them out.

In general, students showed trial-and-error behavior on this item. In some cases, this method seemed to provide the student with global insight into the expression.

From this category we conclude that these students have difficulties in recognizing common factor in nominator and denominator (lack of gestalt); however, the tool offert opportunities for a trial-and-error approach, which can provide insight into these factors.

7 Conclusion and discussion

The issue we wanted to address in this chapter is whether the notions of symbol sense, gestalt view and visual salience, described in a pre-digital era, help us in understanding what students do in a digital environment. The design process and the one-to-one pilot sessions suggest that these concepts remain extremely relevant when deploying digital activities. The observations show that students using a digital environment exhibit both symbol sense behaviors and behavior lacking it. The notions of gestalt view and visual salience are helpful in analyzing student work. Although students work in a digital environment instead of with paper and pencil, these results are in line with past findings in traditional pen-and-paper settings (Arcavi, 1994; Wenger, 1987).

While solving algebraic tasks in the digital environment, the students can use any strategy, and thus can show sensitivity towards gestalt and visual salience aspects, and further develop such sensibility. The tool seems to facilitate this development through its mathematical interface and feedback opportunities, which would be more difficult to offer in a paper-and-pencil environment.

The exemplary tasks also point out that observing symbol sense is not a straightforward affair. It often is quite hard to recognize whether students rely on standard algebraic procedures or are actually showing insight into the equation of expression, in line with the gestalt view or visual salience notion. Using standard procedures at least implies that a student recognizes the form of an expression. Recognizing patterns, and subsequently deciding what action to take, witnesses a gestalt view. However, (over-)reliance on standard procedures can also be seen as a matter of ‘succumbing’ to routine patterns: when a student encounters an expression with brackets he wants them to be eliminated. Extending the concept of visual salience to patterns provided by standard routines students already know could perhaps relieve the tension between the application of standard routines and succumbing to salient patterns.

Are we suggesting that digital tools are the panacea for algebra education? Things are not as simple as that. Crucial on the issue of how to design such activities is of course appropriate content, i.e. items inviting symbol sense, as proposed by the designed categorization. If the tasks are not appropriate, the intended learning will not happen. The potential is in the combination of task design and digital implementation. If the tasks do invite for adequate procedural techniques and appropriate theoretical thinking, a powerful environment is designed. The Task-Technique-Theory model (Chevallard, 1999; Kieran & Drijvers, 2006; Lagrange, 2000) may help designers to keep this aspect in mind.

The digital environment itself is a next crucial factor. High demands are put upon the digital tool in use. Students can get stuck by limitations of the technology. This being said, the potential added value of technology is promising: compared to carrying out the tasks with paper and pencil, we now have opportunities for different levels feedback and correction, for construction and exploration room for students, private and 'endless' practice and room for multiple stepwise strategies. With these conditions, the student is not restricted to strategies proposed by the digital tool itself, but can make his or her own correct or incorrect reasoning steps.

These conclusions suggest some guidelines for further research and development. Three issues for future development emerge: the sequencing of tasks, the extension of feedback, and scaffolding. First, future development should involve the design of *outlined sequences of tasks*, which appeal to symbol sense, and range from 'solvable with procedural skills' through 'inviting symbol sense' to 'only solvable with symbol sense insight'. Second, cues for developing gestalt view and the ability to deal with visual salience could be provided by relevant *feedback*. This issue asks for further elaboration. Feedback needs to be designed in more detail, concerning both the amount and the type of feedback (Hattie & Timperley, 2007). This also includes timing issues. As we saw students just starting a task without taking the time to actually think about it, it might be a good idea to include a cue for first scrutinizing the item carefully. When addressing feedback we can build on research by Nicaud (2004) and Sangwin (2008). Third, it might be worthwhile to build *scaffolding* into the sequence of activities, through initial activities that are structured and provide much feedback, that are then followed by items that gradually offer less structure and feedback. Support for this idea of formative scenarios (Bokhove, 2008) can be found in the notion of fading (Renkl, Atkinson, Maier, & Staley, 2002). It is in the line of these three issues that we plan to continue our research.

We do not pretend that the final word in the debate on procedural skills and symbol sense skills has been said. We do believe, however, that an optimal educational strategy is to focus at both simultaneously, and that technology may provide appropriate environments for this.

Chapter 4

Implementing feedback in a digital tool for symbol sense

Bohove, C. (2010). Implementing feedback in an digital tool for symbol sens. *International Journal for Technology in Mathematics Education*, *17*(3), 121-126.

Implementing Feedback in a Digital Tool for Symbol Sense

1 Introduction

During the last twenty years the relationship between procedural skills and conceptual understanding has been widely debated. This relationship played and still plays a central role in the ‘Math War’ (Becker and Jacob, 1998; Kieran and Drijvers, 2006) discussion. An important issue in this debate is how students should acquire algebraic expertise: by practicing algorithms, or by focusing on reasoning and strategic problem solving activities. One approach sees computational skills as an essential ingredient for understanding mathematical concepts (US Department of Education, 2007). The second approach has more focus on conceptual understanding (ibid.). Although there often is agreement on the fact that both procedural skills and conceptual understanding are important, there is disagreement on their relationship and the priorities among the two. Apart from this discussion, the last decades also saw an advent in the use of technology in mathematics education. The National Council of Teachers of Mathematics (2008) acknowledges the potential of ICT for learning in their position statement. In our research we combine these two issues: we want to use the potential of ICT to address algebraic skills, on both a procedural and a conceptual level. A pre-study concerned the choice and design of a digital tool for algebra tools (Bokhove and Drijvers, 2010a). Figure 1 sums up a selection of these criteria. The next cycle involved one-to-one sessions with students. These sessions had two goals. One was to see whether digital activities could invite symbol sense behaviour; we report on this in a different article (Bokhove and Drijvers, 2010b). A second goal was to elicit the design of feedback for a second revision of the tool. This chapter reports on this second goal.

- Assesses both basic skills and symbol sense;
- Provides an open environment and feedback to facilitate formative assessment;
- Stores both answers and the solution process of the student;
- Steps;
- Freedom to choose own strategy;
- Authoring tool for own questions;
- Intuitive interface ('use to learn' vs. 'learn to use');
- Close to paper-and-pencil notation.

Fig. 1: Selected criteria for algebra tools (Bokhove & Drijvers, 2010)

2 Conceptual Framework

To place our focus on feedback in a more general framework, we briefly describe the three topics that are involved in our study: algebra didactics (symbol sense), theories on tool use (instrumental genesis) and assessment (feedback and formative assessment).

Algebra

According to Arcavi (1994) algebraic skills have a basic skill component and a symbol sense component. Both basic skills and symbol sense should be addressed in education, as they are closely related: understanding of concepts makes basic skills understandable, and basic skills can reinforce conceptual understanding. Here, symbol sense lies in the realm of conceptual understanding: an intuitive feel for when to call on symbols in the process of solving a problem, and conversely, when to abandon a symbolic treatment for better tools. Arcavi (1994) describes several 'behaviours' of symbol sense. Figure 2 shows the dimension Drijvers, Goddijn and Kindt (2010) see: basic skills involve procedural work with a local focus and emphasis on algebraic calculation, while symbol sense involves strategic work with a global focus and emphasis on algebraic reasoning.

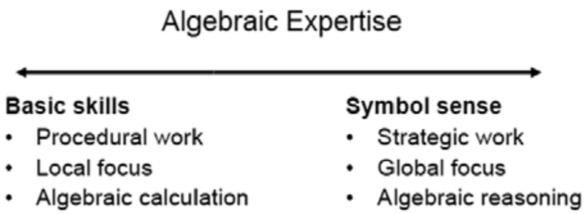


Fig. 2: Algebraic expertise as a dimension: from basic skills to symbol sense

Bokhove and Drijvers (2010b) studied four categories of items with symbol sense opportunities: solving equations with common factors, covering up sub-expressions, resisting visual salience in powers of sub-expressions and recognising common factors. The categories are sometimes subtly interlinked. For example the category on recognizing common factors involves common factors that are not clearly apparent in an equation. The category on equations with common factors, on the other hand, has clearly visible common factors on both hand sides of the equation. One could even say that after recognising common factors (category four) one is left with a category one equation. However, the type of symbol sense involved, is different. One thing all these categories have in common is a combination of gestalt view or resistance to visual salience. A *gestalt view on algebraic expressions* involves the ability to consider an algebraic expression as a whole, to recognize its global characteristics, to 'read through' algebraic expressions and equations, and to foresee the effects of a manipulation strategy. *Resistance to visual salience* refers to the ability to resist visually salient properties of expressions, and their implicit invitation to carry out specific operations. The following task provides an example of the first category:

Category 1: Solving equations with common factors
This category can be characterized by common factors on the left and right hand side of the equation. Example:

Solve the equation:
$(x^2 - 7x + 12) \cdot (8x - 11) = (x^2 - 7x + 12) \cdot (3x + 14)$

Fig. 3: Example task from our digital tool

Later on we will return to this example in our case study.

Tool use

Tool use is an integrated part of human behaviour. Vygotsky (1978) sees a tool as a mediator, a "new intermediary element between the object and the psychic operation, directed at it". Verillon and Rabardel (1995) distinguish artefact and instrument. The *artefact* is just the tool. The *instrument* denotes the relationship between a person and the artefact. Only when this relationship is established, one can call it a "user agent". The mental processes that come with this are called *schemes*. In short: instrument = artefact + instrumentation scheme. Trouche (2003) distinguishes *instrumentation* (how the tool shapes the tool-use) and *instrumentalisation* (the way the user shapes a tool). *Instrumental genesis* is the process of an artefact becoming an instrument. In this process both conceptual and technical knowledge play a role ("use to learn" and "learn to use"). To overcome the contrast between pen-and-paper and ICT based learning, an ICT environment has to correspond with traditional techniques (Kieran and Drijvers, 2006).

Assessment

Black and Wiliam (2004) distinguish three functions for assessment:

- supporting learning (formative)
- certifying the achievements or potential of individuals (summative)
- evaluating the quality of educational programs or institutions (evaluative)

Summative assessment is also characterised as assessment *of* learning and is contrasted with formative assessment, which is assessment *for* learning. Black and Wiliam (1998) define assessment as being 'formative' only when the feedback from learning activities is actually used to modify teaching to meet the learner's needs. From this it is clear that feedback plays a pivotal role in the process of formative assessment.

Feedback

In the learning process adapting instruction to meet students learning needs showed substantial benefits, for example in studies by Carpenter, Fennema, Peterson, Chiang and Loef

(1989) and Black, Harrison, Lee, Marshall and Wiliam (2003). As the role of feedback had to be taken into account, drill-and-practice use of the computer made formative assessment difficult. An interesting question is whether the use of “more intelligent” new technology makes a difference in this respect. Bangert-Drowns, Kulik, Kulik and Morgan (1991) found that not being able to see the answer before trying a question is better. Also giving details of the right answer, instead of just wrong or right, seemed more effective, as other research has also confirmed (Dempster, 1991, 1992; Elshout-Mohr, 1994).

Several studies including Nyquist (2003) seem to show more effectiveness in assessment when using feedback. The stronger the feedback the larger the effect seems to be (Elawar and Corno, 1985). Reviews conducted by Natriello (1987), Crooks (1988), Bangert-Drowns et al. (1991) and Black and Wiliam (1998) showed that that not all kinds of feedback to students about their work are equally effective. It is therefore sensible to distinguish feedback types. There are several models that distinguish feedback types. Nyquist (2003) reviewed 185 studies in higher education, developing a typology of different kinds of formative assessment:

- *Weaker feedback only:* students are given only the knowledge of their own score or grade, often described as “knowledge of results”
- *Feedback only:* students are given their own score or grade, together with either clear goals to work towards or feedback on the correct answers to the questions they attempt, often described as “knowledge of correct results”.
- *Weak formative assessment:* students are given information about the correct results, together with some explanation.
- *Moderate formative assessment:* students are given information about the correct results, some explanation, and some specific suggestions for improvement.
- *Strong formative assessment:* students are given information about correct results, some explanation, and specific activities to undertake in order to improve.

This distinction emphasizes the important role feedback plays in formative assessment. Hattie and Timperley (2007) did a meta-review of the effectiveness of different types of feedback. The feedback effects of cues and corrective feedback are deemed best. Seeking feedback is governed by a cost/benefit ratio. In general, feedback is psychologically reassuring, and people like to obtain feedback about their performance, even if it has no impact on their performance. The model provided by Hattie and Timperley (ibid.) distinguishes three questions that effective feedback answers:

Where am I going? (the goals)	*FeedUp*
How am I going?	*FeedBack*
Where to next?	*Feed Forward*

Each feedback question works at four levels (focus of the feedback):

- Task level: how well tasks are understood/performed (FT)
- Process level: the main process needed to understand/perform tasks (FP)
- Self-regulation level: self-monitoring, directing and regulating of actions (FR)
- Self level: personal evaluations and affect (usually positive) about the learner (FS)

. Hattie and Timperley (2007) also provide some statements on the effectiveness of (combinations of) feedback types, including that FS feedback is least effective, simple FT feedback is more effective than complex FT feedback , FT and FS do not mix well ("Well done, that is correct" is worse than "Correct" only), FT is more powerful when it's about faulty interpretations, not lack of information. Furthermore they state that we should be attentive to the varying importance of the feedback information during study of the task.

These principles are the basis of the choice of our tool and will also be the basis of our student-inquired feedback design for an algebra tool.

3 Research questions

The study aims at elaborating possible and desired feedback for our digital tool. The questions we set out to answer in this chapter are:
(1) Can the feedback design of a digital tool be improved through student inquiry?
(2) What methodology can be used to elaborate feedback desired by students?

4 Methodology

The methods include one-to-one sessions with students, post-analysis of feedback occurrences and subsequent revision of the prototype. Apart from validating whether the proposed activities in the prototypical digital environment indeed invite symbol sense (Bokhove and Drijvers, 2010b), five one-to-one sessions with students were also held to explicate what feedback was desired on what moment. As the tool was also intended for home-use we chose to focus on individual usage of the tool, and thus one-to-one sessions. Four pre-university grade 12 students were in the so-called nature/health stream. A fifth student took the advanced math for science course. The students all had C+ grades for mathematics. An additional sixth one-to-one session was held with a mathematics student, to compare the data with expert behaviour. The sessions were recorded with audio and camera, and the content of the computer screens by screen recording software (Camtasia).

During the sessions, the students worked through the digital activities. They were asked to think aloud while working. If a student was not able to do (part of) an item, the observer asked what information would help in proceeding. On occasions where student used wrong strategies or made specific procedural choices, the observer asked the student what he or she was thinking. This informed possible feedback for a future revision of the prototype. After the first part of each session, the observer and the student went through the student´s work and reflected on the solutions, discussing the student´s arguments and alternative solution paths, as well as possible (feedback) improvements of the prototype.

Analysis of screen casts took place with software for qualitative data analysis (ATLAS TI), and involved aggregating data on desired feedback. This meant that for every task we constructed an overview of all the feedback that was desired by the students. Next, this feedback was implemented in the revised prototype. In the next paragraph we give an example of a feedback-design-in-progress for the example from figure 2.

5 Results

The results consist of (i) the described student-driven methodology for elaborating desired feedback for digital tools; (ii) a feedback design for our initial prototypical intervention; (iii) a revised intervention.

These results are exemplified in a case study example.

Input	Feedback	Description	Feedback
$x = 3$ or $x = 4$ or $x = 5$	Correct	Only if it's exactly this input, the students scores maximum points. Note that the answer is not accompanied by FS feedback, as FS feedback influences FT feedback in a negative way (Hattie & Timperley, 2007)	FT
$(8x - 11) = (3x + 14)$	You probably divided by $(x^2 - 7x + 12) = 0$. You lost some solutions for $x^2 - 7x + 12 = 0$.	The student gets part of the score for one of the solutions, but is also served feedback on losing some solutions.	FT
$(8x - 11) = (3x + 14)$	Plot graphs of the functions on both sides of the equations and determine how many points of the intersection there are.	By providing students with a plotting tool, students are invited to find out for themselves whether the answer they have given is correct.	FR
$5x^3 - 60x^2 + 235x - 300 = 0$	You probably don't know how to solve this third order equation. If not, then look at the initial equation. Do you see corresponding factors on the left and right hand side?	The student has expanded the expression and rewritten it to the form expression = 0. The students involved in the target group do not know how to solve a third order equation, and is therefore advised to scrutinize the initial equation some more.	FP
$(x^2 - 7x + 12) = 0$ or $(8x - 11) = (3x + 14)$	You recognized similar factors on both sides.	The student recognizes the common factors on both hand sides of the equation.	FT

Fig. 4: Step one. The table shows a part of the feedback design

6 Case study example

In this paragraph we provide an example of the feedback elaboration process through a feedback design for one of the tasks in our prototypical tool. The task we will present is the task from figure 2.

Solve the equation

$$(x^2 - 7x + 12) \cdot (8x - 11) = (x^2 - 7x + 12) \cdot (3x + 14)$$

Subsequently the steps that are involved are:
Step 1: we *distilled desired feedback* occurrences from the screencasts and made a feedback design.

Step 2: we *implemented* the design in our second revision. See figure 5.

Step 3: we *tested* our implementation. See figure 6.

Step 1: Feedback from one-to-one sessions

In this step we analyzed the screencast recordings of several student sessions. We distinguished remarks on desired feedback in all of the sessions, other welcome additions for the second revision of the tool, and other technical remarks. Grouping similar feedback resulted in a feedback design. Figure 4 shows a selection of the design for the case study example. The first column is the input from the student, the second column denotes the feedback the students want and thus we desire in our tool, the third column provides some explanation on expected student behaviour. The last column shows the feedback type (Hattie and Timperley, 2007).

Step 2: Implementation of the feedback design

Figure 5 shows the feedback design being implemented in our second revision. The figure shows part of the authoring environment that is used 'behind the scenes'. We can recognize several elements from figure 4. At the top of the figure tabs denote the different feedback instances one exercise holds. In this case instance four of in total eight feedback instances is active and highlighted in white. Every feedback instance consists of an *answer model*. Every step in the solution process of the student is evaluated according to the answer models that are in the feedback implementation. Note that the answer model is *randomized*. This means that *a* through *f* are random integers (within boundaries that are defined elsewhere). In this case one can recognize the Quadratic Formula in the answer model, as applied to the quadratic term in the initial equation. Of course, our case study example with the quadratic part yields two solutions, the linear term of the equation yields one solution.

Under this box the authoring environment provides the opportunity to apply rules and limitations: "Equivalent" is checked. This denotes that the answer given by the student should be equivalent to the answer model. “Vorm” is checked. This option is linked to the box on the bottom right and demands that the student answer is in a certain algebraic form. As the digital tool has an emphasis on symbol sense we desire that a student recognizes similar terms. The “Vorm” box therefore contains the form after recognizing these similar terms. The four radio buttons signify what symbol is appropriate: a green “goed” (correct) symbol, an orange “half” correct symbol or a red “fout” (incorrect) symbol. “Door” can be used for answers that the tool deems not entirely correct but require the possibility to ‘follow-through’ on the answer. This opposed to the situation where the solution process suddenly stops. If all these conditions apply, the feedback in the box on the bottom left is shown and the score provided is given.

Step 3: Testing the revision

Figure 6 shows part of what the feedback instance from figure 5 looks like when giving in the corresponding step. In this case the student has recognized common factors on both sides of the equation. The task is embedded in a full-fledged equation editor, and allows the student to make algebraic steps.

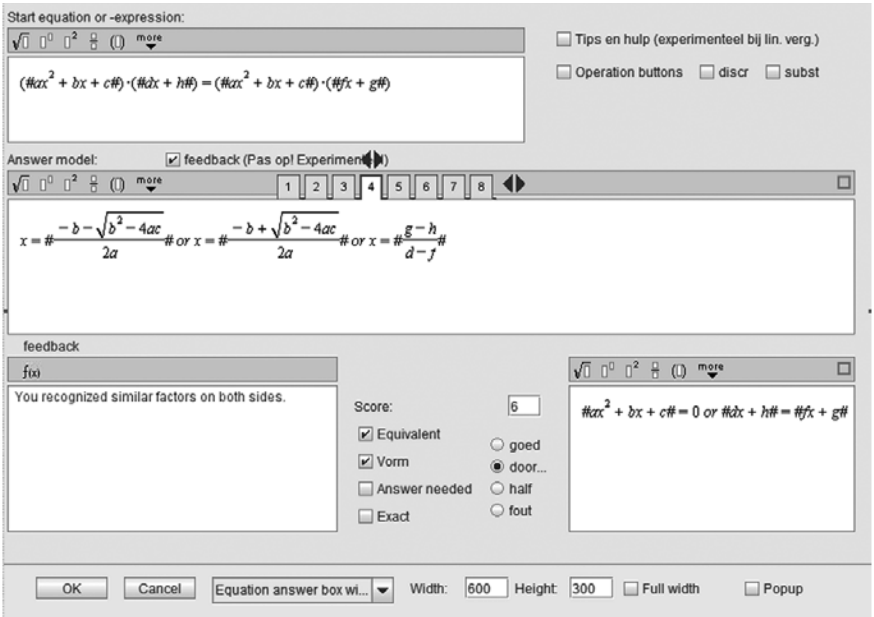


Fig. 5: Step two: authoring custom feedback for one of the items

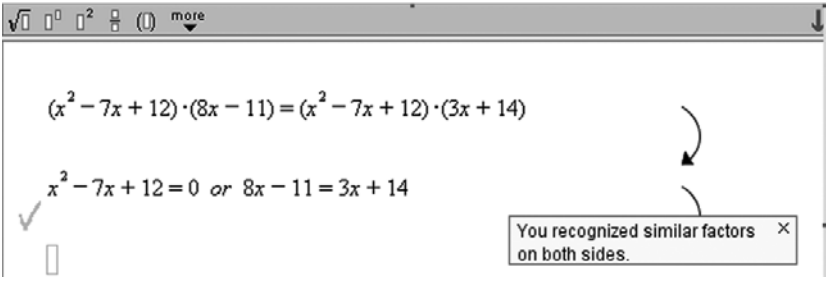


Fig. 6: Step three: custom feedback in action.

7 Conclusion and discussion

Our first question was whether the feedback design of a digital tool can be improved through student inquiry. We conclude that asking students when to use what feedback can improve a digital tool.

The second question was what methodology can be used to elaborate feedback desired by students. We have found our methodology fruitful:
(1) Deploy the tool and ask students to ‘think aloud’;
(2) Tabulate all the responses;
(3) Make a feedback design;
(4) Implement the feedback design in the tool;
(5) Test the tool with help of the feedback design, and make another iteration.

This provides a revision of the tool. It is possible to greatly improve a tool this way. In our one-to-one sessions and explorations on the role of feedback we would like to bring up three points of discussion.

Firstly, the role of scores. Scoring seems to have a negative impact (Butler and Nisan, 1986). Butler (1988) even concluded that the effects of diagnostic remarks completely disappeared when grades were added. So here we have a dilemma: do we use grades/marks and feedback together, hoping motivation will overcome the disadvantages or not.

Secondly, it is very difficult to anticipate all the mistakes students can make. We in no way would claim that this methodology and this tool caters for all common misconceptions. The methodology described does, however, add student input to the domain of feedback design.

Thirdly, technology is always ‘on the move’. Along with feedback, technological developments influence the improvement of our tool. Whether these are actually implemented in our tool will depend on the added value. In the author’s opinion tools for education should not focus on new features or technologies but on its relevance in a classroom setting. Both education and design can profit from each other.

Chapter 5

Effects of feedback in an online algebra intervention

Bokhove, C., & Drijvers, P. (2011). *Effects of feedback in an online algebra intervention*. Manuscript submitted for publication.

Effects of feedback in an online algebra intervention

1 Introduction

During the last decade the assumed dichotomy between procedural skills and conceptual understanding has been widely debated. It has been a focal point in the so-called 'Math War' discussion (Schoenfeld, 2004). The debate also influenced the type of expertise that algebra education should target: should students focus on practicing algorithms or on reasoning and problem solving strategies? One approach stresses the fact that computational skills are an essential ingredient for understanding mathematical concepts (US Department of Education, 2007). Another approach starts off with more focus on conceptual understanding (US Department of Education, 2007). Although most experts seem to agree that essentially both are needed, there is no clear agreement on the relationships and priorities among the two. Apart from this debate, recent years also saw an evolution in the use of technology in mathematics education. The National Council of Teachers of Mathematics (2008) formulated the potential of Information and Communication Technology (ICT) for learning in their position statement. The present research combines the aforementioned elements: we want to use the potential of ICT to address algebraic skills, on both a procedural and a conceptual level, and to design and test an intervention exploiting just that potential. In particular, we want to use the 'power of feedback' (Hattie & Timperley, 2007) in these computer tools, as it allows students to work at home independently. The use of feedback, however, comes with its own challenges: how much feedback is needed, when is it needed and when should we avoid feedback? To address these challenges, in this paper we try to identify crucial feedback design principles guiding such an intervention, and investigate their effects on learning. As the nature of feedback depends on the research context and the way it is implemented in an intervention, it is wise to first sketch a broader framework for feedback. Therefore we first address the research context of algebraic expertise, and then describe the notion of feedback as an essential part of formative assessment.

2 Research context: algebraic expertise

This research on the effects of feedback in an online algebra intervention takes place within the context of mathematics, targeting the design of an intervention that caters for both the acquisition of procedural skills and conceptual understanding. The distinction between procedural skills and conceptual understanding is a highly researched field of interest. The book Adding it up (Kilpatrick, Swafford, & Findell, 2001) synthesizes the research on this issue in the concept of mathematical proficiency, which comprises five strands: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning and productive disposition. Here, conceptual understanding is defined as “the comprehension of mathematical concepts, operations, and relations” (p. 116), and procedural fluency as the “skill in carrying out procedures flexibly, accurately, efficiently, and appropriately” (ibid.). Furthermore, “the five strands are interwoven and interdependent in the development of proficiency in mathematics” (ibid.). Arcavi (1994) introduced the notion of symbol sense, which includes “an intuitive feel for when to call on symbols in the process of solving a problem, and conversely, when to abandon a symbolic treatment for better tools” (p. 25). Arcavi describes eight behaviors in which symbol sense manifests itself. These behaviors show the intertwinement between procedural skills and conceptual understanding as complementary aspects of algebraic expertise. Both procedural skills and symbol sense need to be addressed in algebra education, as they are intimately related: understanding of concepts makes basic skills understandable, and basic skills can reinforce conceptual understanding (Arcavi, 2005). In line with this work, Drijvers, Goddijn and Kindt (2010) see algebraic expertise as a dimension ranging from basic skills to symbol sense (see Figure 1). Basic skills involve procedural work with a local focus and emphasis on algebraic calculation, while symbol sense involves strategic work with a global focus and emphasis on algebraic reasoning.

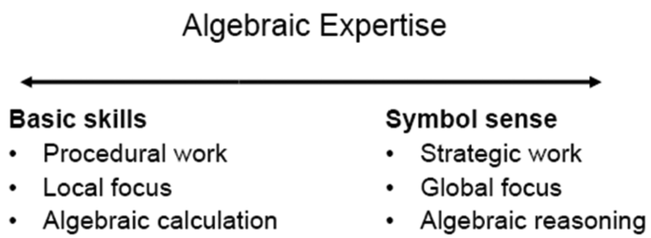


Fig.1: Algebraic expertise as a dimension (Drijvers, Goddijn & Kindt, 2010)

3 Formative assessment and feedback

Black and Wiliam (2004) distinguish three functions for assessment: supporting learning (formative), certifying the achievements or potential of individuals (summative), and evaluating the quality of educational programs or institutions (evaluative). Summative assessment is also characterized as assessment of learning and is contrasted with formative assessment, which is assessment for learning. Black and Wiliam (1998) define assessment as being 'formative' only when the feedback from learning activities is actually used to modify teaching to meet the learner's needs. From this it is clear that feedback plays a pivotal role in the process of formative assessment. According to an overview by Mason and Bruning (2001) determining factors for effectiveness of feedback are: elaboration, student achievement levels, depth of understanding, attitude toward feedback, learner control, response certitude, and timing. To give an example, Morrisson et al. (1995) found that delayed and knowledge-of-correct-response feedback may be more beneficial than answer-until-correct or no-feedback for lower level learning, while Clariana (1990) concluded that answer-until-correct feedback may be more effective for higher order learning than for lower level processing. While lower ability students may benefit from more immediate, specific forms of feedback, higher ability students may gain more knowledge from feedback that allows for active processing by the student. A review by Vasilyeva et al. (2007) also provides recommendations for the effective use of feedback in web-based application. In general, keeping in mind this body of knowledge, we want to address the general question:

Which design principles may guide feedback design in an online algebra intervention and what are the effects of these design choices?

In answering this question, we focus on three promising concepts concerning feedback. The first is the concept of timing and fading, which states that the amount and timing of feedback should differ according to the moment and context. Second, the concept of a crisis, in which no help is offered in the form of feedback whilst experiencing a conceptual crisis, after which feedback is presented to overcome that crisis. Third, feedback variation: providing more elaborate, ‘intelligent’ and varied feedback. These three concepts lead to three sub-questions:

(1) Timing and fading: are there indications that formative scenarios with fading feedback improve the acquisition of algebraic expertise?

(2) Crises: do crises in algebraic task sequences improve the acquisition of algebraic expertise?

(3) Variation: does variation in feedback influence scores and student behavior?

4 Method

Type of study

This complete study globally followed a design research approach with four phases. The preliminary research phase concerned the choice and design of a digital tool for algebra (Bokhove & Drijvers, 2010a). The first intervention cycle focused on whether the prototypical digital activities would invite symbol sense behavior through one-to-one sessions (Bokhove & Drijvers, 2010b). The second cycle consisted of a small scale field experiment in one school. The third and final cycle involved a large-scale classroom experiment. This research set-up shows a progress from small-scale to large-scale in 'layers of formative evaluation' (Tessmer, 1993), with an accompanying shift from more qualitative to more quantitative analyses. The last cycle, aiming at the intervention effects, is the focus of this article, although we refer to Bokhove & Drijvers (in press) for the overall effects of the intervention.

Participants

The participants were 324 12th grade 17-18 year old students from fifteen classes from nine Dutch secondary schools, involving eleven mathematics teachers. The schools were spread across the country and showed a variation in school size and pedagogical and religious backgrounds. The participating classes consisted of pre-university level ‘wiskunde B’ students (comparable to grade 12 in Anglo-Saxon countries). 43% of the participants were female and 57% were male. The participating schools reacted to an open invitation in several bulletins for mathematics education. Schools received an exemplary course planning and some hints on using the intervention. They were however free to adapt the intervention to their own requirements. Schools deployed the intervention in the last three months of 2010, just before preparations for the final national exams would start. Teachers received mailings on a regular basis, and could visit the project website www.algebra-metinzicht.nl. In addition, schools were randomly allocated to two feedback conditions c1 and c2. Data collection included results from a pre- and post-test, and the scores and log files of the digital activities. The log files record information on students' individual item scores, feedback, answers, and number of attempts.

Content of the intervention

In the design of this study's intervention, we want to address both ends of the dimension. To do so, we contend that an online algebra intervention needs to offer symbol sense opportunities. For the task design, we used sources related to the transition from secondary to tertiary education, such as exit and entry examinations, text books, journals and remedial courses. Several suitable 'symbol sense type items' were identified and selected. The main criterion was that items would address both procedural skills and symbol sense. This yielded a collection of items, grouped into four categories: (1) on solving equations with common factors, (2) on covering up sub-expressions, (3) items asking for resisting visual salience in powers of sub-expressions, and (4) items that involve recognizing 'hidden' factors. Even if these categories may seem quite specific, they share the overall characteristic of an intertwinement between local and global, procedural and strategic focus.

Category 1: Solving equations with common factors.

Items in this category are equations with a common factor on the left and right-hand side, such as:

Solve the equation:

$$(x^2 - 7x + 12) \cdot (8x - 11) = (x^2 - 7x + 12) \cdot (3x + 14)$$

Category 2: Covering up sub-expression.

In this category, sub-expressions need to be considered as algebraic entities that can be covered up without caring for their content. A well-known example from Wenger (1987) is:

Solve for v:

$$v \cdot \sqrt{u} = 1 + 2v \cdot \sqrt{1 + u}$$

Category 3: Resisting visual salience in powers of sub-expressions.
This category is about recognizing when to expand expressions and when not. It contains equations with sub-expressions that just beg to be expanded because they are raised to a power:

Solve the equation:

$$(x - 3)^2 + 4 = 40$$

Category 4: Recognizing 'hidden' factors.
This category concerns the recognition of factors that are not immediately apparent (gestalt). An example is the following item adapted from Tempelaar (2007):

Rewrite:

$$\frac{x^2 - x}{x^2 - 2x + 1}$$

The intervention consists of tasks taken from these four categories.

Structure of the intervention

The intervention called 'Algebra met Inzicht' [Algebra with Insight] was designed in the Digital Mathematical Environment. The intervention is depicted graphically in Figure 2 and consists of a pen-and-paper pre-test, four digital modules, called d1-d4, each covering one of the four task categories described above (Bokhove & Drijvers, 2010b), digital module d5 as a diagnostic tool, digital module d6 as a final digital test and, finally, a pen-and-paper post-test.The time needed to complete the module was estimated at six hours work in total, excluding pre- and post-tests. Table 1 shows whether an intervention element was presented digitally or with pen-and-paper, the number of items and Cronbach's alpha. The alpha values indicate that all parts of the intervention can be considered reliable (Garson, 2011).

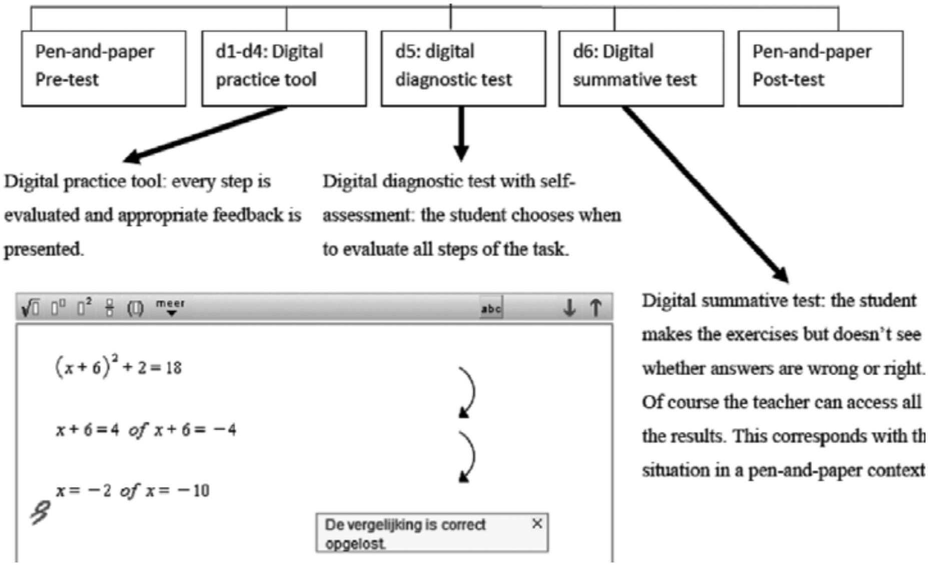


Fig. 2: Outline of the formative scenario underpinning the intervention

Table 1: Reliability of the parts of the intervention

Part	**Pre**	**d1**	**d2**	**d3**	**d4**	**d5**	**d6**	**Post**
Items N	8	12	6	11	16	23	23	10
Alpha	.724	.687	.917	.909	.880	.843	.828	.709
Digital (d)/pen-and-paper (p)	p	d	d	d	d	d	d	p

5 Three feedback design principles and their elaborations

We now describe the three main feedback design principles, that correspond to the three sub-questions phrased above, in more detail.

Feedback design principle 1: timing and fading with formative scenarios

The first feedback design principle, that corresponds to the first research sub-question, concerns timing and fading with formative scenarios. With formative scenarios we mean a structured collection of learning activities, starting off with a lot of feedback to aid learning, but decreasing the amount towards the end, to facilitate transfer. It is based on the concept of *fading* (Renkl, Atkinson, & Große, 2004). Fading refers to the idea that one starts off with completely elaborated tasks and moves to less elaborated tasks by removing intermediate steps. According to Jones and Fleischman, "faded examples cause effective learning by forcing the student to encounter and overcome an impasse" (Jones & Fleischman, 2001, p. 5). An impasse occurs when a learner realizes a lack of complete understanding of a piece of knowledge. In the original research by VanLehn et al. impasses are part of the CASCADE model (VanLehn, Jones, & Chi, 1992) of cognitive skill acquisition. Based on the CASCADE model, Jones and Fleischman (2001) assume that when learners study complete examples they often superficially read over the examples without experiencing impasses that would trigger self-explanations. It signifies the situation in which a rule for solving a task is missing. Formative scenarios (Bokhove, 2008) are a variation of this concept, starting off with much feedback, and providing a gradually decreasing level of feedback.

In the intervention we developed in the frame of this study, at the start feedback is provided for all intermediate steps, and in an additional condition also through options the student can use to get hints and worked examples. During the intervention, the amount of feedback decreases towards no feedback in the step level and only overall feedback after finishing the task. The principle applies to the intervention as a whole. More specifically, the intervention after a pretest begins with a practice tool. The tool provides feedback on every step. The next step concerns self-assessment and diagnostics: the student performs the steps without any feedback and chooses when to check his or her solution. Feedback is then given for the whole exercise. Finally, students get a final exam, with no means to see how they performed. Just as is the case with a paper test, the teacher will be able to check and grade the exam (in this case automatically) and give students feedback on their performance. After all, a student needs to be able to accomplish tasks independently, without the help of a computer. This approach also corresponds with ideas on graduated help, scaffolding and 'the power of feedback' (Hattie & Timperley, 2007).

Feedback design principle 2: crises

The second feedback design principle, that corresponds to the second research sub-question, concerns evoking crises. Learning often takes place when we are unsuccessful in performing a task or when things go wrong. One underlying principle for the design of our intervention is the concept of a *crisis*. With a crisis we refer to a principle that the poet John Keats so eloquently described at the beginning of the nineteenth century:

> Don't be discouraged by a failure. It can be a positive experience. Failure is, in a sense, the highway to success, inasmuch as every discovery of what is false leads us to seek earnestly after what is true, and every fresh experience points out some form of error which we shall afterwards carefully avoid.

Although crises can be described at a societal level, for example when a whole scientific discipline is thrown into a state of crisis causing a paradigm shift (Kuhn, 1996), here we refer to the individual level. We pose that in a sense the individual has to deal with his or her own paradigm shift. When confronted by a conceptual crisis critical thinking about problems and evaluating conflicting contexts, dialectical reasoning could help to overcome the crisis. The principle of a crisis also resembles other concepts, such as cognitive conflicts (Tall, 1977) , Van Hiele's 'crises of learning' (Van Hiele, 1985) and 'productive failure' (Kapur, 2008). It also ties in with the impasses that we described in the section on formative scenarios.

In the present intervention, we used this concept at the level of a series of items, in the following way. As soon as we expect students to rely on standard algorithms, this is challenged by an item that evokes a conceptual crisis: a task that is hard or impossible to accomplish with the algorithms that worked so well earlier. We illustrate this with an example from one-to-one sessions (Bokhove & Drijvers, 2010b).

Figure 3 shows a student elaboration of an exemplary task solving equations with common factors. This task followed several tasks with only linear components, for example the equation $(3x-2)\cdot(6x+2) = (3x-3)\cdot(3x-2)$. In this case students could either recognize common factors on two sides of the equation, or expand to work towards the Quadratic Formula. In both cases students possess the knowledge that is needed for solving the task. Note that, although assignments have the same structure, the tasks are randomized, which means that every student gets a different assignment.

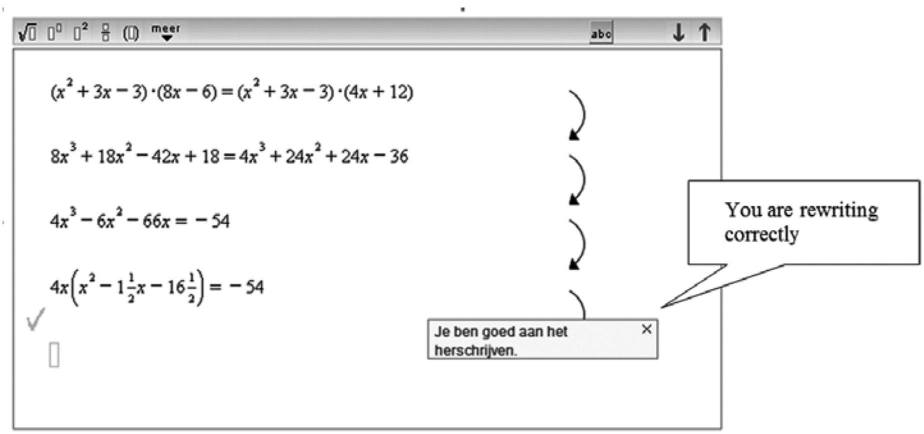


Fig. 3: Solving an equation with common factors (Bokhove & Drijvers, 2010b)

This student belonged to a group of students that did not recognize the common factors on the left- and right-hand side. When this happened, a much used strategy would consist of an algorithm that they apparently described as 'would always work': expand and work towards the Quadratic Formula. The yellow tick denotes that the expression is algebraically equivalent to the initial equation, but that it is not the final correct answer. This is accompanied by only part of the total score and feedback (in Dutch) on correctly having rewritten the expression. Although these students showed good rewriting skills, in the end they were not able to continue, as they did not possess the skill mastery to solve a third order equation. Right after this task a similar task is provided, but now with feedback pointing at the structure of the initial equation, and instructional screencast clips.

Throughout the module, the setup of the four series of tasks had the following sequential design: first 0-7 tasks that could be solved with pre-existing knowledge, then a crisis task, followed by tasks that are accompanied by screencast clips with a worked example plus additional feedback. One difference with existing definitions of crises is that they mostly describe the crises that occur while solving regular tasks, pointing out a gap in the student's knowledge. In this case crises are provoked intentionally to (hopefully) activate knowledge and/or to make the student aware of this. By subsequently providing aid in overcoming the crisis, the crisis is addressed. We contend that this approach could aid the acquisition of algebraic expertise and aim to scrutinize the fraction of correctly answered items and number of attempts right during and after a crisis task.

Feedback design principle 3: variation

The third feedback design principle, that corresponds to the third research sub-question, concerns variation of feedback. Hattie and Timperley (2007) conducted a meta-review of the effectiveness of different types of feedback. The feedback effects of cues and corrective feedback are deemed best. Feedback questions work at four levels (focus of the feedback): (i) the task level: how well tasks are understood/performed (FT), (ii) the process level: the main process needed to understand/perform tasks (FP), (iii) the self-regulation level: self-monitoring, directing and regulating of actions (FR), (iv) and the self-level: personal evaluations and affect (usually positive) about the learner (FS). Hattie and Timperley (ibid.) also provide some statements on the effectiveness of (combinations of) feedback types, including that FS feedback is least effective, simple FT feedback is more effective than complex FT feedback, FT and FS do not mix well ("Well done, that is correct" is worse than "Correct" only), and that FT is more powerful when it involves faulty interpretations. Furthermore they state that we should be attentive to the varying importance of feedback information during study of the task.

In this intervention, custom feedback and so-called IDEAS feedback are used to provide more elaborate, 'intelligent' feedback. The Digital Mathematical Environment has a provision for feedback by connecting to the IDEAS web service (Heeren & Jeuring, 2010), as well as the feature of providing custom feedback, the latter of which is described by Bokhove (2010). The IDEAS web service is also implemented for other online mathematical environments.

Figure 4 shows the essential characteristics of the IDEAS implementation in the DME. Firstly, the general characteristics of IDEAS feedback, which include what feedback is shown when and where. These settings were used to create two different feedback conditions c1 and c2. Secondly, IDEAS implements a block of diagnostic messages, which concerns feedback on strategy, the 'correct step', but also possible 'detours'. The third block of feedback concerns rewrite rules, rules that can be applied to an expression. Finally there are buggy rules, which describe the feedback that appears when a mistake is made. Using the authoring environment we implemented two series of tasks, d1 and d3, with both custom and IDEAS feedback.

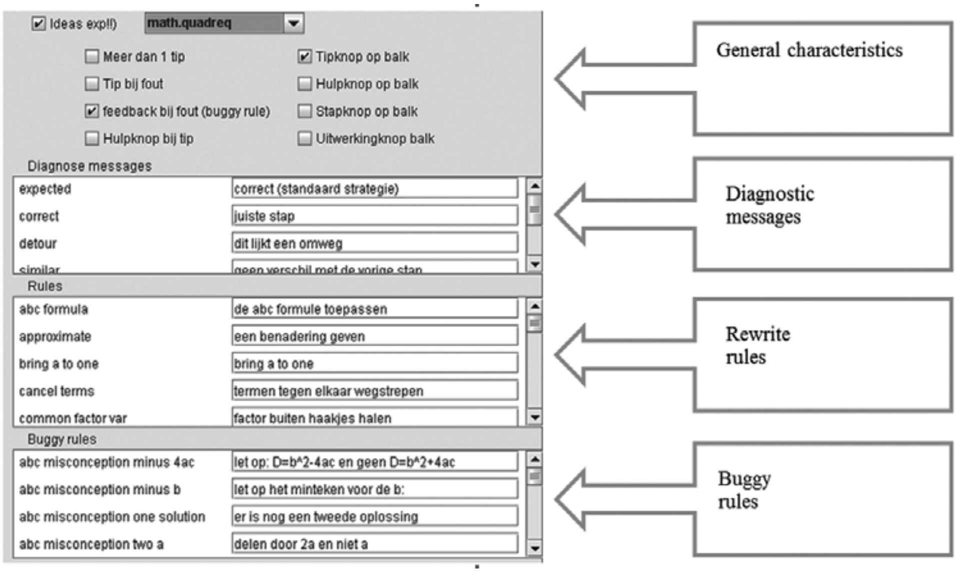


Fig. 4: Screenshot of DME's authoring environment for IDEAS feedback

Two different feedback conditions were implemented in the two series of tasks: one condition c1, and one condition c2. The first feedback condition c1 consisted of IDEAS and custom feedback without buttons in the interface. Feedback is only provided in the stepwise approach.

To illustrate this, figure 5 shows the solution process for a polynomial equation. The student loses solutions for the equation along the way, and appropriate feedback warns the student that this is happening: "You are about to lose two solutions. Keep in mind that the expression $(2x^2 + 4x - 3)$ also yields two solutions. Please complete". This feedback is along the lines of 'feedback about the task' (FT). The second feedback condition c2 is the same as c1, but *additionally* provides several buttons on the screen that could be used for getting hints and solutions of the exercises.

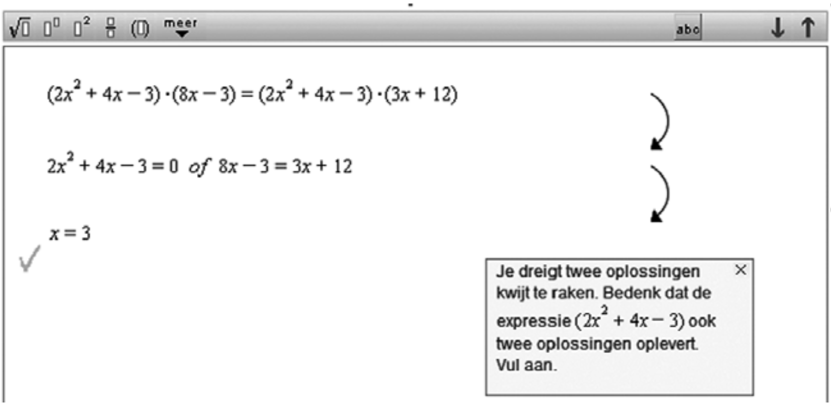


Fig. 5: Stepwise custom feedback

Figure 6 shows these buttons for (i) tip, which provides a hint for the next step, (ii) stap, which provides the next step in the solution process, and (iii) los op, which solves the whole equation and thus provides a 'worked example'.

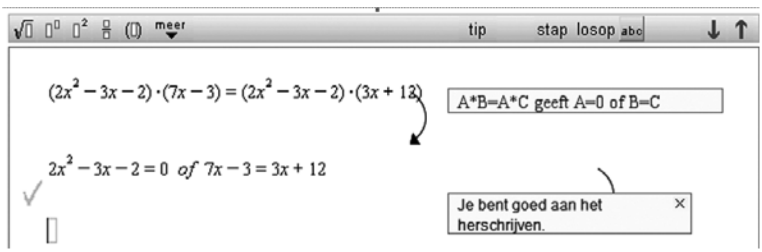


Fig. 6: Feedback condition c2, including buttons

These buttons can be used at will by the student, providing self-regulatory tools (feedback type FR). In the case of Figure 6 the student used the 'stap' button to obtain the next step in the solution process, and feedback 'A*B=A*C gives A=0 or B=C'.

6 Results of the feedback design

Results for design principle 1: timing and fading

Table 2: Fractions of correct items per category(1)

Category	Pre	d1	d2	d3	d4	d5(2)	d6	Post
Cat1	0.532	0.929	-	-	-	0.728	0.630	0.872
Cat2	0.248	-	0.938	-	-	0.841	0.776	0.825
Cat3	0.566	-	-	0.950	-	0.794	0.627	0.755
Cat4	0.571	-	-	-	0.853	0.759	0.656	0.872
Control	-	-	-	-	-	-	-	0.629
Total	0.507	0.929	0.938	0.950	0.853	0.780	0.672	0.787

(1) Skipped items were not included in the "fraction correct".

(2) Note that the final d5 score is a result of both scores and point deductions for attempts. Therefore the final score for d5 is not the same as the "fraction correct". For example, with deductions, the score is 49 out of 100 points. This is much less than 0.780 as tabulated above. This discrepancy is not applicable to d6, as there were no point deductions.

. Table 2 shows the fraction of correct items per category. It suggest that the formative scenario with timing and fading (i) shows a relatively low pretest score, (ii) high tool-aided scores in d1-d4, (iii) d5, on their own feet, worse (iv) d6 even worse, but finally the posttest significantly better than the pretest. Apparently, the fading feedback did not hinder the effectiveness of the intervention. Rather, we contend that the fact that scores declined after the practice module, but still maintained an overall significant effect, suggests that the addition of 'faded' modules d5 and d6 makes a difference. This hypothesis is also supported by the scores on the four separate item categories.

Results for design principle 2: crises

Figure 7 shows the fraction of correctly answered items per part of the intervention, d1 to d4. The arrows denote the location of the crisis task. Looking at the scores the crisis are not really noticeable. Therefore we also looked at the number of attempts. Figure 8 shows similar graphs, but now with the total number of attempts plotted on the vertical axis (N remained 324 during the tasks). Attempts seem to give a better indicator for a crisis with the trend for d1, d2, and d4 being that the number of attempts needed decrease after a crisis task. Part d3 did not show this trend. Thus, for d1, d2 and d4 performance, measured as number of attempts needed, increased after a crisis. For d3 the number of attempts increased after crisis item 3.5. A qualitative post-analysis showed that this was caused by the use of negative and broken exponents in items 3.7 and 3.8.

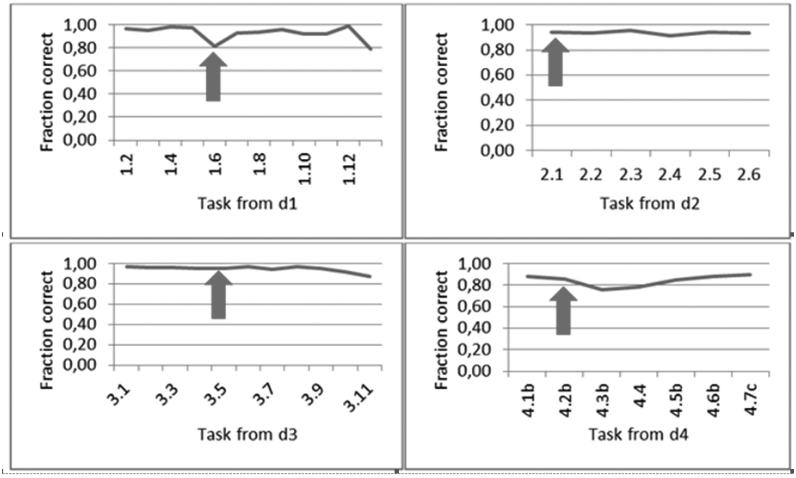


Fig. 7: Fractions of correct items for tasks from d1-d4

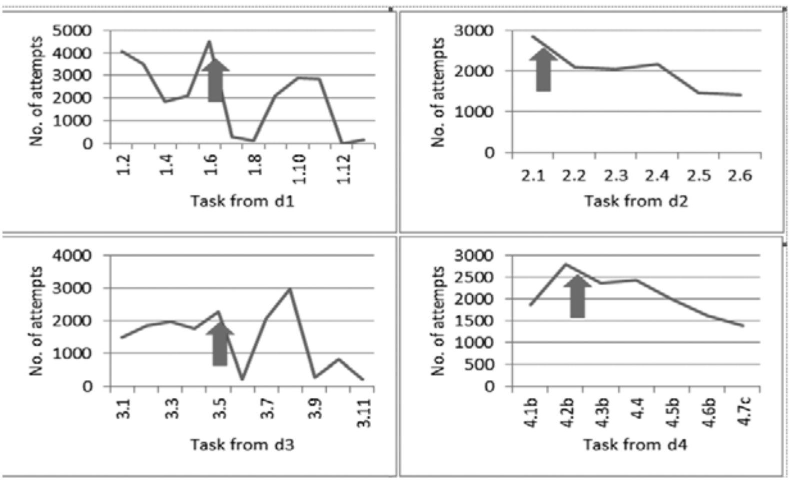


Fig. 8: Number of attempts spent on tasks from d1-d4

Results for design principle 3: variation

For the results on the two feedback conditions, we first present the quantitative findings. Next, these will be illustrated by some case examples. As a Kolmogorov-Smirnov test shows that both d1 and d3 scores are not distributed normally (Z=5.408, p<0.001; Z=6.768, p<0.001 respectively), we apply a non-parametric Mann-Whitney test. This test shows that there is a significant difference between the feedback conditions when we look at the score for d1 (U=7680.00, p<0.001, r=-.321) with condition c2 scoring higher than condition c1. According to Cohen (1992) this accounts for a medium effect size. The second series of tasks d3, however, did not show a significant difference (U=10560.00, p=.531). When considering the number of attempts, this was significantly higher for feedback condition c1 (*Mdn*=126.00) (without extra buttons) than for feedback condition c2 (*Mdn*=105.00), U=9904.50, p<.001, r=.202. Although we can classify this as just a small effect, it suggests that the addition of buttons for hints and solutions results in less attempts.

Apart from these quantitative results there also is a substantial body of case examples where students have successfully or unsuccessfully used feedback options. We now illustrate the use of IDEAS feedback with three examples from both feedback conditions, one example from c1, one example of a student using the hint button from c2, and one example of a student using the button for a worked solution in c2. Note that feedback was translated from Dutch into English for the purpose of this article, and that randomization of the tasks means that students received different equations to solve.
The symbols ✗, ✓ and respectively designate incorrect answers, correct but not final answers and correct answers.

Table 3: Feedback example from condition c1 (stepwise feedback)

Step	**Student**		**Feedback**
0	$(-4x+5)(8x-5) = (-4x+6)(4x+14)$		
1	$-32x^2+60x-25 = -16x^2-32x+84$	✓	You are rewriting correctly
2	$-16x^2+92x+190 = 0$	✗	Hint: rewrite to [expression]=0
3	$-16x^2+92x+109 = 0$	✓	You are rewriting correctly
4	$x = \frac{-92-\sqrt{1488}}{-32}$ or $x = \frac{-92+\sqrt{1488}}{-32}$		You have solved the equation correctly

In the example in Table 3 the student starts by expanding the left and right hand side of the equation. As there are no feedback buttons available, the system evaluates the expression in step 1 as correct (but not the final solution). The student makes a calculation error in rewriting in the form [expression]=0. Now the system hints that the expression is incorrect and gives feedback. The difficulty of judging what mistake was made instantly becomes apparent: in this case the student already understood he/she should rewrite to [expression]=0 but makes a calculation error. In step 3 this error is corrected. The student then remembers to apply the Quadratic formula and solves the equation correctly. We see that the system does still have difficulties with judging what format of the expression is 'good enough' and what is not. The evaluation is quite 'liberal', giving the notation in step 4 full marks.

The second example, shown in Table 4, reveals that the additional buttons help a student to overcome initial difficulties with the given equation.

Table 4: Feedback example from condition c2 (hints)

Step	Student		Feedback
0	$(x^2 + x - 6)(7x - 6) = (x^2 + x - 6)(3x + 12)$		Hint: AB=AC => A=0 or B=C
1	$x^2 + x - 6 = 0$ or $7x - 7 = 3x - 12$	✗	
2	$x^2 + x - 6 = 0$ or $7x - 6 = 3x - 12$	✓	You are rewriting correctly
3	$x = \frac{-1-\sqrt{25}}{2}$ or $x = \frac{-1+\sqrt{25}}{2}$ or $x = 4\frac{1}{2}$	✓	This is not quite the exact format
4	$x = -3$ or $x = 2$ or $x = 4\frac{1}{2}$		You have solved the equation correctly

After being given a new equation the student uses the ‘tip’ button to get a hint. The student uses the hint to apply the correct strategy, first making a notational error, but correcting this in step 2. After this the student concludes the task in steps 3 and 4. In step 3 the system prompts that the expressions can be simplified.

Table 5: Feedback example from condition c2 (solve)

Step	Student		Feedback
0	$(2x^2 - 4x - 4)(6x - 11) = (2x^2 - 4x - 4)(3x + 14)$		AB=AC => A=0 or B=C
1	$2x^2 - 4x - 4 = 0$ or $6x - 1 = 3x - 14$	✓	Rewrite in form [expression]=0
2	$2x^2 - 4x - 4 = 0$ or $3x - 25 = 0$	✓	Move constants to the right
3	$2x^2 - 4x - 4 = 0$ or $3x = 25$	✓	Free up variable by dividing on both sides
4	$2x^2 - 4x - 4 = 0$ or $x = \frac{25}{3}$	✓	Simplify by factoring
5	$x^2 - 2x - 2 = 0$ or $x = \frac{25}{3}$	✓	Use quadratic formula
6	$x = \frac{2+\sqrt{12}}{2}$ or $x = \frac{2-\sqrt{12}}{2}$ or $x = \frac{25}{3}$	✓	Simplify roots
7	$x = 1 + \sqrt{3}$ or $x = 1 - \sqrt{3}$ or $x = \frac{25}{3}$		You have solved the equation correctly

The third and last example, shown in Table 5, concerns a student that uses the ‘los op’ button to automatically solve the given tasks, with the system adding the strategies as feedback for every step in the solution process. This example shows the difficulties of evaluating student answers. Most solutions from students combined several steps into one. For example, steps 2, 3 and 4 could easily be combined in one step. The notational issue mentioned earlier also crops up: many teachers would perhaps have given full marks for the solution in step 6, but because the square root *can* be simplified this is not seen as the final solution.

7 Conclusion

The general question we wanted to answer was in what way can feedback be used in the design of an intervention and what are its effects? The question was elaborated in three sub-questions that involved the framework for feedback: timing and fading, crises and variation.

(1) Timing and fading: are there indications that formative scenarios with fading feedback improve the acquisition of algebraic expertise?
(2) Crises: do crises in algebraic task sequences improve the acquisition of algebraic expertise?
(3) Variation: does variation in feedback influence scores and student behavior?

Conclusion on timing and fading

The score progress for the intervention shows that an initial pre-knowledge score is followed by a large improvement of scores for algebraic expertise when in practice mode. After this, in self-test mode, the scores dropped markedly, and in the final digital test they dropped even further. Overall, however, the intervention showed a significant improvement across all four task categories. The improvement is biggest for the second category; the third category shows the least improvement. Thus, although our formative scenario causes ‘pain’ in gradually fading feedback in the course of the intervention, the learning effect is still there. The design criterion of using formative scenarios seems to improve the acquisition of algebraic expertise.

Conclusion on crises

The results are mixed for this design principle. When looking at the fraction of items answered correctly there seemed to be no effect for a crisis task. This is probably caused by the fact that difficulty with tasks cannot be seen if the final score is high, although achieved with difficulty. The number of attempts after a crisis task does decrease in three of four series of tasks. In series category one, two and four it seems the crisis is followed (trend) by a drop in number of attempts and an increase in score. For category three, however, this is not the case. From a qualitative post-analysis we conjecture that this was caused by the use of negative and broken exponents in items in the tasks. Whether this improves the student's actual expertise is unclear from the results.

Conclusion on variation

When observing the scores students obtained in the two conditions, we can see that there is a medium effect for the feedback condition including self-regulatory feedback (condition c2). This effect, however, only was apparent in one of the series of tasks. We think we can explain the difference between d1 and d3 in the fact that both series address the category of polynomial equations (Bokhove & Drijvers, 2010). Having solved polynomial equations with feedback in d1 meant that the students were already capable of solving similar types of equations (in d3, which followed after d1), and subsequently no additional form of self-regulation was needed in d3. The addition of buttons for feedback also had an, albeit small, effect on the number of attempts. This makes sense as the additional feedback that can be requested discourages more attempts. When looking in more detail at the use of the feedback in the three case examples it is clear that both task-related and self-regulatory (FT and FR) feedback can be used in a formative way *for* the learning of algebra. Students can use the feedback to overcome difficulties and check whether they are on the correct solution path or not.

Taken together we can see that the use of feedback in an online algebra intervention can have a significant effect. Furthermore, the principles provide a broad approach to the potential of feedback use in digital tools. However, there still is a lot to work out. One difficulty in regard to feedback is whether it was actually the feedback (formative scenario, crisis or feedback variation) that improved scores. Especially with students using the online algebra intervention at home, it is quite difficult to control for many of the variables. Another word of caution with regard to ‘intelligent’ feedback is needed. Although research suggests that worked examples are effective (Sweller & Cooper, 1985) students could easily be tempted to ‘just push the button’ (this is an actual statement from a student) to get full marks. Although the results of using custom and IDEAS feedback for algebraic expertise are promising, there still are many improvements to be made. These improvements should focus primarily on notational aspects as shown in the case examples. Student motivation declines when they do not get full marks from the system, and no explanation is given. Secondly, feedback should be adapted to the target audience and math curriculum. Clearly, the mathematical language of higher education is different from that of secondary education. It is not a viable option to let all teachers author their own set of feedback comments. One goal in the near future will be to try and provide default values for feedback that applies to the most common student errors and behaviors, resulting in feedback ‘out-of-the-box’. It is imperative that the appropriateness and quality of ‘intelligent’ feedback is improved before we can reap the benefits. One good method to achieve this would be to use student inquiry (Bokhove, 2010).

The three feedback design principles were implemented because we wanted to answer the general question which design principles may guide feedback design in an online algebra intervention and what are the effects of these design choices? We have demonstrated this in the previous sections. Nevertheless they are are only a cautious first step on the road to actual use and implementation of feedback in computer tools for education. If we acknowledge the ‘power of feedback’ (Hattie & Timperley, 2007) in technology we should be willing to think about the implications this has for educational practice, otherwise this power could easily turn out to be a hindrance, rather than an advantage. Therefore the conditions under which computer feedback flourishes should be pursued further. As one student said: “although there are some bugs in the feedback this is a good program which I would like to use more often.”

Chapter 6

Effects of a digital intervention on the development of algebraic expertise

Bokhove, C., & Drijvers, P. (in press). Effects of a digital intervention on the development of algebraic expertise.*Computers & Education.* http://dx.doi.org/10.1016/j.compedu.2011.08.010

Effects of a digital intervention on the development of algebraic expertise

1 Introduction

According to a plethora of available resources, the potential of computers for K-12 education has been widely recognized (Voogt & Knezek, 2008). However, for a successful integration of ICT many factors have to be taken into account. Some factors concern the individual level, such as students' computer attitudes, computer experience and gender; other variables operate at class and school level and stress the gap between the proposed and implemented ICT curricula (Tondeur et al.,2008). Therefore, more knowledge is required about both the optimal conditions to benefit from ICT's potential and the relevant characteristics of ICT interventions.

The optimization of ICT's potential also concerns the field of mathematics education. In line with recent research findings (e.g. Goos et al., 2009; Heid & Blume, 2008a, 2008b; Pierce & Stacey, 2010), the U.S. National Council of Teachers of Mathematics (2008) acknowledges the potential of ICT for mathematics education in its recent position statement. As a result, the last decade has seen a rise of (online) environments for algebra and accompanying research (Berger, 2010; Kim & Wei, 2011). However, educational use of ICT for mathematics often consists of 'drill-and-practice' activities and as such, seems to focus on procedural skills rather than on conceptual understanding. This raises the important question of how students best acquire algebraic expertise: by practicing algorithms, or by focusing on reasoning and strategic problem solving activities? Research within the framework of instrumental and anthropological approaches shows that there is an interaction between the use of ICT tools and conceptual understanding and that the successful integration of ICT into the classroom is more complex than expected (Artigue, 2002). For example, students need to be able to reconcile conventional pen-and-paper techniques and ICT techniques (Kieran & Drijvers, 2006). Also, the use of ICT can easily shift from ‘using it to learn’ towards ’learning to us’ a tool, and subsequently just ‘pushing the buttons’ (Zorn, 2002). Furthermore, although ICT can provide a space to ‘learn by doing’ (Papert, 1980), its use should also be guided by instructions, so as to avoid the ‘play paradox’ of accomplishing a task without learning the intended concepts (Noss & Hoyles, 1992).

Important characteristics of ICT tools that can be used for addressing both procedural skills and conceptual understanding in mathematics include options for the recording of the student's solution process, and the possibility for students to use different strategies through a stepwise approach. This enables the student to apply his or her paper-and-pencil reasoning steps and strategies (Bokhove & Drijvers, 2010a). Building on Buchberger's

(1990) whitebox/blackbox distinction, Beeson's glass box principle (1998) stresses the importance of transparency of the solution process for educational practice.

In sum, the aforementioned literature reveals many challenges for the inclusion of ICT within algebra education. In this light, the aims of this study are (1) to design an online environment and an intervention for the learning supported by formative assessment for both procedural skills and conceptual understanding in algebra; (2) to investigate its effects; and (3) to identify decisive factors that influence the intervention's outcome.

2 Conceptual framework

As the study aims to design an intervention for learning algebra supported by formative assessment and to study its effects, the conceptual framework consists of principles from formative assessment and from algebra didactics..

Formative assessment

Black and Wiliam (2004) distinguish three functions for assessment: supporting learning (formative), certifying the achievements or potential of individuals (summative), and evaluating the quality of educational programs or institutions (evaluative). Summative assessment is characterized as assessment *of* learning and is contrasted with formative assessment, which is assessment *for* learning. Black and Wiliam (1998) define assessment as being 'formative' only when the feedback from learning activities is actually used to modify teaching to meet the learner's needs. From this it is clear that feedback plays a pivotal role in the process of formative assessment. Hattie and Timperley (2007) conducted a meta-review of the effectiveness of different types of feedback. The feedback effects of cues and corrective feedback are deemed best. The model provided by Hattie and Timperley (2007) distinguishes three questions that effective feedback answers: the question ‘Where am I going?’ (feedup), the question ‘How am I going?’ (feedback) and the question ‘Where to next?’ (feedforward). As Table 1 denotes, each feedback question works at four levels task, process, self-regulation and self.

Table 1: Four levels of feedback (Hattie & Timperley, 2007)

Id	Level	Description
FT	Task	how well tasks are understood/performed
FP	Process	the main process needed to understand/perform tasks
FR	self-Regulation	self-monitoring, directing and regulating of actions
FS	Self	personal evaluations and affect

Hattie and Timperley (2007) also provide statements on the effectiveness of (combinations of) feedback types. FS feedback is least effective, simple FT feedback is more effective than complex FT feedback, FT and FS do not mix well, and FT is more powerful when it concerns faulty interpretations. Furthermore, they stress the varying importance of the feedback information during a task. In this study, these different faces of feedback form the basis for the intervention's main design principles..

Algebraic expertise

The distinction between procedural skills and conceptual understanding is a highly researched field of interest. The book *Adding it up* (Kilpatrick, Swafford, & Findell, 2001) synthesizes the research on this issue in the concept of *mathematical proficiency*, which comprises five strands: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning and productive disposition. Here, conceptual understanding is defined as “the comprehension of mathematical concepts, operations, and relations” (p. 116), and procedural fluency as the “skill in carrying out procedures flexibly, accurately, efficiently, and appropriately” (p. 116). Furthermore, “the five strands are interwoven and interdependent in the development of proficiency in mathematics” (p. 116). Arcavi introduces the notion of *symbol sense*, which includes “an intuitive feel for when to call on symbols in the process of solving a problem, and conversely, when to abandon a symbolic treatment for better tools” (Arcavi, 1994, p. 25). He describes eight behaviors in which symbol sense manifests itself. These behaviors show the intertwinement between procedural skills and conceptual understanding as complementary aspects of algebraic expertise. Both procedural skills and symbol sense need to be addressed in algebra education, as they are intimately related: understanding of concepts makes basic skills understandable, and basic skills can reinforce conceptual understanding (Arcavi, 2005). Synthesizing this work, Drijvers, Goddijn and Kindt (2010) see algebraic expertise as a dimension ranging from basic skills to symbol sense (see Figure 1). Basic skills involve procedural work with a local focus and emphasis on algebraic calculation, while symbol sense involves strategic work with a global focus and emphasis on algebraic reasoning.

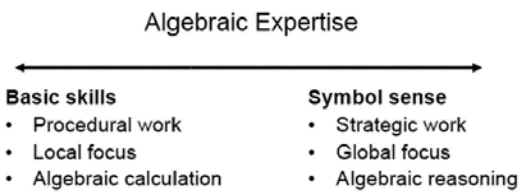


Fig.1 : Algebraic expertise as a dimension (Drijvers, Goddijn & Kindt, 2010)

Figure 1 provides the lens through which algebraic expertise is considered in this study. To address both ends of the dimension, the online activities are meant to offer both procedural and symbol sense opportunities. From literature on symbol sense, four types of items are identified (Bokhove & Drijvers, 2010b). The main criterion is that items would invite both procedural skills and symbol sense, with an emphasis on the latter. The four item categories are as follows..

Category 1: Solving equations with common factors.
Items in this category are equations with a common factor on the left and right-hand side, such as:

Solve the equation:

$$(x^2 - 7x + 12) \cdot (8x - 11) = (x^2 - 7x + 12) \cdot (3x + 14)$$

Category 2: Covering up sub-expressions.
In this category, sub-expressions need to be considered as algebraic entities that can be covered up without caring for their content. A well-known example from Wenger (1987) is:

Solve for *v*:

$$v \cdot \sqrt{v} = 1 + 2v \cdot \sqrt{1 + u}$$

Category 3: Resisting visual salience in powers of sub-expressions.
This category is about recognizing when to expand expressions and when not. It contains equations with sub-expressions that just beg to be expanded because they are raised to a power:

Solve the equation:

$$(x - 3)^2 + 4 = 40$$

Category 4: Recognizing 'hidden' factors.
This category concerns the recognition of factors that are not immediately apparent (gestalt). An example is the following item adapted from Tempelaar (2007):

Rewrite $\dfrac{x^2 - x}{x^2 - 2x + 1}$

Even if these categories may seem quite specific, they share the overall characteristic of an intertwinement between local and global, procedural and strategic focus (Drijvers et al., 2010; Hoch & Dreyfus, 2004; Kilpatrick et al., 2001; Pierce & Stacey, 2007).

The study's aim is to investigate the effects of activities in an online exercise environment for the learning of both procedural skills and conceptual understanding in algebra - the latter further defined by the four categories above - and to identify possible predictors for its successful use. These aims are elaborated into two research questions:

(1) What is the effect of an intervention on the development of algebraic expertise, including both procedural skills and symbol sense?
(2) Which factors predict the resulting algebraic performance?

3 Methods

Type of study

This study globally followed a design research approach with four phases. The preliminary research phase concerned the choice and design of a digital tool for algebra (Bokhove & Drijvers, 2010a). The first intervention cycle focused on whether the prototypical digital activities would invite symbol sense behavior through 1-to-1 sessions (Bokhove & Drijvers, 2010b). The second cycle consisted of a small scale field experiment in one school (N=31). The third and final cycle concerned a large-scale classroom experiment (N=324). This research methodology shows a progress from small-scale to large-scale in 'layers of formative evaluation' (Tessmer, 1993), with an accompanying shift from more qualitative to more quantitative analyses. The last cycle, aiming at the intervention effects, is the focus of this chapter.

Participants

Participants were 324 12th grade 17-18 year old students from fifteen classes from nine Dutch secondary schools, involving eleven mathematics teachers. The schools were spread across the country and located in seven provinces. There also a spread in school size and pedagogical and religious backgrounds. The participating classes consisted of pre-university level 'wiskunde B' students (comparable to grade 12 in Anglo-Saxon countries). 43% of the participants were female and 57% were male. The participating schools reacted to an open invitation made in several bulletins for mathematics education. Schools received an exemplary course planning guide and some hints on using the intervention's software tools. They were however free to adapt the intervention to their own requirements. Registration of key decisions through questionnaires and logging made sure that the way in which the intervention was deployed in each class could be taken into account in data analysis. Schools deployed the intervention in the last three months of 2010, just before the commencement of preparations for the final national exams. Teachers received mailings on a regular basis, and could visit the project website www.algebrametinzicht.nl.

Intervention

The intervention is depicted graphically in the upper part of Figure 2 and consists of a pen-and-paper pretest, four digital modules, called d1-d4, each covering one of the four item categories described above (Bokhove & Drijvers, 2010b), digital module d5 as a diagnostic tool, digital module d6 as a final digital test and, finally, a pen-and-paper posttest.

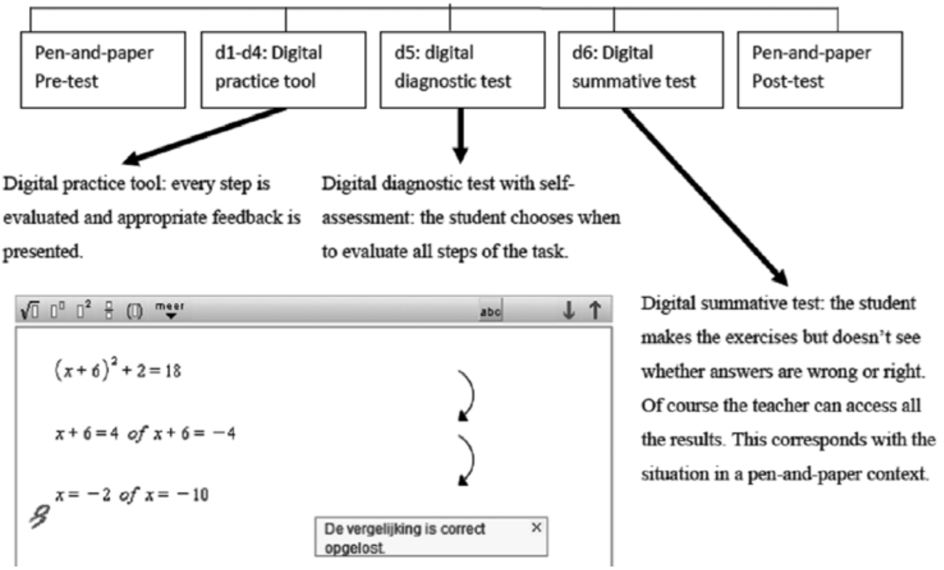


Fig. 2: Outline of the formative scenario underpinning the intervention

The total time needed to complete the module was estimated to be six hours work, excluding pre- and posttests. For the item design, we used sources related to the transition from secondary to tertiary education, such as high school exit and university entry examinations, remedial courses, text books and professional journals. Table 2 shows whether an intervention element was presented digitally or with pen-and-paper, the number of items and Cronbach's alpha. The reliability of the test items was determined in the second cycle of the study, showing alphas ranging from .7 to .9, which is confirmed by the alphas in the third and final cycle. The items for pretest, d5, d6, and posttest are taken from the 45 items that comprised d1-d4 , with a proportion distributed over the four item categories, and in random order. In modules d1-d4, the order of the items followed the structure (i) pre-crisis items, (ii) crisis item, and (iii) post-crisis items. For a full understanding of this structure we refer to the design principle on crises in section 3.3.2. The alpha values show that all parts of the intervention can be seen as reliable (Garson, 2011).

Table 2:: Reliability of the parts of the intervention

Part	Pre	d1	d2	d3	d4	d5	d6	Post
Items N	8	12	6	11	16	23	23	10
Alpha	.724	.687	.917	.909	.880	.843	.828	.709
Digital (d)/pen-and-paper (p)	p	d	d	d	d	d	d	p

Three intervention design principles guided the design of the intervention: (1) the notions of timing and fading within formative scenarios, and (2) crises. They both make use of the principle on (3) feedback elaboration, which was described in section 2.1. We now discuss the other two principles.

Design principle: timing and fading within formative scenarios

Black and Wiliam (1998) define assessment as being 'formative' only when the feedback from learning activities is actually used to modify teaching to meet the learner's needs. From this it is clear that feedback plays a pivotal role in the process of formative assessment. For the purpose of this chapter we mention that the intervention was used in two feedback conditions c1 and c2, both consisting of a varying amount of feedback. We refer to Bokhove and Drijvers (2011) for more detail on this. Based on the principles of formative assessment we propose the concept of a formative scenario defined as a structured collection of learning activities. Initially these provide a lot of feedback to foster learning, but decrease the amount of feedback towards the end, to facilitate transfer. The intervention is based on the concept of fading (Renkl, Atkinson, & Große, 2004). Fading refers to the idea that one starts off with completely elaborated items and moves to less elaborated items by removing intermediate steps. According to Jones and Fleischman "faded examples cause effective learning by forcing the student to encounter and overcome an impasse" (Jones & Fleischman, 2001, p. 5). An impasse occurs when a learner realizes a lack of complete understanding of a piece of knowledge (see Figure 2). The intervention - following a pen-and-paper pretest - starts with a practice module d1-d4. In these modules, the tool provides feedback on every step. The next step in the intervention concerns self-assessment and diagnostics: the student performs the algebraic manipulations without any feedback and chooses when to check his or her solution. Feedback is then provided for the item as a whole. The number of attempts needed for finding the solution determines the credits given: success after one attempt earns bonus points, and when many attempts are needed points are deducted. Finally, students get an online final test, in which they cannot see how they perform. This online test is graded automatically and the teacher provides students with feedback on their performance. After all, a student needs to be able to accomplish tasks independently, without the help of a computer. This approach also emphasizes the importance of timing in feedback as mentioned in ‘the power of feedback’, when they discuss the varying importance of the feedback information during a task (Hattie & Timperley, 2007). To conclude the intervention the students do a pen-and-paper posttest.

Design principle: crises

With a crisis we refer to a principle that the poet John Keats so eloquently described in the early 19th century:

> Don't be discouraged by a failure. It can be a positive experience. Failure is, in a sense, the highway to success, inasmuch as every discovery of what is false leads us to seek earnestly after what is true, and every fresh experience points out some form of error which we shall afterwards carefully avoid. (John Keats, 1795 - 1821)

This principle, with dialectic characteristics, corresponds to similar concepts, such as cognitive conflicts (Tall, 1977) , Van Hiele's 'crises of learning' (Van Hiele, 1985) and 'productive failure' (Kapur, 2008). It also ties in with the impasses that we described in the section on formative scenarios. The idea in the intervention is that when students rely on algorithms that seem to always work, this conviction is challenged by a conceptual crisis: an item that is hard or impossible to accomplish with the algorithms that worked so well so far. Figure 3 shows items from the sequence for category one, which illustrate the way in which crisis items are integrated within d1 to d4. The general structure of a sequence is: pre-crisis items, crisis item, post-crisis items. Note that, although assignments have the same structure, the items are randomized, which means that every student receives a different assignment.

Data

The following data were collected. Examples of these data can be found in appendix B.

(1) General characteristics of students and classes, including gender and exam results.

(2) Pre- and posttest results (scores out of 100 points). Both pre- and posttest were scored on (partly) correct answers through a uniform grading model.

(3) Scores and log files of the digital module (including time, number of attempts, number of errors). All activity within the environment was logged for each of the students.

(4) Pre questionnaire for students, based on the work of Reed, Drijvers and Kirschner (2010), which probed attitudes and behaviors concerning mathematics.

(5) Post questionnaires for both students and teachers addressing the quality of the module, the students' motivation and the way they perceived the intervention in the classroom. In this chapter only a few variables from the student questionnaire are included in the analysis.

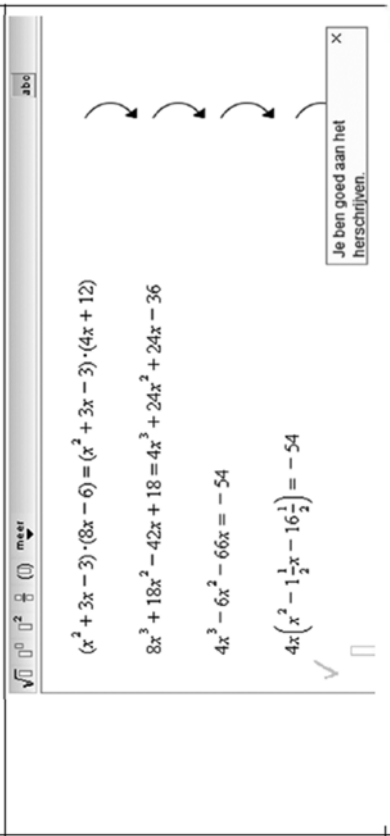
Pre-crisis items In the initial items students are confronted with equations they have experience with. Students may choose their own strategy. Many students choose to expand brackets as that is the strategy that they have used often: work towards the form $ax^2 + bx + c = 0$ and use the Quadratic Formula. There is some limited feedback on the task (FT).	
Crisis item Students are then confronted with an intentional crisis: if a student uses his/her conventional strategy of expanding the expression. The yellow tick at the bottom of the screen denotes that the equation is algebraically equivalent to the initial one, but that it is not the final answer. This is accompanied by a partial score for an item and some feedback in Dutch: 'You are rearranging correctly'. Although these students showed good rearranging skills, in the end they were not able to continue, as they did not master the skill to solve a third order equation. There is some limited feedback on the task (FT).	
Post-crisis items After the crisis item students are offered help by providing a 'voorbeeldfilm', an instructional screencast, and buttons to get hints ('tip'), the next step in the solution ('stap') or a worked solution ('losop'). These features have in common that they provide feedforward information at the task level (FT) and self-regulation (FR).	

Fig. 3: Sequence of items illustrating the design principle with crises

Data Analysis

Two analyses were carried out to answer the research questions. For the first questions concerning effect, we used t-tests (Field, 2005). For determining predictors explaining the effect, we used multilevel analysis, for which a rationale is given in the next section.

Variables for research question 1

For the first research question on the effect of the intervention, the dependent variables were the posttest score (post-test, score out of 100) and scores for symbol sense behavior. Symbol sense behavior was determined per item through a three value code, with 1 signifying symbol sense behavior as operationalized in the four categories (Bokhove & Drijvers, 2010b), -1 signifying a lack of symbol sense behavior as missing symbol sense opportunities, and 0 for none of the two (e.g., the student not even starting the solution process). Appendix B gives some examples of these codings. Through these codes, pretest and posttest sums for symbol sense behavior were calculated, and called symsensepre and symsensepost respectively. To check for inter-rater reliability Krippendorff's alpha was calculated for all dependent variables by having a second rater grade and code the pre- and posttest as well. The results (alpha=.91 for all items of 5% of the students' pretests and 5% of the posttests) indicates a very good inter-rater reliability (Hayes & Krippendorff, 2007). The independent variables for research question 1 are the separate classes and the categories described in section 2.2.

Multilevel analysis for research question 2

In the reported study, there were 324 students, divided into 15 classes from 9 different schools. A major advantage of multilevel models compared to single-level regression analysis is the possibility to explore relationships among variables at different levels simultaneously (Goldstein, 2002; Raudenbush & Bryk, 2001; Snijders & Bosker, 1999). A typical multilevel model involves outcome variables that are related to several predictor variables at the individual level (level one; in this case the student level) and at the group level (level two; in our case the class level). In modeling we followed Dedrick et al.'s guidelines (2009) for multilevel analyses, which include the descriptions of the models including predictors and covariates, the centering that is used, distributional assumptions, outliers, missing data, software and method of estimation used, and a list of parameter estimates and standard errors (section 4).

Gelman and Hill (2006) emphasize that multilevel models allow us to study effects that vary by group, for example an intervention that is more effective in some schools than others (perhaps because of unmeasured school-level factors such as teacher morale), allowing the estimation of group averages and group-level effects. One concern was the relatively low N of 324, but Gelman and Hill (2006) state that the key concern with multilevel modeling is the estimation of variance parameters, but that it should still work at least as well as classical regression. The analysis of the null model will show that there are significant differences between classes, and thus multilevel modeling is appropriate. As it was not possible to increase the number of classes, the available power was improved by collapsing the number of levels to two: classes and students.

The software used was MLwiN 2.22 with estimation procedure RIGLS. Before performing the whole sample analysis, extreme outliers in independent variables were changed to 'MISSING' (e.g. the time spent in the tool: one session of 10 hours). To facilitate interpretation of results, explanatory variables were rescaled to have a meaningful zero point by using Grand Mean centering. Missing data was treated with macros for multiple imputations (Carpenter & Kenward, 2010). After this data treatment, removing extreme outliers (decreasing N to 286), the residual plot for student level shows that residuals are approximately normally distributed. The assumption of homoscedasticity also was met therefore we performed the analysis.

Subsequent models were improved in an iterative fashion. First, the intercept only one-level and two-level models were compared to determine whether there was variance at the class level. Incrementally, first- and second-level predictor variables were then entered into the model. Next, random slope variation was added to the significant predictors to test whether regression slopes varied across classes. Significance testing was one-sided.

The added value of a new model was checked by looking at the proportion of explained variance, while model improvement is assessed by comparing deviance values, the log-likelihood, treating the difference as being chi-squared distributed (Hox, 1995; Snijders & Bosker, 1999).

Variables for research question 2

For the multilevel analysis the outcome variable is the posttest score (post-test, score out of 100). The independent variables or predictors are grouped for multilevel analysis as follows:

Pre-knowledge. This set of variables concerns knowledge that influences student competency upfront, and include the trend of all exam grades on a school for the last three years (trendexam, score out of 10, two digits) and the pretest score (pre-test, score out of 100). Trendexam is at the class level, pre-test at the student level.

Time spent. This set of variables concerns the total time spent in the module, time_total, also broken down in time per part of the digital module: time_manual for the online manual, time_d14 for the time spent working on parts 1 to 4, time_d5 for time spent on part d5 and time_d6 for time spent on part d6. The variables initially were recorded in seconds, but were also modified to time in hours, adding '_h'' to the variable names. All variables are at the student level.

Tool-related. This set of variables pertains to the use of the tool, and contains variables, taken from the post-questionnaire: a self-report of students concerning the percentage of use of the tool at home (vs_phome, percentage out of 100) and the student verdict on the tool (vs_general), a scale consisting of 1 (negative), 2 (slightly negative), 3 (neutral), 4 (slightly positive) and 5 (positive). Both variables are at the student level.

Student attitude. This set of variables concerns the attitudes and behaviors of the students towards mathematics and the use of the tool, taken from the pre-questionnaire. One variable concerns students' self-reported motivation after the intervention (vs_motivation, 3 items). Four other scales, in line with Reed et al. (2010), concern (1) their general attitude towards mathematics (MATH, 5 items), (2) their attitude towards using computers for mathematics (ICT, 4 items), (3) the extent to which they reported purposeful and investigative behaviors when undertaking mathematical activities (PURINV, 6 items), and (4) the extent to which they reported reflecting upon and communicating about their thinking and actions (REFCOM, 8 items). The scales range from 1 (very negative attitude/low level of self-reported behavior) to 5 (very positive attitude/ high level of self-reported behavior). The midpoint represented a neutral attitude or medium level of self-reported behavior. All variables are at the student level. Example item are presented in appendix A.

4 Results

Intervention effects

Table 3 shows the descriptives of the output variables for the first research question, each signifying a score between 0 and 100, with the exception of d5 showing a few scores above 100 because of possible bonus points. The lower N's for the digital modules d5 and d6 are due to implementation differences between schools, probably as the result of some teachers' stronger encouragement for their students to work on the modules compared with others. This factor is taken into account in subsequent analyses by using the time spent in the separate modules within the intervention. As only 7 out of 292 students who sat the posttest scored 100 out of 100 (2.4%), we conclude that there was almost no ceiling effect.

Table 3: Descriptives of the parts of the intervention

	Min	Max	Mdn	SD	N
Symsensepre	-6.00	3.00	-1.00	2.35	318
Pre-test	2.00	98.00	51.00	21.37	318
d1-d4	0.00	100.00	97.25	21.08	311
d5	0.00	106.00	48.50	31.89	254
d6	1.00	100.00	68.00	28.44	223
Post-test	10.00	100.00	82.00	15.46	292
Symsensepost	-5.00	3.00	1.00	1.50	292

The results show that the score on the posttest (*M*=78.71, *SE*=15.175) is significantly higher than the pretest score (*M*=51.55, *SE*=21.094), *t*(286)=-22.589, *p*<.001, *r*=.801. According to Cohen's benchmark (1992) this suggests a large effect. Despite of this positive overall effect, thirty students out of 286 (10.49%) scored lower in the posttest than the pretest. Out of these thirty students, eight did not do any work in the module (total time spent = 0), and just made the pre- and posttest. A comparison between the remaining 22 students and the other students shows that these 22 students (*M*=77.18, *SE*=19.237) scored significantly better in the pretest than the rest of the students (*M*=49.17, *SE*=19.722), *t*(276)=6.404, *p*<.001, *r*=.360, showing a medium effect. These students did not spend (*M*=4.027, *SE*=2.557) significantly less time in the digital tool than the rest of the students (*M*=4.994, *SE*=3.357), *t*(275)=-1.318, *p*=.189, *r*=.079. They did, however, have (*M*=3.912, SE=0.561) a more positive attitude towards mathematics (*M*=3.609, *SE*=0.581), *t*(177)=2.001, *p*<.05, *r*=.149, only a small to medium effect. Comparison of symbol sense pre- and post-scores reveals that the symbol sense score on the posttest (*M*=1.462, *SE*=1.504) is significantly higher than the pretest score (*M*=-1.493, *SE*=2.339), *t*(285)=-20.602, *p*<.001, *r*=.773. According to Cohen's benchmark (1992) this is a large effect.

To investigate differences between the different types of items, Table 4 shows the fraction of correct answers for all students, all parts of the digital module and for each of the four item categories described in section 2. It shows that the improvement between pre- and posttest is significant across all four categories. The improvement is greatest for the second category; the third category shows the least improvement.

Table 4: Fraction of items correct per category (1)

Category	Pre	d1	d2	d3	d4	d5 (2)	d6	Post
Cat1	0.532	0.929	-	-	-	0.728	0.630	0.872
Cat2	0.248	-	0.938	-	-	0.841	0.776	0.825
Cat3	0.566	-	-	0.950	-	0.794	0.627	0.755
Cat4	0.571	-	-	-	0.853	0.759	0.656	0.872
Control	-	-	-	-	-	-	-	0.629
Total	0.507	0.929	0.938	0.950	0.853	0.780	0.672	0.787

(1) Items that were not attempted were not included in the 'fraction correct'.
(2) Also note that the final d5 score includes point reductions for attempts. Therefore the final score for d5 is not the same as the 'fraction correct'. For example, with deductions, the score is 49 out of 100 points. This is much less than the 0.780 tabulated above. This discrepancy is not applicable to d6, as there were no point deductions.
(3) The asterisks behind the values in the post column denote the significance of the difference between pre- and posttest scores. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
(4) Two new control items, not within one of the four categories, were added to the posttest to see whether students could solve these.

Multilevel model

Table 5 shows the development of the model from the null model to the final model. Each of the models is addressed in the sub-sections below. The models with *a* added are initial models with all predictors. The models with *b* added are the final models for that iteration excluding non-significant predictors.

Null model

Comparison of model-fit of the one- and two-level *intercept only models* revealed significant variation between classes. The between-class variance is estimated as 65.150, and the variance between students within classes is estimated as 172.197. The value of the likelihood ratio statistic, obtained from the two models' loglikelihoods, is 63.383, which is compared to a chi-squared distribution ($df = 1$). We conclude that there is significant variation between the classes. The intraclass correlation then is .27 which means that 27% of variance is explained by group differences and 73% by student differences.

Model 1: preknowledge model

In the first model we added predictors for pre-knowledge. The variable trendexam ($\chi^2 = 0.569$, $df = 1$, $p > 0.05$) did not contribute significantly and was excluded from the model. The variable pre-test ($\chi^2 = 109.462$, $df = 1$, $p < 0.001$) showed significant effects. Compared to the null model, the inclusion of pre-knowledge resulted in a significant model improvement ($\chi^2 = 109.462$, $df = 1$, $p < 0.001$).

Model 2: timespent model

In the second model we added predictors for time spent in the tool. The variable time_total_hr ($\chi^2 = 20.552$, $df = 1$, $p<0.001$) showed a significant contribution to the model. As this variables consists of separate time components we also added separate time components, respectively time_manual_hr ($\chi^2 = 2.512$, $df = 1$, $p>0.05$), time_d14_hr ($\chi^2 = 2.957$, $df = 1$, $p>0.05$), time_d5_hr ($\chi^2 = 18.142$, $df = 1$, $p<0.001$) and time_d6_hr ($\chi^2 = 5.842$, $df = 1$, $p<0.05$). Thus, only time_d5_hr and time_d6_hr improved the model and showed a significant contribution. The other two were subsequently removed, leaving only time_d5_hr and time_d6_hr to improve the model significantly ($\chi^2 = 20.309$, $df = 1$, $p<0.001$).

Model 3: tool model

In the third model we added predictors concerning the tool. None of the predictors vs_phome (B=0.057, SE=0.042, $p>0.05$) and vs_general (B=1.031, SE=0.769, $p>0.05$), however, contributed significantly to the model, so they were excluded. For this reason, model 3 is not mentioned in Table 5.

Model 4: student attitude model

In the fourth model we added predictors for student attitude, respectively MATH (B=6.301, SE=2.276, p<.01), ICT (B=-1.943, SE=1.539, p>0.05), REFCOM (B=1.241, SE=2.204, p>0.05), PURINV (B=1.244, SE=2.829, p>0.05) and vs_motivation (B=-2.776, SE=1.596, p<.05). Thus, only MATH and vs_motivation seemed to improve the model significantly. However, when both were added to the model, vs_motivation fell under the threshold (B=-1.379, SE=1.288, p>0.05) and was subsequently removed, leaving only MATH to improve the model significantly ($\chi^2 = 10.202$, $df = 1$, $p<0.001$).

Model 5: random slopes

Testing for random slopes indicated significant variance in the regression slope of the predictor pre-test across classes ($\chi^2 = 12.565$, $df = 1$, $p < 0.001$). The covariance -1.233 (standard error 0.500) shows that there is negative relation between intercept and slope, i.e. the higher the intercept, the less steep the slope. This again strengthens the findings that there are significant differences between the classes. Furthermore, a high average score for a class is accompanied by a gentle slope, and a low average by a steep slope.

These results imply that students with an average pre-test score, who spend an average amount of time in d5, and have an average general attitude towards mathematics, score 79.528 (SE=1.994) out of a possible 100 on the final test. For every point (out of 100) higher on the pre-test score, the post-test score also increases 0.248 (SE=0.049, p<.001). Furthermore, for every hour spent in part d5 an additional 1.189 points (SE=0.539, p<.05) are obtained. Finally, test scores are 5.257 points ($SE = 1.414$, $p < .001$) higher when general attitude towards mathematics (MATH) increases by 1 point on a 5-point scale, equivalent to a 21.14 point difference in test scores between the least and the most positive students on a 100-point final test. Taken together, these attitudes could potentially account for significant point differences in test scores between individuals.

	Null model	Model 1a	Model 1b	Model 2a	Model 2b	Model 4 (1)	Model 5
Fixed							
Intercept	78.893 (2.230)***	79.023 (2.046) ***	79.023 (2.046) ***	79.069 (1.960) ***	79.069 (1.960) ***	79.014 (1.941) ***	79.528 (1.994) ***
(pre-test-gm)		0.290 (0.038)***	0.290 (0.037)***	0.308 (0.037) ***	0.308 (0.037) ***	0.271 (0.038) ***	0.248 (0.049) ***
(trendexam-gm)		-0.969 (10.878)	-----	-----	-----	-----	-----
(time_manual_hr-gm)				1.389 (1.246)	-----	-----	-----
(time_d14_hr-gm)				-0.264 (0.314)	-----	-----	-----
(time_d5_hr-gm)				1.931 (0.515) ***	1.931 (0.515) ***	1.189 (0.539) *	1.189 (0.539)*
(time_d6_hr-gm)				7.814 (4.986) *	7.143 (4.482) *	7.321 (4.757) *	7.282 (4.631) *
(MATH-gm)						5.034 (1.562) **	5.257 (1.414)***
Random							
Level 2 class	65.150 (27.235)**	55.996(22.857) **	55.040 (22.909) **	48.310 (20.965)**	50.538 (21.093)**	49.794 (20.718) *	50.270 (11.999)***
(pre-test-gm)/cons							-1.233 (0.500)*
(pre-test-gm)/(pre-test-gm)							0.027 (0.014) *
Level 1 student	172.197 (14.631)***	138.676 (11.914) ***	138.674 (11.913) ***	130.317 (11.195) ***	129.258 (11.104) ***	124.603 (10.704) ***	120.113 (10.342) ***
Model fit							
Deviance (-2 loglikelihood)	2363.508	2254.038	2254.046	2235.904	2233.737	2223.535	2203.204
χ^2		109.470	109.462	18.142	20.309	10.202	20.331
Df		2	1	4	2	1	2
P		p<0.001	p<0.001	p<0.01	p<0.001	p<0.001	p<0.001
Reference		M0	M0	M1b	M1b	M2b	M4
N	286	286	286	286	286	286	286

Table 5: The multilevel model

Values between brackets are the standard errors. 'gm' refers to Grand Mean centering.
(1) Only the final model 4 was added to the table. $^{*}p < 0.05$., $^{**}p < 0.01$., $^{***}p < 0.001$.

5 Conclusions and discussion

We set out to answer two questions: (1) What is the effect of an intervention on the development of algebraic expertise, including both procedural skills and symbol sense? and (2) Which factors predict the resulting algebraic performance?

The answer to the first research question is that use of the intervention for an average of five hours has a large effect on improving algebraic expertise. This means that there is not only an improvement in score, but also an improvement in recognizing patterns and having a sense for symbols. A vast improvement is apparent in all four categories of tasks, with the category inspired by Wenger (1987) showing the greatest progress.

The answer to the second research question is that previous knowledge, time spent in self-test and summative test mode, and general attitude towards mathematics are the largest predictors for a high posttest outcome. The fact that the time spent in the self-test mode (d5) and digital summative test (d6) are more significant explanatory variables than the practice mode (d1-d4) suggests the design strategy to decrease the amount of feedback when moving towards summative assessment makes sense. Without having implemented d5 and d6 two influential parts of the intervention would be missed. The other two significant explanatory variables, pretest score and attitude towards mathematics, seem less ground-breaking. It was not clear-cut beforehand that these two variables would also impact the outcome when using an online tool for algebraic expertise. The fact that there are indeed no significant predictors for the posttest score that have to do with attitude towards ICT suggests that conventional pen-and-paper techniques and ICT techniques are reconciled (Kieran & Drijvers, 2006). In other words: contrary to earlier research (e.g. Tondeur et al., 2008), in this intervention ICT and mathematics seem to be integrated in such a way that the student's attitude towards ICT in itself does not influence the outcome. Compared to only a pen-and-paper approach the ICT adds advantages such as being able to learn anytime anyplace, receiving feedback and randomizing items. In contrast to this, variables such as overall quality of the group (operationalized by trend exam grades), gender, total practice time and whether teachers and students worked more at home or at school, did not significantly predict the outcome. We contend that this signifies that the success of the intervention is predominantly independent of many of these variables that differed between classes. For example, the percentage of time that students spent using the tool in class or at home did not influence the outcome. Some teachers spent almost all their time with their class in the computer lab; others let the students work both at home and at school. Another class only used the module at home; they had a much smaller gain from it. Because this group was quite small, we cannot draw strong conclusions from this, but the interpretation that paying no attention to the module and just having students work at home is less effective than providing a mix of home and class work does seem to make sense. In addition, attitudes towards ICT tools for mathematics, negative or positive did not significantly influence the outcome. It is important to note that higher average scores for a class go hand in hand with less added value for the student. The same holds for higher pretest scores. In other words: the more skilled a student is at the start, the less beneficial this digital intervention seems. This is in line with research on the expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003), which holds that instructional methods that are highly effective with novice learners can lose their effectiveness and even have negative effects when used with more experienced learners. This was also confirmed by the analysis of the students who scored worse on the posttest than the pretest.

The conclusion that the intervention ‘works’ invites some recommendations for the use of technology for algebraic expertise in mathematics education. First, students who have already acquired good algebraic expertise should not be forced to 'practice till they drop'. Second, the module’s online availability and home access should allow for differentiation: students who need practice can practice as much as they want (due to item randomization) and students who do not need practice can show this, as results are stored. This differentiation approach implies that not all the work within the online module is done within the confinement of the classroom’s four walls. Use of the tool at school can induce whole class discussions and preparatory instructions for individual use of the tool. An emphasis on work at home might even be more effective than an emphasis on use in school. Practically this means that teachers can differentiate in level of expertise between students. The online tool makes differentiation easier, as students can use it at their own pace, anytime and anyplace. Teachers can scrutinize the results and use one-to-one communication for individual feedback, for example through social media. Third, design principles concerning formative scenarios and crises seem to work. This implies that causing intentional crises by offering non-standard test items is a fruitful approach in item design. Students should not be served standard questions but also be challenged by non-standard questions. The crises they cause may evoke learning. Also, the idea that students have to 'stand on their own two feet' should be kept in mind when designing interventions. Therefore, technological tools that are implemented in the curriculum should take into account that the use of these tools prepares students for final examinations, even if the tool use is not normally allowed during examination sessions. As we conjecture these implications hold beyond just the mathematical domain, teachers and designers alike should be aware of these didactical implications.

We are aware of the study’s limitations. First, we have used a relatively small population for multilevel analysis. The reasons for this, and for continuing anyway, were addressed in the methodology section. Second, there is still quite a large percentage of variance that remains unexplained. Third, a note should be made about all-too-easy assumptions of causality. The experimental setup did not include a control group, because the aim was not to prove whether an ICT intervention is better than another approach, but to show that an online intervention based on specific design principles can foster algebraic expertise. The focus on explanatory factors reveals which variables influence outcome and which ones do not. A comparative study on a larger scale is advisable.

There are other challenges for studies on the use of web-based interventions at home. Clearly, it is expected that the availability of the online module at home is a positive thing. Methodologically speaking, however, the lack of experimental control on the home situation poses a problem. Herein we discern a paradox: how can we control a situation that essentially is uncontrolled? The absence of control, in contrast to a school situation, might even be the strength of web-based interventions. Fourth and finally, although the study informs discussion and debate on the dichotomy of conceptual understanding versus skills, the question remains whether the designed intervention affects both of these aspects. Perhaps these students are 'only' conditioned to follow - higher level - algorithms and recipes? We contend that recognizing patterns is a different level to expanding expressions routinely. The added value lies in making more explicit the algebraic expertise involved, and thus in demystifying the insight that experts have acquired.

In spite of these limitations, the intervention's effects are manifest. Still, further replicative as well as explorative research is needed. However, this should not withhold us from being moderately optimistic or -as one student remarked understatedly: "seeing that at first I could not solve the equation without fault and now I can, I must conclude that my expertise has improved."

Appendix A

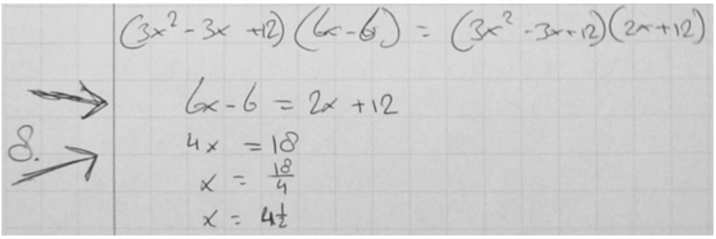
This appendix provides examples of the data that were collected.

<table>
<tr><th>#</th><th>Data source</th><th>Example item</th></tr>
<tr><td>1</td><td>Pre- and posttest results (scores out of 100 points).</td><td>Example item:
Solve the equation:
$(x^2 - 7x + 12) \cdot (8x - 11) = (x^2 - 7x + 12) \cdot (3x + 14)$
Example fragment of results:
<table>
<tr><td>maxpoints</td><td>5</td><td>3</td><td>10</td><td>6</td><td>5</td><td>5</td><td>3</td><td>6</td><td>43</td><td>100,0</td></tr>
<tr><td>id</td><td>pre1</td><td>pre2</td><td>pre3</td><td>pre4</td><td>pre5</td><td>pre6</td><td>pre7</td><td>pre8</td><td>totaal</td><td></td></tr>
<tr><td>Student</td><td>5</td><td>3</td><td>9</td><td>6</td><td>4</td><td>5</td><td>3</td><td>6</td><td>41</td><td>95,3</td></tr>
</table>
The first line denotes maximum points to be obtained. The cells behind ‘Student’ denote scores, with green ‘+1’, white ‘0’ and orange ‘-1’ (see appendix B). The final cell in the row is the % scored.
See appendix B for examples of student work.</td></tr>
<tr><td>2</td><td>Scores and log files of the digital module.</td><td>Fragment of a log-file:
log scores | log answers | log errors | log attempts count | log attempts | log data
<table>
<tr><th>ami_1_10</th><th>ami_1_11</th><th>ami_1_12</th><th>ami_1_13</th><th>ami_1_2</th><th>ami_1_3</th><th>ami_1_4</th><th>ami_1_5</th><th>ami_1_6</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td>0</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>5</td><td>3</td><td></td><td></td><td>34</td><td>12</td><td>4</td><td>4</td><td>8</td></tr>
<tr><td>3</td><td>4</td><td>0</td><td>0</td><td>3</td><td>6</td><td>2</td><td>2</td><td>8</td></tr>
<tr><td></td><td></td><td></td><td></td><td>41</td><td>2</td><td></td><td></td><td></td></tr>
<tr><td>5</td><td>5</td><td>0</td><td>0</td><td>78</td><td>38</td><td>18</td><td>19</td><td>32</td></tr>
<tr><td>2</td><td>12</td><td>0</td><td>0</td><td>16</td><td>11</td><td>10</td><td>2</td><td>2</td></tr>
</table>
Note that the layout of this table was slightly changed to improve the presentation. These files were processed to aggregate totals.</td></tr>
<tr><td>3</td><td>Pre questionnaire for students (Reed et al., 2010). (a)</td><td>General attitude towards mathematics: “I enjoy doing mathematics’”
Attitude towards using computers for mathematics: “I like to use computers in mathematics’”
Purposeful and investigative behaviors: “I like to do self-tests to see if I really understand mathematics’”
Reflective and communicative behaviors: “I like to share solutions with other students”</td></tr>
<tr><td>4</td><td>Post questionnaires for both students and teachers.</td><td>Example question:
How do you judge your general experiences with this module? (1=negative/2=slightly negative/3=neutral/4=slightly positive/5=positive)</td></tr>
</table>

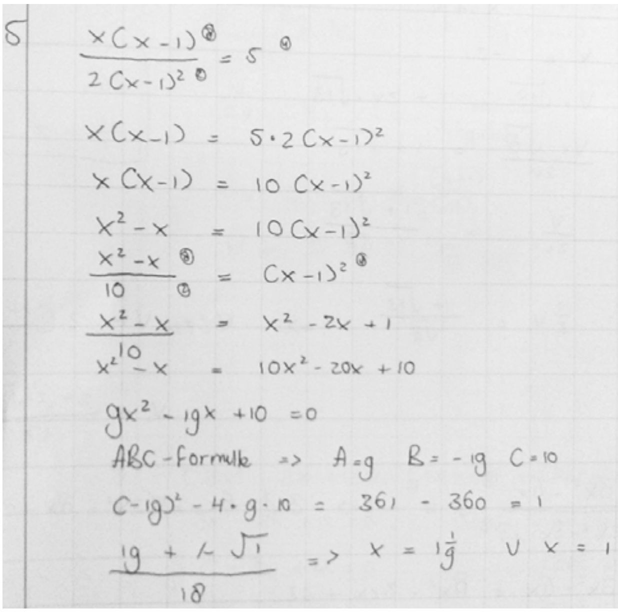
Note. English-language items were translated to Dutch for this study.
(a) These items were measured using 5-point rating scales, with the range of answers from „strongly disagree‟ to „strongly agree‟

Appendix B

This appendix provides some examples of students' symbol-sense (1) and non-symbol-sense (-1) solutions.



In the case above the student recognized common terms on the left and right hand side of the equation. Therefore we scored this solution as '1'. Of course, the student loses solutions, which results in a score reduction.



In this example the equation's left hand side already has a factored form. Nevertheless, the student chose to work out the parentheses. This solution was valued as '-1': the student did not show Gestalt (pattern salience) in this solution step. Note that solving this equation correctly should not be valued as '1', as the equation is already in the 'correct' format. In our valuation this case would score a '0'. Note that apart from these scores for symbol sense students also obtained a score for correctness. In this case the student did receive points, only missing out on those given for solutions that were not allowed.

Chapter 7

Conclusion

Conclusion

1 Conclusions

In this chapter we summarize the findings from the previous chapters and the study as a whole. The general question in this study is:

In what way can the use of ICT support acquiring, practicing and assessing algebraic expertise?

This general question led to several sub-questions, each addressed in an appropriate cycle in the study:

1a) Which criteria are relevant for the evaluation of digital tools for algebra education?
1b) Which digital algebra tool best meets these criteria?
2a) Do the concepts of symbol sense, gestalt view and pattern and local visual salience, described in a pre-digital era, help us in understanding what students do in a digital environment?
2b) Can the feedback design of a digital tool be improved with students?
2c) Which methodology can be used to elaborate feedback desired by students?
3a) Timing and fading: are there indications that formative scenarios improve the acquisition of algebraic expertise?
3b) Crises: do crises in algebraic tasks improve the acquisition of algebraic expertise?
3c) Variation: does variation in feedback influence scores and student behavior?
4a) What is the effect of an intervention on the development of algebraic expertise, including both procedural skills and symbol sense?
4b) Which factors predict the resulting algebraic performance?

The previous chapters led to the following answers to these sub-questions. Chapter 2 concerns the identification of criteria that are relevant for the evaluation of digital tools for algebra education. We constructed an evaluation instrument consisting of 27 criteria grouped in four categories. The categories are based on a conceptual framework that matched the goals and intentions of the study, and consist of key aspects from theories on tool use, algebra didactics (algebraic expertise), and assessment (feedback and formative assessment). A fourth category concerns general tool characteristics. The modified Delphi approach, conducted to validate the criteria, revealed a large agreement among external experts on these criteria. The construction of criteria weights led to the identification of the most important criteria: stability and performance, correct display of mathematical formulas, ease of use, mathematical soundness, and ability to store the work. We conclude that the designed list of criteria provides a good evaluation instrument for describing characteristics of digital tools for algebra that we consider relevant for the purpose of the study. The instrument provides insight into the different features of various tools. It could also be very helpful for software development in mathematics education, and particularly useful in algebra education. Furthermore, using the evaluation instrument, we rated seven tools that met the minimal criteria and had our codes validated by an external expert scoring. This resulted in the Digital Mathematics Environment (DME) having the highest overall score and the conclusion that it would be best suited for addressing the research goals on algebraic expertise. Key features of DME include enabling stepwise problem solving strategies, being easy to use, and storing the solution process of the student; additionally, it is well suited for formative assessment, as it offers several strategy modes, feedback and self-review.

Chapter 3 addresses whether the notions of symbol sense, gestalt view and visual salience, described in a pre-digital era, could help us in understanding what students do in a digital environment. The design process of digital content and the one-to-one pilot sessions suggest that these concepts remain extremely relevant when deploying digital activities. The observations show that students using a digital environment exhibit some symbol sense behaviors but also exhibit behaviors lacking symbol sense. The notions of gestalt view and visual salience are helpful in analyzing student work. Although students worked in a digital environment, the results were in line with past findings in traditional pen-and-paper settings (Arcavi, 1994; Wenger, 1987). While solving algebraic tasks in the digital environment, the students could use any strategy, and thus show sensitive towards gestalt and visual salience aspects, and further develop such sensibility. The DME seems to facilitate this development through its mathematical interface and feedback opportunities, which is more difficult to offer in a paper-and-pencil environment. The exemplary tasks also point out that observing symbol sense is not a straightforward affair. It often is quite hard to recognize whether students rely on standard algebraic procedures or are actually showing insight into the equation of expression, in line with the gestalt view or visual salience notion.

In chapter 4 we focus on the implementation of feedback in the DME. We conclude that the feedback design of a digital intervention can be improved through student inquiry. A fruitful methodology we utilized to elaborate suitable feedback in the digital intervention is: (1) deploy the intervention and ask students to ‘think aloud’; (2) tabulate all the responses; (3) make a feedback design; (4) implement the feedback design in the digital environment; (5) test the environment with help of the feedback design, and make another iteration.

In chapter 5 we describe three feedback design principles: timing and fading with formative scenarios, crises and variation in feedback. Regarding the first feedback principle, timing and fading with formative scenarios, the score progress during the intervention shows that an initial pre-knowledge score was followed by a large improvement of scores for algebraic expertise when in practice mode. After this, in self-test mode, the scores drop markedly, and in the final digital test they drop even further. Overall, however, the intervention shows a significant improvement across all four task categories identified in chapter 3. The largest improvement is apparent in the second category; the third category shows the least improvement. Although the formative scenario causes 'pain' in gradually fading feedback in the course of the intervention, the learning effect remains. The design principle of using timing and fading with formative scenarios seems to improve the acquisition of algebraic expertise. For the second feedback principle, crises, the results are mixed. When looking at the fraction of items answered correctly there seems to be no effect for a crisis task. This is probably caused by the fact that difficulty with tasks cannot be seen if the final score is high, although achieved with difficulty. The number of attempts needed after a crisis task, however, decreases in three of four series of tasks. In categories one, two and four it seems the crisis is followed by a drop in number of attempts and an increase in score. For category three this is not the case. Perhaps this is caused by the use of root signs and logarithms in the tasks. Whether the design principle improves student's expertise remains unclear from the results. The third design principle concerns feedback variation. The students´ scores in the two conditions reveal a medium effect for the feedback condition including self-regulatory feedback (condition c2). This effect, however, only is apparent in one of the series of tasks. We explain the difference between d1 and d3 by the fact that both series address the category of polynomial equations (Bokhove & Drijvers, 2010). Having solved polynomial equations with feedback in d1 meant that the students were already capable of solving similar types of equations (in d3, which followed after d1), and, subsequently, no additional form of self-regulation was needed in d3. The addition of buttons for feedback also has an, albeit small, effect on the number of attempts. This makes sense as the additional feedback that can be requested discourages more attempts. When looking in more detail at the use of the feedback in the three case examples, we conclude that both task-related and self-regulatory (FT and FR) feedback can be used in a formative way for the learning of algebra. Students can benefit from the design principle of feedback variation, to overcome difficulties and to check whether or not they are on the correct solution path. Taking the three feedback principles together it is clear that the use of feedback in an online algebra intervention can have a significant effect. Furthermore, the three feedback design principles provide a broad approach to the potential of feedback use in digital environments.

In chapter 6 we consider the overall effect of the intervention. The use of the intervention for an average of five hours has a large effect on improving the students´ algebraic expertise. Students not only show an improvement in score, but also an improvement in recognizing patterns and in their sense for symbols. A significant improvement is apparent in all four categories of questions, with the category based on Wenger (1987) showing the biggest improvement. A multilevel analysis reveals that previous knowledge, time spent in self-test and summative test mode, and general attitude towards mathematics are the largest predictors for a high posttest outcome. The fact that no significant predictors that have to do with ICT are found may indicate that indeed conventional pen-and-paper techniques and ICT techniques are reconciled (Kieran & Drijvers, 2006), with the digital environment adding advantages such as providing feedback and randomized tasks. Variables such as overall quality of the school (operationalized by trend exam grades), total practice time and home or school use of the digital environment do not significantly predict the outcome. This suggests that the intervention´s success is predominantly independent of many of these variables. The role of the teacher is only a moderate factor, whereby the way in which teachers had to use the environment was only loosely prescribed. Some teachers spent almost all their time with their class in the computer lab, whereas others let the students both work at home and at school. One class only worked the intervention at home; this resulted in a much smaller gain. Because the N is quite small, we can't infer strong conclusions from this, but it is logical to conclude that paying no attention to the intervention and letting students work at home is less effective than providing a mix of home and class work. It is important to note that higher average scores of a school go hand in hand in hand with a lesser degree of added value for the student. The same holds for higher pre-test scores. In other words: the smarter a student is, the less beneficial this digital intervention seems. This is in line with research on the expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003), which holds that instructional methods that are highly effective with novice learners can lose their effectiveness and even have negative effects when used with more experienced learners.

To synthesize these findings into a general conclusion we return to the main research question. We provide a short conclusion for every part.

In what way can the use of ICT support acquiring, practicing and assessing algebraic expertise?

In what way. There are many ways to design an environment. This study shows that a framework based on formative assessment, with feedback design as guiding principle, can be a successful approach.
Use of ICT. We have seen that we can use ICT, an online environment in particular, for learning algebra. We have to keep in mind the tool's affordances and constraints, and the purpose for which it is deployed. Furthermore, because of instrumentation, student characteristics must be taken into account. In our study, on the one hand there was no gender difference in effect. Students' pre-knowledge, on the other hand, did make a difference: the better students scored on the pre-test, the less gain they had.
Acquiring. In developing algebraic expertise, we have found that the use of intentional crises seems to aid the transfer of algebraic expertise.
Practicing. A didactical scenario that enables students to practice in the beginning stages of an intervention, and using timing and fading, ensures that students learn to 'stand on

their own feet', and to be independent of the practice tool.
Assessing. A didactical scenario also needs room for summative testing. Combining formative, summative assessment and feedback in one scenario works.
Algebraic expertise. Summative assessment shows that the intervention has a large effect on both sides of the algebraic expertise dimension, basic skills and symbol sense.

What does this study add to the existing vast body of research knowledge? For this we return to the three key theoretical foundations of the study: ICT tool use, algebraic expertise, and formative assessment. On the first topic of tool use we have noticed that it mainly played a part in the preliminary cycle of our study: by explicitly adding characteristics concerning tool use, i.e. ease of use, to the evaluation instrument we ensured that the tool we chose would be 'best' for our purpose. The second topic concerns algebraic expertise. The dimension with basic skills and symbol sense proved to be useful for identifying item categories and interpreting student work. Both the extension of theory with pattern salience and the four item categories we identified contributed to the operationalization of the notion of symbol sense in relation to digital activities. Notions of formative assessment, the third perspective, helped us in designing the intervention, and the three feedback design principles, in particular. These three feedback principles can also be seen as an operationalization of formative assessment in the realm of feedback design in online activities.

2 Recommendations for designers and teachers

Based on the conclusions in the previous section we can phrase the following recommendations for designing and using digital environments for algebra education. The recommendations are for both designers of digital environments and content, as well as for teachers who want to use them in their teaching. As the distinction between a designer and a teacher in this context greatly depends on the software in use, only certain software come with an authoring environment, for example, it is hard to categorize the recommendations easily into 'for designers' and 'for teachers'. An additional aim for these recommendations is to bridge the gap between theory and practice, as stated in the preface of the thesis.

Use student and teacher feedback to improve the design of digital activities.
Often feedback is designed within the context of a research or designer setting. Mostly researchers hypothesize that their well-thought-of interventions also work in practice. To increase the chance that this is actually the case, it is sensible to use the domain knowledge of practitioners. Environments that provide feedback can be improved significantly if we involve the actors that participate in the knowledge domain. Therefore, we recommend involving students and teachers early in the design process.

Make full use of the possibility of feedback generated by a digital environment.
It is clear that feedback that is generated by a digital environment can aid students when they are working. Feedback can point out whether a solution is correct, point out mistakes, and give instructions and keep students on the right track of learning. This means that teachers should be aware of the advantages of feedback and of the types of feedback that are available in digital environments. Both designers and teachers should take into account the 'good practices' for feedback.

Context: choose an appropriate tool and make sure you know what it can and cannot do for you. No digital tool is universal or perfect. What we set out to do in this study, and in Chapter 2 in particular, is to provide a framework with which one can evaluate algebra tools so that, depending on the context, its affordances and constraints can be determined. By being aware of this, the best tool for a certain task can be selected, and this choice can be based on sound arguments. Furthermore, it enables us to compare contexts, analogous to design research where the main research question is based on the format: "What are the characteristics of an intervention A for the purpose/outcome B in context C?". Therefore, we recommend using Chapter 2's evaluation framework when choosing digital tools.

Embed activities in pedagogical and didactical scenarios that facilitate transfer.
Do not do digital activities just for the fun of them. Integrate and embed these activities in a broader curriculum: they should not be a goal in themselves but a means towards a goal. One successful way of embedding them is in a formative scenario with feedback timing and fading: a scenario based on formative assessment, starting off with activities with elaborate feedback, and then fading towards activities without feedback.

Algebraic expertise: do not bet on just skills or just insight, as both are closely interlinked.
In discussions, participants often put too much emphasis on either insight or skills. We recommend acknowledging the importance of both, and stop wasting time and energy in discussing a prevalent viewpoint. Teachers in particular should pay attention to both basic skills and symbol sense by providing appropriate activities that cater to both.

Use 'unusual' tasks to improve learning, for example by intentionally causing a crisis.
We learn a lot when we are confronted with cognitive conflicts. As many books provide students with 'safe' tasks, we should allow for 'unusual' or crisis tasks. In doing so, however, we should provide a 'safety net' of extra instruction and feedback for students, which can be implemented in (online) digital activities.

See the use of online environments as an incentive to rethink home and classroom work.
The results also imply that not all the work with the digital environment is done within the confinement of four walls. An emphasis on work at home can even be more effective than

it used to be, with time at school meant for class discussions, instruction and starting up individual use of the digital environment.

Differentiate, because good students may benefit less.
Students who have already acquired a good algebraic expertise should not be forced to 'practice till they drop'. The fact that an environment is used online and at home can facilitate differentiation: students who *need* practice *can* practice as much as they want (another advantage of randomization), students who do *not need* practice *can* show this, as results and work are all stored.

Results-at-a-glance

The main research question is: in what way can the use of ICT support acquiring, practicing and assessing algebraic expertise?

In what way? There are many ways to design an environment. This study shows that a framework based on formative assessment, with feedback design as guiding principle, can be a successful approach.

Use of ICT. We have seen that we can use ICT, an online environment in particular, for learning algebra. We have to keep in mind the tool's affordances and constraints, and the purpose for which it is deployed. Furthermore, because of instrumentation, student characteristics must be taken into account. In our study, on the one hand there was no gender difference in effect. Students' pre-knowledge, on the other hand, did make a difference: the better students scored on the pre-test, the less gain they had.

Acquiring. In developing algebraic expertise, we have found that the use of intentional crises seems to aid the transfer of algebraic expertise.

Practicing. A didactical scenario that enables students to practice in the beginning stages of an intervention, and using timing and fading, ensures that students learn to 'stand on their own feet', and to be independent of the practice tool.

Assessing. A didactical scenario also needs room for summative testing. Combining formative, summative assessment and feedback in one scenario works.

Algebraic expertise. Summative assessment shows that the intervention has a large effect on both sides of the algebraic expertise dimension, basic skills and symbol sense.

Recommendations-at-a-glance

- Use student and teacher feedback to improve the design of digital activities.
- Make full use of the possibility of feedback generated by a digital environment.
- Context: choose an appropriate tool and make sure you know what it can and cannot do for you.
- Embed activities in pedagogical and didactical scenarios that facilitate transfer.
- Algebraic expertise: do not bet on just skills or just insight, as both are closely interlinked.
- Use 'unusual' tasks to improve learning, for example by intentionally causing a crisis.
- See the use of online environments as an incentive to rethink home and classroom work.
- Differentiate, because good students may benefit less.

Fig. 1: Results and recommendations at a glance

3 Discussion points from a personal stance

In this paragraph we would like to touch on some personal discussion points that concern the study.

In the course of this study we have had numerous discussions about the nature and essence of insight. When do we exactly have insight in mathematical concepts? From the literature that has appeared on this topic, and that is described in the previous chapters, one thing became clear to us: skills and insight go hand in hand. Therefore we acknowledge that training of certain equations primes the structure of these equations. This does not imply that there is no conceptual development. As a matter of fact, we would contend that insight is almost always strengthened by practice, and vice versa. One experience we remember concerns the initial lack of understanding of the epsilon-delta definition of limit. It seemed to be a 'trick' when one ended up with ε^2 and then repeated the process with a delta being the square root of epsilon. It was only after several proofs, and accompanying graphs, that understanding actually set in. In this case this occurred even after passing an exam on epsilon-delta limits. This is why the theoretical framework on algebraic expertise, and notions such as symbol sense and salience, was extended with pattern salience (see chapter 3). We wrote: "extending the concept of visual salience to patterns provided by standard routines students already know could perhaps relieve the tension between the application of standard routines and succumbing to salient patterns." Thus the extended theoretical framework serves as a bridge between skills, understanding, structures and patterns. To be sure that these are flexible we used the principle of intentional crises through non-standard tasks. Still, we acknowledge that the items and the sequences of items we constructed only cover a very limited part of the algebra curriculum. Claiming that algebra performance as a whole has improved, would be too bold a statement.

As a consequence, a second point concerns the observation that we have only touched on a small subset of the vast body of algebraic knowledge. The focus has only been on a limited part of four categories of tasks concerning flexible manipulation skills. We are aware that it will be hard to generalize the findings to other mathematical topics. However, in designing our intervention we did make use of general principles, for example regarding feedback design. Therefore, we do conjecture that many of the findings can be applied to other interventions for other knowledge domains.

There are other challenges for studies on the use of web-based interventions. Clearly, it is expected that the availability of the online module at home is positive. Methodologically speaking, however, the lack of experimental control of the home situation poses a problem. Herein we discern a paradox: how can we control a situation that essentially is uncontrolled? The absence of control, in contrast to a school situation, might even be the strength of web-based interventions. During the study we have undertaken some informal polls to see what the research field thinks about this. The results were ambivalent, to say the least. In general, there were two types of viewpoints. The first recommends a controlled environment at school and seems to assume that effects that are proved this way will also hold for work done at home. The second viewpoint acknowledges the fact that distance learning is hard to research, but recommends using instruments such as log files and student self-reports. We think that ideally these different viewpoints should be aligned.

Finally, we should watch for what we call didactics of technological shortcomings: features of software that are greatly missed, but are trivialized by some sort of didactical explanation. Of course, these explanations can be valid, but this should not imply we do not take notice of glaring shortcomings of the software. Two examples may illustrate this point. First, there are algebra tools that cater for multi-step solutions and tools that do not. Although we would not want to discount advantages of tools that only check for the final solution, we do contend that the ability to input in-between steps is an advantage, as it adheres to the glass box principle (Beeson, 1998). However, as many tools do not have this feature, this is sometimes played down by stating that a correct solution can only be found if the solution path is correct. We would beg to differ. Furthermore, looking at the reverse situation, an incorrect answer does not always imply a wrong solution path. Taking into account the road to the answer is a pedagogical point that is widely accepted in educational practice. A second example concerns the way in which students have to input a solution. Although curricular demands could lead to statement like ‘students should be able to use Maple syntax’, mathematically speaking there should be cognitive fidelity (Dick, 2007) between the way in which students input algebraic expressions and these expressions' semantic meaning. For secondary education this would imply that a graphical ‘pretty print’ input editor is more suitable than a linear one, which requires learning syntax. Of course, students will be able to learn that x^2 is the same as x^2, but with current technology students can easily input the first. We propose to stop justifying the use of linear syntax input when syntax is not a learning objective. This also touches on the difference between ‘use to learn’ and ‘learn to use’. Both examples are cases of software limitations that are defended by didactical explanations. This is no disqualification of software, because it provides insight in the way that this particular piece of software can be deployed in the classroom. We should, however, not accept these shortcomings and aim for improvement. Didactics of technological shortcomings should not be used to justify a status quo.

4 Further research

The study set out to design an online intervention for algebraic expertise, and to investigate the transfer from digital environment towards pen-and-paper. It is clear that there are still many features of digital tools that can improve. One factor to take into account concerns recent software and hardware developments.

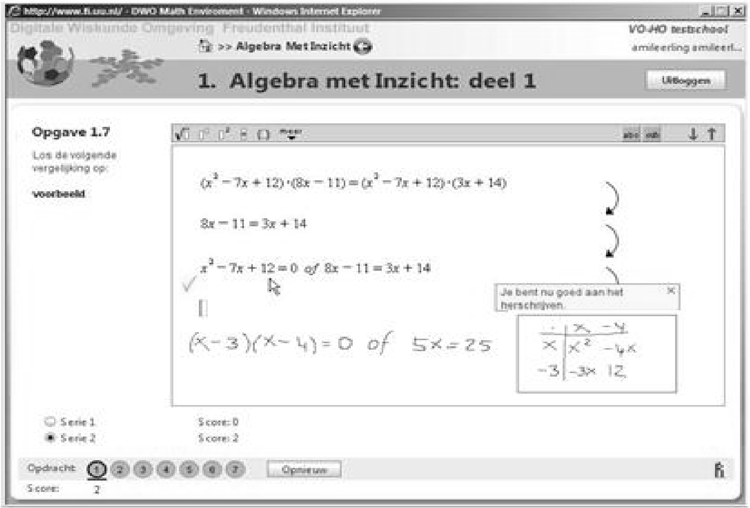


Fig. 2: Combining algebra tools and tablet technology: the future?

Concerning software, many new programs or new releases of existing software have appeared. Some will perhaps offer new opportunities for algebra education. Hardware has also evolved. The appearance of tablet computers, some even with formula recognition for mathematics, has a potential for further integrating digital and traditional pen-and-paper environments.

Figure 2 shows a still from a movie mock-up of the combination of an algebra tool and tablet technology; the full animation can be found on www.algebrametinzicht.nl. Whatever the future technological developments, one thing is clear: that theory and practice should go hand in hand. This means that educational research has to proceed to see what works, why it works and how digital tools can be improved.

References

Arcavi, A. (1994). Symbol Sense: Informal Sense-Making in Formal Mathematics. *For the Learning of Mathematics*, *14*(3), 24-35.

Arcavi, A. (2005). Developing and using symbol sense in mathematics. *For the Learning of Mathematics*, *25*(2), 42–47.

Artigue, M. (2002). Learning Mathematics in a CAS Environment: The Genesis of a Reflection about Instrumentation and the Dialectics between Technical and Conceptual Work. *International Journal of Computers for Mathematical Learning*, *7*(3), 245-74.

Awtry, T. & Kirshner, D. (1994). *Visual salience in algebraic transformations*. Paper presented at the 16th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education. Baton Rouge, LA: Louisiana State University.

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

Becker, J. P. & Jacob, B. (1998). Math War Developments in the United States (California). *ICMI Bulletin*, *44*.

Bedaux, J., & Boldy, M. (2007). *Ontwerpversie Aanbevelingen Toetsstandaarden Wiskunde*. Retrieved from http://www.surfspace.nl/nl/Redactieomgeving/Publicaties/Documents/ontwerpversie_toetsstandaarden_wiskunde.pdf

Beeson, M. (1989). Logic and computation in MATHPERT: An expert system for learning mathematics. In E. Kaltofen & S. Watt (Eds.), *Computers and mathematics* (pp. 202-214). New York: Springer.

Beeson, M. J. (1998). Design Principles of Mathpert: Software to support education in algebra and calculus. In N. Kajler (Ed.), *Computer-Human Interaction in Symbolic Computation* (pp. 89-115). Berlin/ Heidelberg/New York: Springer-Verlag.

Berger, M. (2010). Using CAS to solve a mathematics task: A deconstruction. *Computers & Education*, *55*(1), 320-332.

Black, P., & Wiliam, D. (1998). Inside the Black Box: Raising Standards Through Classroom Assessment. *Phi Delta Kappan*, *80*(2), 139-149.

Black, P. & Wiliam, D. (2004). The formative purpose: assessment must first promote learning. In Wilson, M. (Ed.), *Towards Coherence Between Classroom Assessment and Accountability - 103rd Yearbook of the National Society for the Study of Education* (pp. 20-50). Chicago: University of Chicago Press.

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

Bliss, J., & Ogborn, J. (1989). Tools for exploratory learning. *Journal of Computer Assisted Learning, 5*(1), 37-50.

Bokhove, C. (2008, June). *Use of ICT in formative scenarios for algebraic skills*. Paper presented at the 4th conference of the International Society for Design and Development in Education, Egmond aan Zee, the Netherlands.

Bokhove, C. (2010). Implementing Feedback in a Digital Tool for Symbol Sense. *International Journal for Technology in Mathematics Education, 17*(3), 121-126.

Bokhove, C. (2011). Digitaal werken aan algebraïsche vaardigheid en inzicht [Working digitally on algebraic skills and insight]. *Euclides, 86*(5), 186-188.

Bokhove, C. & Drijvers, P. (2010a). Digital tools for algebra education: criteria and evaluation. *International Journal of Computers for Mathematical Learning, 15*(1), 45-62.

Bokhove, C., & Drijvers, P. (2010b). Symbol Sense Behavior in Digital Activities. *For the Learning of Mathematics, 30*(3), 43-49.

Bokhove, C., & Drijvers, P. (2011, July). *Effects of feedback conditions for an online algebra tool*. Paper presented at the 10th International Conference on Technology in Mathematics Teaching, Portsmouth, UK.

Bokhove, C., & Drijvers, P. (in press). Effects of a digital intervention on the development of algebraic expertise. *Computers & Education*. http://dx.doi.org/10.1016/j.compedu.2011.08.010

Bokhove, C., Koolstra, G., Heck, A., & Boon, P. (2006). *Using SCORM to Monitor Student Performance: Experiences from Secondary School Practice*. Retrieved from http://mathstore.ac.uk/articles/maths-caa-series/apr2006/index.shtml

Buchberger, B. (1990). Should Students Learn Integration Rules? *SIGSAM Bulletin, 24*(1), 10–17.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving evaluation on interest and performance. *British Journal of Educational Psychology*, *58*, 1-14.

Butler, R. & Nisan, M. (1986). Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance. *Journal of Educational Psychology*, *78*(3), 210-216.

Carpenter, J. R., & Kenward, M. G. (2010). *Instructions for MLwiN multiple imputation macros*. Retrieved from http://missingdata.lshtm.ac.uk/

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C. P. & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: an experimental study. *American Educational Research Journal, 26*(4), 499-531.

Chevallard, Y. (1999). L'analyse des pratiques enseignantes en théorie anthropologique du didactique. *Recherches en Didactique des Mathématiques, 19*, 221–266.

Clariana, R. B. (1990). A Comparison of Answer until Correct Feedback and Knowledge of Correct Response Feedback under Two Conditions of Contextualization. *Journal of Computer-Based Instruction*, *17*(4), 125-29.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159.

Consortium NKBW. (2007). *Eindrapport NKBW 1*. Utrecht: NKBW. Retrieved from http://www.science.uva.nl/amstel/nkbw/archief/nkbw1/nkbw/documents/20071130_nkbw_eindrapport.pdf

Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, *58*(4), 438-481.

cTWO. (2007). *Rijk aan betekenis [Rich in meaning]*. Retrieved from http://www.fi.uu.nl/ctwo/publicaties/docs/Rijkaanbetekenisweb.pdf

De Wever, B., Schellens, T., Valcke, M., & Van Keer, H. (2006). Content Analysis Schemes to Analyze Transcripts of Online Asynchronous Discussion Groups: A Review. *Computers & Education*, *46*(1), 6-28.

Dedrick, R. F., Ferron, J. M., Hess, M. R., Hogarty, K. Y., Kromrey, J. D., Lang, T. R., Niles, J. D., et al. (2009). Multilevel Modeling: A Review of Methodological Issues and Applications. *Review of Educational Research*, *79*(1), 69-102.

Dempster, F. N. (1991). Synthesis of research on reviews and tests. *Educational Leadership, 48*(7), 71-76.

Dempster, F. N. (1992. Using tests to promote learning: A neglected classroom resource. *Journal for Research and Development in Education*, *25*(4), 213-217.

Dick, T. (2007). Keeping the faith. Fidelity in technological tools for mathematics education. In G. W. Blume & M. K. Heid (Eds.), *Research on technology and the teaching and learning of mathematics: Syntheses, cases, and perspectives. Vol. 2: Cases and perspectives* (pp. 333-339). Greenwich, CT: Information Age Publishing.

Drijvers, P. (2006). Context, abstractie en vaardigheid in schoolalgebra [Context, abstraction and skills in school algebra]. *Nieuw Archief voor Wiskunde*, *5*(7), 198–203.

Drijvers, P. & Kop, P. (2008). *Katern 1: Vergelijkingen vergelijken [Part 1: Comparing equations]*. Retrieved from http://www.fi.uu.nl/elwier/materiaal/handboek/documents/2008-10-13Vergelijkingen-Vergelijken.pdf.

Drijvers, P., Goddijn, A., & Kindt, M. (2010). Algebra education: exploring topics and themes. In P. Drijvers (Ed.), *Secondary algebra education. Revisiting topics and themes and exploring the unknown* (pp. 5-26). Rotterdam: Sense.

Elawar, M. C. & Corno, L. (1985). A Factorial experiment in teachers' written feedback on student homework: Changing teacher behavior a little rather than a lot. *Journal of Educational Psychology*, *77*(2), 162-173.

Ellington, A. J. (2003). A Meta-Analysis of the Effects of Calculators on Students'Achievement and Attitude Levels in Precollege Mathematics Classes. *Journal for Research in Mathematics Education, 34*(5), 433-463.

Elshout-Mohr, M. (1994). Feedback in self-instruction. *European Education*, *26*(2), 58-73.

Engineering Council. (2000). *Measuring the mathematics problem*. Retrieved from http://www.engc.org.uk/ecukdocuments/internet/document%20library/Measuring%20the%20Mathematic%20Problems.pdf

Field, A. (2005). *Discovering Statistics Using SPSS (2nd ed.)*. London: Sage Publications Ltd.

Garson, D. (2011). *Structural Equation Modeling*. Retrieved from http://faculty.chass.ncsu.edu/garson/PA765/standard.htm

Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and teaching in higher education, 1*, 3-31.

Goldstein, H. (2002). *Multilevel Statistical Models (4th ed.)*. Chichester: John Wiley & Sons Ltd.

Goos, M., Soury-Lavergne, S., Assude, T., Brown, J., Kong, C. M., Glover, D., Grugeon, B., et al. (2009). Teachers and Teaching: Theoretical Perspectives and Issues Concerning Classroom Implementation. In C. Hoyles & J.-B. Lagrange (Eds.), *Mathematics Education and Technology-Rethinking the Terrain Vol. 13* (pp. 311-328). Boston, MA: Springer US.

Gravemeijer, K. (1990). Globaal kijken, een kenmerk van algebraïsche deskundigheid [Global viewing: a characteristic of algebraic expertise]. *Nieuwe Wiskrant*, *10*(2), 29–33.

Guin, D., & Trouche, L. (1999). The Complex Process of Converting Tools into Mathematical Instruments: The Case of Calculators. *International Journal of Computers for Mathematical Learning, 3*(3), 195-227.

Guin, D., Ruthven, K., & Trouche, L. (Eds.). (2005). *The didactical challenge of symbolic calculators: Turning a computational device into a mathematical instrument*. New York: Springer.

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

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of Learning Skills Interventions on Student Learning: A Meta-Analysis. *Review of Educational Research*, *66*(2), 99-136.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication methods and measures, 1*(1), 77-89.

Hearnshaw, H. M., Harker, R. M., Cheater, F. M., Baker, R. H., & Grimshaw, G. M. (2001). Expert consensus on the desirable characteristics of review criteria for improvement of health care quality. *Quality in Health Care*, *10*(3), 173-178.

Heck, A., & Van Gastel, L. (2006). Mathematics on the Threshold. *International Journal of Mathematical Education in Science & Technology, 37*(8), 925-945.

Heeren, B., & Jeuring, J. (2010). Adapting Mathematical Domain Reasoners. In *Proceedings of MKM 2010, the 9th International Conference on Mathematical Knowledge Management Vol. 6167. Lecture Notes in Computer Science* (pp. 315 - 330). Berlin: Springer.

Heid, M. K., & Blume, G. W. (2008a). *Research on Technology and the Teaching and Learning of Mathematics: Vol. 1, Research Syntheses*. Charlotte, NC: Information Age Publishing.

Heid, M. K., & Blume, G. W. (2008b). *Research on Technology and the Teaching and Learning of Mathematics: Vol. 2, Cases and Perspectives*. Charlotte, NC: Information Age Publishing.

Hoch, M., & Dreyfus, T. (2004). Structure Sense in High School Algebra: The Effect of Brackets. In M.J. Høines & A.B. Fuglestad (Eds.) *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education Vol. 3* (pp. 49-56). Cape Town: International Group for the Psychology of Mathematics Education.

Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam: TT-Publikaties.

Jones, R. M., & Fleischman, E. S. (2001). Cascade Explains and Informs the Utility of Fading Examples to Problems. In J.D. Moore & K. Stenning (Eds.) *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 459-464). Mahwah, NJ: Erlbaum.

Jonker, V. (2007). *Digitaal authentiek toetsen van wiskunde [Digital authentic testing of mathematics]*. Retrieved from http://www.fi.uu.nl/wisweb/kloo/digitaaltoetsen/documents/20070329_klo_authentiek.pdf

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*(1), 23-31.

Kapur, M. (2008). Productive Failure. *Cognition and Instruction*, *26*(3), 379-424.

Kieran, C., & Drijvers, P. (2006). The Co-Emergence of Machine Techniques, Paper-and-Pencil Techniques, and Theoretical Reflection: A Study of CAS Use in Secondary School Algebra. *International Journal of Computers for Mathematical Learning*, *11*(2), 205-263.

Kilpatrick, J. E., Swafford, J. E. & Findell, B. E. (2001). *Adding it up: helping children learn mathematics*. Washington, DC: National Academy Press.

Kim, Y., & Wei, Q. (2011). The impact of learner attributes and learner choice in an agent-based environment. *Computers & Education*, *56*(2), 505-514.

Kirshner, D. (1989). The visual syntax of algebra. *Journal for Research in Mathematics Education*, *20*, 274–287.

Kirshner, D. & Awtry, T. (2004). Visual salience of algebraic transformations. *Journal for Research in Mathematics Education, 35*(4), 224–257.

Klein, D. (2007). A quarter century of US 'math wars' and political partisanship. *BSHM Bulletin: Journal of the British Society for the History of Mathematics*, *22*(1), 22 - 33.

Kolovou, A. (2011). *Mathematical problem solving in primary school*. Utrecht: FIsme Scientific Library, Utrecht University.

Kuhn, T. S. (1996). *The Structure of Scientific Revolutions (3rd ed.)*. Chicago: University Of Chicago Press.

Lagrange, J.-B. (2000). L'intégration d'instruments informatiques dans l'enseignement: une approche par les techniques. *Educational Studies in Mathematics, 43*, 1–30.

Li, Q., & Ma, X. (2010). A Meta-analysis of the Effects of Computer Technology on School Students' Mathematics Learning. *Educational Psychology Review*, *22*(3), 215-243.

London Mathematical Society. (1995). *Tackling the mathematics problem*. Retrieved from http://202.38.126.65/oldmirrors/www.lms.ac.uk/policy/tackling_maths_prob.pdf

Mason, B. J., & Bruning, R. (2001). *Providing Feedback in Computer-based Instruction: What the Research Tells Us*. Retrieved from http://dwb.unl.edu/Edit/MB/MasonBruning.html

Morrison, G. R., Ross, S. M., Gopalakrishnan, M., & Casey, J. (1995). The Effects of Feedback and Incentives on Achievement in Computer-Based Instruction. *Contemporary Educational Psychology*, *20*(1), 32-50.

National council of teachers of mathematics. (2008). *The Role of Technology in the Teaching and Learning of Mathematics*. Retrieved from http://www.nctm.org/about/content.aspx?id=14233

National Mathematics Advisory Panel. (2007). *Preliminary Report*. Retrieved from http://www.ed.gov/about/bdscomm/list/mathpanel/pre-report.pdf

Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, *22*(2), 155-175.

Nicaud, J.-F., Bouhineau, D. & Chaachoua, H. (2004). Mixing Microworld and Cas features in building computer systems that help students learn algebra. *International Journal of Computers for Mathematical Learning, 9*(2), 169–211.

Nicaud, J.-F., Bouhineau, D., Chaachoua, H., & Trgalova, J. (2006, December). *Developing Interactive Learning Environments that can be used by all the classes having access to computers. The case of Aplusix for algebra*. Paper presented at the 17th ICMI study: Digital Technologies and Mathematics Teaching and Learning: Rethinking the Terrain, Hanoi, Vietnam.

Nicol, D. J., & MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199-218.

Noss, R., & Hoyles, C. (1992). *Looking back and looking forward. Learning mathematics and Logo*. Boston, MA: MIT.

Nyquist, J. B. (2003). *The benefits of reconstructing feedback as a larger system of formative assessment: a meta-analysis*. Nashville: Vanderbilt University.

Papert, S. (1980). *Mindstorms: children, computers, and powerful ideas*. New York: Basic Books Inc.

Pierce, R., & Stacey, K. (2007). Developing Algebraic Insight. *Mathematics Teaching Incorporating Micromath*, *203*, 12-16.

Pierce, R., & Stacey, K. (2010). Mapping Pedagogical Opportunities Provided by Mathematics Analysis Software. *International Journal of Computers for Mathematical Learning*, *15*(1), 1-20.

Raudenbush, D. S. W., & Bryk, A. S. (2001). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage Publications, Inc.

Reed, H. C., Drijvers, P., & Kirschner, P. A. (2010). Effects of Attitudes and Behaviours on Learning Mathematics with Computer Tools. *Computers & Education*, *55*(1), 1-15.

Renkl, A., Atkinson, R. K., & Große, C. S. (2004). How Fading Worked Solution Steps Works – A Cognitive Load Perspective. *Instructional Science*, *32*(1/2), 59-82.

Renkl, A., Atkinson, R. K., Maier, U. H. & Staley, R. (2002). From example study to problem solving: smooth transitions help learning. *Journal of Experimental Education. 70*(4), 293–315.

Sangwin, C. & Naismith, L. (2008). Implementing computer algebra enabled questions for the assessment and learning of mathematics. *International Journal for Technology in Mathematics Education*, *15*(1), 3–18.

Sangwin, C., & Grove, M. (2006, January). *STACK: addressing the needs of the "neglected learners"*. Paper presented at the WEBALT conference, Eindhoven, the Netherlands.

Schoenfeld, A. H. (2004). The math wars. *Educational Policy*, *18*(1), 253–286.

Snijders, P. T. A. B., & Bosker, P. R. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage Publications Ltd.

Stacey, K., & Chick, H. (2000). Discussion Document for the Twelfth ICMI Study: The Future of the Teaching and Learning of Algebra. *Educational Studies in Mathematics, 4*2(2), 215-224.

Sweller, J., & Cooper, G. (1985). The Use of Worked Examples as a Substitute for Problem Solving in Learning Algebra. *Cognition and Instruction*, *2*(1), 59-89.

Tall, D. (1977). Cognitive Conflict and the Learning of Mathematics. In *Proceedings of the 1st Conference of The International Group for the Psychology of Mathematics Education*. Utrecht, Netherlands: PME.

Tempelaar, D. (2007). Onderwijzen of bijspijkeren? [Teach or remediate?]. *Nieuw Archief voor Wiskunde*, *8*(1), 55–59.

Tempelaar, D., & Caspers, W. (2008). De rol van de instaptoets [The role of entry exams]. *Nieuw Archief voor Wiskunde*, *9*(1), 66-71.

Tessmer, M. (1993). *Planning and Conducting Formative Evaluations*. Abingdon: Routledge.

Tondeur, J., van Keer, H., van Braak, J., & Valcke, M. (2008). ICT integration in the classroom: Challenging the potential of a school policy. *Computers & Education*, *51*(1), 212-223.

Trouche, L. (2004). Managing the Complexity of Human/Machine Interactions in Computerized Learning Environments: Guiding Students' Command Process through Instrumental Orchestrations. *International Journal of Computers for Mathematical Learning*, *9*(3), 281-307.

Underwood, J. S., Hoadley, C., Lee, H. S., Hollebrands, K., DiGiano, C., & Renninger, K. A. (2005). IDEA: Identifying Design Principles in Educational Applets. *Educational Technology Research and Development*, *53*(2), 99-112.

University of Waterloo. (2000). *Activities matrix - feedback type*. Retrieved from http://tlc.uwaterloo.ca/is301a/groups/sample/feedback.html

US Department of Education. (2007). *National Mathematics Advisory Panel preliminary report*. Retrieved from http://www.ed.gov/ about/bdscomm/list/mathpanel/pre-report.pdf

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

Van den Heuvel-Panhuizen, M. (1996). *Assessment and realistic mathematics education*. Utrecht: CD-β Press / Freudenthal Institute, Utrecht University.

Van Hiele, P. M. V. (1985). *Structure and Insight: A Theory of Mathematics Education*. Orlando, Fl: Academic Press.

Van Streun, A. (2000). Representations in applying functions. *International Journal of Mathematical Education in Science and Technology, 31*(5), 703-725.

VanLehn, K., Jones, R. M., & Chi, M. T. H. (1992). A Model of the Self-Explanation Effect. *The Journal of the Learning Sciences*, *2*(1), 1-59.

Vasilyeva, E., Puuronen, S., Pechenizkiy, M., & Rasanen, P. (2007). Feedback adaptation in web-based learning systems. *International Journal of Continuing Engineering Education and Life-Long Learning*, *17*(4/5), 337-357.

Verillon, P. & Rabardel, P. (1995) Cognition and artifacts: A contribution to the study of though[t] in relation to instrumented activity. *European Journal of Psychology in Education*, *9*, 77-101.

Voogt, J., & Knezek, G. (Eds.). (2008*). International Handbook of Information Technology in Primary and Secondary Education (Vol. 20)*. Boston, MA: Springer US.

Vos, P. (2007). Algebra-prestaties van tweedeklassers [Algebra performance of second graders]. *Euclides*, *82*(4), 129-132.

Vygotsky, L. S. (1978) *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wenger, R. H. (1987). Cognitive science and algebra learning. In A. Schoenfeld (Ed.), *Cognitive science and mathematics education* (pp. 217–251). Hillsdale, NJ: Lawrence Erlbaum Associate.

Zorn, P. (2002). *Algebra, computer algebra and mathematical thinking*. Retrieved from http://www.stolaf.edu/people/zorn/cretepaper.pdf

Summary

Samenvatting

Acknowledgements

Curriculum Vitae

Summary

This PhD thesis documents the development and outcomes of a mathematics education research that was carried out from 2007 until 2011 at the Freudenthal Institute for Science and Mathematics Education within the DUDOC Program funded by Platform Beta Techniek. This Platform facilitates secondary school teachers to carry out scientific research in the domain of science and mathematics education. An important emphasis in the DUDOC Program is on bridging the gap between theory and practice when it comes to educational research. This thesis comprises a series of articles each addressing a different perspective of this study on the use of ICT for acquiring, practicing and assessing algebraic expertise.

For several years the skill level of students leaving secondary education in the Netherlands has been discussed. Lecturers in higher education –for example– complain about their freshmen's apparent lack of algebraic skills. Another development in recent years is the advent of the use of technology in mathematics education (Li & Ma, 2010). Combining algebraic expertise and ICT use, the aim of this study is to design an online environment for learning supported by formative assessment of both procedural skills and conceptual understanding in algebra, to investigate the effects of the environment, and to identify decisive factors that influence the outcome. The central research question therefore is: i*n what way can the use of ICT support acquiring, practicing and assessing algebraic expertise?* This general question leads to several sub-questions, each related to an appropriate cycle in the study:

1a) Which criteria are relevant for the evaluation of digital tools for algebra education?
1b) Which digital algebra tool best meets these criteria?
2a) Do the concepts of symbol sense, gestalt view and pattern and local visual salience, described in a pre-digital era, help us in understanding what students do in a digital environment?
2b) Can the feedback design of a digital tool be improved with students?
2c) Which methodology can be used to elaborate feedback desired by students?
3a) Timing and fading: are there indications that formative scenarios improve the acquisition of algebraic expertise?
3b) Crises: do crises in algebraic tasks improve the acquisition of algebraic expertise?
3c) Variation: does variation in feedback influence scores and student behavior?
4a) What is the effect of an intervention on the development of algebraic expertise, including both procedural skills and symbol sense?
4b) Which factors predict the resulting algebraic performance?

The theoretical framework is based on the three key perspectives of ICT tool use, algebraic expertise, assessment and feedback. ICT tool use concerns characteristics of tools and the

ways they are adopted. The main concepts involved are notions on instrumentation and effectiveness of ICT tools. With algebraic expertise we mean both skills and understanding, described by Arcavi as 'symbol sense' (Arcavi, 1994). We regard algebraic expertise as a dimension ranging from basic skills to symbol sense (Drijvers, Goddijn, & Kindt, 2010), and contend that both should be addressed. The framework concerning assessment refers Black and Wiliam's distinction (Black & Wiliam, 1998) of assessment *for* learning (formative assessment) and assessment *of* learning (summative assessment). As feedback plays a pivotal role in formative assessment it also forms the core of this theoretical lens. As we aim to design an intervention in several iterations, the research method is based on the principles of design research (Van den Akker et al., 2006). The development throughout the cycles can be characterized by a shift in focus: from more qualitative formative towards quantitative summative. This involves scaling up from a small target audience towards a larger target audience (Tessmer, 1993). This pragmatic approach requires our methodology to be mixed: at first more qualitative, and more quantitative later on, using a more quasi-experimental approach with pre- and post-tests. Research takes place in one preparatory cycle and three subsequent cycles, which are described in chapters 2-6.

Chapter 2 focuses on the preliminary cycle concerning the design of criteria for an evaluation instrument for digital algebra tools. It aims to give a clear picture of what characteristics are deemed important by experts (N=27) and as such provides a framework for studying the plethora of tools that are currently available. We conclude that the designed evaluation instrument is appropriate for describing the characteristics of digital tools for algebra that we consider relevant for the purpose of the study. The evaluation instrument provides insight in the different features of a tool, as well as in our priorities in interest. We also conclude that the Digital Mathematics Environment (DME) scored highest overall and is best suited for addressing the research goals on the co-emergence of procedural skill fluency and symbol sense expertise. A key feature of DME is that it enables stepwise problem solving strategies. It is easy to use, stores the students´ solution processes, and is well suited for formative assessment, by offering several strategy modes, feedback and self-review. The results of the second chapter are used as an initial argument for the tool choice: why choose *this* tool and what characteristics of the tool are paramount? With this DME tool we designed a prototypical online intervention for the development of algebraic expertise.

Chapter 3 reports on the first research cycle with one-to-one think-aloud sessions that were conducted with five 12th grade students studying 'wiskunde B' in the pre-university science stream. The design process and the one-to-one pilot sessions suggest that the concepts symbol sense, gestalt view and visual salience remain extremely relevant when deploying digital activities. The observations show that students using a digital environment exhibit some symbol sense behaviors but also exhibit behaviors lacking symbol sense. The notions of gestalt view and visual salience are helpful in analyzing student work. Although students work in a digital environment instead of with paper and pencil, these results are in line with past findings in traditional pen-and-paper settings (Arcavi, 1994; Wenger, 1987). While solving algebraic tasks in the digital environment, students can use any strategy, and thus can show sensitivity towards gestalt and visual salience aspects, and further develop such sensibility. The digital environment facilitates this development through its mathematical interface and feedback opportunities, which would be more difficult to offer in a paper-and-pencil environment. Furthermore, extending the concept of visual salience to patterns provided by standard routines students already know may relieve the tension between the application of standard routines and succumbing to salient patterns.

In chapter 4 we describe how the qualitative observations presented in the previous chapter inform us about ways in which we can design feedback as part of the intervention, which relates to one of the criteria in chapter 2. After studying student behavior, we conclude that asking students when to use what feedback can improve a digital intervention. A fruitful methodology is: (i) deploy the digital intervention and ask students to 'think aloud'; (ii) tabulate all the responses; (iii) make a feedback design; (iv) implement the feedback design in the environment; (v) test the intervention with help of the feedback design, and make another iteration. Based on the initial characteristics presented in chapter 2, the digital activities on symbol sense (chapter 3), and the student-informed methodology for feedback design (chapter 4), we redesigned the intervention in an iterative fashion. The revised intervention was field tested in a second cycle for two classes (12^{th} grade, wiskunde B, N=31), after which we made the final improvements based on several design principles.

In chapter 5 we elaborate on these design principles and on the way these principles worked or didn't work in the third and last cycle in nine different schools (N=324). The three design principles all focus on feedback and include timing and fading with formative scenarios, crises and variation. Firstly, timing and fading with formative scenarios means that the digital intervention starts off with a practice mode with much feedback, then provides a diagnostic mode only giving feedback at *the end* of the solution process, and finally ending up with a final summative digital test without feedback except for the final score. Secondly, crises, which are created intentionally in series of test items, are difficult non-standard items that can only be solved if students understand the structure of an expression. Right after a crisis item students are provided with feedback of different types, such as screencast clips, hints and worked examples. This variation in feedback type is a third design principle. Taken together we conclude that the use of feedback in an online algebra intervention can have a significant effect. Furthermore, the design principles provide a broad approach to the potential of feedback use in digital environments. The variation principle led to two separate experimental conditions for the study. The first condition only provides task-related feedback on all steps of the solution process. The second, self-regulated condition provides additional buttons for hints, next solution steps and worked examples. Although gradually fading feedback in the course of the intervention – in line with the notion of formative scenario – causes 'pain' to the students, the learning effect is there.

The design criterion of using formative scenarios improves the acquisition of algebraic expertise. The results for the design of crises are mixed. When looking at the fraction of items answered correctly there seems to be no effect for a crisis task. This is probably caused by the difficulty to assess skill level if the final score after practice is high, although achieved with difficulty. The number of attempts after a crisis task, however, did decrease in three out of four series of tasks. The use of square root signs and logarithms in the tasks may have played a role in the one series where number of attempts did not decrease. However, whether crises improve student's actual expertise remains unclear from these results. Comparing the two feedback conditions, the student scores show a medium effect for the feedback condition including both task-related and self-regulatory feedback. When looking in more detail at the use of the feedback in the three case examples it is clear that both can be used in a formative way *for* the learning of algebra. Students can use the feedback to overcome difficulties and check whether they are on the correct solution path or not.

In chapter 6 we describe the overall effects of the online intervention in the third and final cycle (N=324). For the analysis multilevel modeling is used. The use of the intervention for an average of five hours has a large effect on improving algebraic expertise, as post-test score (M=78.71, SE=15.175) is significantly higher than the pretest score (M=51.55, SE=21.094), $t(286)=-22.589$, $p<.001$, $r=.801$. Comparison of symbol sense pre- and post-scores reveals that the symbol sense score on the posttest (M=1.462, SE=1.504) also is significantly higher than the pretest score (M=-1.493, SE=2.339), $t(285)=-20.602$, $p<.001$, $r=.773$. According to Cohen's benchmark (Cohen, 1992) this is a large effect. This means that there is not only an improvement in score, but also an improvement in recognizing patterns and having a sense for symbols. A vast improvement is apparent in all four categories of questions, with the category described by Wenger (1987), an example being "solve for v: $v \cdot \sqrt{v} = 1 + 2v \cdot \sqrt{1+u}$", showing the biggest improvement. Furthermore, previous knowledge, time spent in self-test and summative test mode, and general attitude towards mathematics are the largest predictors for a high posttest outcome. The fact that the time spent in the self-test mode and digital summative test are more significant explanatory variables than the practice mode suggest that students should be able to 'stand on their own two feet', and that environments that offer feedback should decrease the amount of feedback when moving towards summative assessment. The other two significant explanatory variables, pre-test score and attitude towards mathematics, seem less groundbreaking. However, it was not clear-cut beforehand that these two variables would also influence the outcome in the case of using an *online* environment for algebraic *expertise*. The fact that these variables have nothing to do with ICT may indicate that indeed conventional pen-and-paper techniques and ICT techniques are reconciled (Kieran & Drijvers, 2006). In line with this, the variables overall quality of the school (operationalized by trend exam grades), total practice time and whether teachers and students worked more at home or at school, also did not significantly predict the outcome.

Chapter 7 synthesizes the thesis through conclusions, recommendations, point of discussion and suggestions for further research. We describe the conclusions by returning to the main research question and scrutinizing its parts:

In what way. There are many ways to design an environment. This study shows that a framework based on formative assessment, with feedback design as guiding principle, can be a successful approach.
Use of ICT. We have seen that we can use ICT, an online environment in particular, for learning algebra. We have to keep in mind the tool's affordances and constraints, and the purpose for which it is deployed. Furthermore, because of instrumentation, student characteristics must be taken into account. In our study, on the one hand there was no gender difference in effect. Students' pre-knowledge, on the other hand, did make a difference: the better students scored on the pre-test, the less gain they had.
Acquiring. In developing algebraic expertise, we have found that the use of intentional crises seems to decrease the number attempts that are needed for aid the transfer of algebraic expertise.
Practicing. A didactical scenario that enables students to practice in the beginning stages of an intervention, and using timing and fading, ensures that students learn to 'stand on their own feet', and to be independent of the practice tool.
Assessing. A didactical scenario also needs room for summative testing. Combining formative, summative assessment and feedback in one scenario works.
Algebraic expertise. Summative assessment shows that the intervention has a large effect on both sides of the algebraic expertise dimension, basic skills and symbol sense.

What does this study add to the existing vast body of research knowledge? For this we return to the three key theoretical foundations of the study: ICT tool use, algebraic expertise, assessment and feedback. The first topic of tool use mainly played a part in the preliminary cycle of our study: by explicitly adding characteristics concerning tool use, e.g. ease of use, to the evaluation instrument we made sure that the tool we would choose would best meet our purposes. The second topic concerns algebraic expertise. The dimension basic skills - symbol sense proved to be useful for identifying item categories and for interpreting student work. Both the extension of theory with pattern salience and the four item categories we identified contribute to the operationalization of the notion of symbol sense in relation to digital activities. The notions of formative assessment, the third perspective, helped us in designing the intervention, and the three feedback design principles in particular. These three feedback principles can also be seen as an operationalization of the idea of formative assessment in the realm of feedback design in online activities.
The study as a whole leads to the following recommendations for designers and teachers:

1. Use student and teacher feedback to improve the design of digital activities.

2. Make full use of the possibility of feedback generated by a digital environment.
3. Context: choose an appropriate tool and make sure you know what it can and cannot do for you.
4. Embed activities in pedagogical and didactical scenario that facilitate transfer.
5. Algebraic expertise: do not bet on just skills or just insight, as both are closely interlinked.
6. Use 'unusual' tasks to improve learning, for example by intentionally causing a crisis.
7. See the use of online environments as an incentive to rethink home and classroom work.
8. Differentiate, because good students may benefit less.

One of the discussion points addressed in the final chapter concerns the interplay between acquiring skills and understanding: where do skills end and where does understanding start, and can a sensible distinction be made? We think that utilizing a holistic approach, and considering algebraic skills and symbol sense as one continuous dimension, does away with unfruitful discussions about cause and effect. Also, the fact that this study just focuses on a small sub-domain of algebraic knowledge raises the question whether the findings also hold for other algebra, and even other mathematical subjects. We think that we have laid down general principles that can also be applied to other subjects, and would encourage further research in this field. Another discussion point concerns the study's methodology: we want to be in control of what goes on in our research, but the home context of using an online intervention can hardly be controlled. For the moment we will have to deal with this through instruments such as log-files and students' self-reports. A final point regarding 'didactics of technological shortcomings' is made: glaring disadvantages of digital (algebra) tools should not be trivialized and downplayed by providing a didactical defense of a drawback.

To conclude, further research is needed to replicate the findings and to improve our understanding of the use of ICT for algebra. We certainly should take into account recent software and hardware developments, e.g. tablet technology.

Samenvatting

Dit proefschrift beschrijft de ontwikkeling en uitkomst van onderzoek naar het wiskundeonderwijs, dat plaatsvond in de periode 2007 - 2011 op het Freudenthal Instituut van de Universiteit Utrecht. Het onderzoek is deel van het Dudoc-programma van het Platform Bèta Techniek, een programma dat docenten in het voortgezet onderwijs in staat stelt om wetenschappelijk onderzoek te doen naar onderwijs in wiskunde en natuurwetenschappen. Een belangrijke accent in dit programma ligt op het overbruggen van de kloof tussen theorie en praktijk. Dit proefschrift bestaat uit een reeks artikelen, waarvan elk artikel een andere invalshoek beschrijft van deze studie naar het gebruik van ICT voor het verwerven, oefenen en toetsen van algebraïsche expertise.

Al meerdere jaren staat het vaardigheidsniveau van studenten die van het voortgezet onderwijs het hoger onderwijs instromen ter discussie. Zo klagen universitaire docenten bijvoorbeeld over het vermeende gebrek aan algebraïsche vaardigheden van eerstejaars studenten. Een andere recente ontwikkeling is de opkomst van technologiegebruik in het wiskundeonderwijs (Li & Ma, 2010). Deze twee ontwikkelingen op het gebied van algebraïsche expertise en ICT-gebruik combinerend, is het doel van deze studie om een online omgeving te ontwerpen voor de formatieve toetsing van zowel algebraïsche vaardigheid als inzicht, om de effecten van deze omgeving te onderzoeken en om vast te stellen welke factoren een positieve invloed hebben op deze effecten. De centrale onderzoeksvraag is dan ook: op welke wijze kan ICT worden gebruikt voor het verwerven, oefenen en toetsen van algebraïsche expertise? Deze algemene vraag wordt onderverdeeld in de volgende deelvragen, die alle verbonden zijn met een fase in het onderzoek:

1a) Welke criteria zijn relevant voor de evaluatie van digitale tools voor het algebraonderwijs?
1b) Welke digitale algebra tool voldoet het beste aan deze criteria?
2a) Kunnen de begrippen symbol sense, gestalt en pattern en local visual salience, afkomstig uit een pre-digitaal tijdperk, helpen bij het begrijpen van wat leerlingen doen in een digitale omgeving?
2b) Kan het ontwerpen van feedback in een digitale omgeving verbeterd worden door leerlingen zelf in te zetten?
2c) Welke methodologie kan gebruikt worden om door leerlingen gewenste feedback te identificeren?
3a) *Timing en fading*: kunnen formatieve scenario's het verwerven van algebraïsche expertise verbeteren?
3b) Crises: kunnen crises in algebrataken de verwerving van algebraïsche expertise verbeteren?
3c) Variatie: beïnvloedt variatie in feedback scores en gedrag van studenten?

4a) Wat is het effect van een interventie die zich richt op de ontwikkeling van algebraïsche expertise, inclusief basisvaardigheden en symbol sense?
4b) Welke factoren voorspellen de algebraïsche prestaties van leerlingen?

Het theoretische kader van het onderzoek steunt op drie pijlers: ICT tool use, algebraïsche expertise, en toetsing en feedback. Bij tool use gaat het over eigenschappen van digitale tools en de manieren waarop deze worden gehanteerd. De belangrijkste begrippen die hier een rol spelen, zijn instrumentatie en de effectiviteit van ICT-hulpmiddelen. Met algebraesche expertise bedoelen we zowel vaardigheden als inzicht, zoals beschreven als 'symbol sense' (Arcavi, 1994). Wij beschouwen algebraïsche expertise als een dimensie die loopt van basisvaardigheden naar symbol sense (Drijvers, Goddijn, & Kindt, 2010), en stellen dat beide van belang zijn. Het raamwerk dat we voor toetsing gebruiken, is gebaseerd op het onderscheid dat Black en Wiliam maken (Black & Wiliam, 1998) tussen toetsing *voor* het leren (formatieve toetsing) en toetsing *van* het leren (summatieve toetsing), waarbij in deze studie de nadruk ligt op het eerste. Aangezien feedback een centrale rol speelt in formatieve toetsing maakt dit begrip ook onderdeel uit van ons theoretische kader.

Daar het doel van het onderzoek is om een interventie te ontwerpen, is de onderzoekmethode gebaseerd op principes van design research (Van den Akker et al., 2006). De ontwikkeling van de interventie door de verschillende iteraties heen kan worden gekarakteriseerd door een verschuiving in de focus van het onderzoek: van meer kwalitatief en formatief in het begin naar meer kwantitatief en summatief op het einde. Dit houdt ook in dat er een opschaling plaatsvindt van een kleine groep participanten in het begin van de studie naar een grotere groep deelnemers in de eindfase (Tessmer, 1993). Deze aanpak vereist dat ook de methodologie ontwikkelt: eerst kwalitatief, later kwantitatief met een meer quasi-experimentele aanpak met pre- en posttests. Het onderzoek omvat één voorbereidende cyclus, en drie opeenvolgende cycli, die worden beschreven in de hoofdstukken 2 tot en met 6.

Hoofdstuk 2 richt zich op de voorbereidende cyclus en beschrijft het ontwerpen van criteria voor een evaluatie-instrument voor digitale algebra tools. Het schetst een beeld van de eigenschappen van dergelijke tools, zoals die door experts (N=27) belangrijk worden gevonden en vormt daarmee een raamwerk voor de bestudering van de vele tools die op de markt zijn. De conclusie is dat het ontworpen evaluatie-instrument geschikt is voor het beschrijven van eigenschappen van digitale algebra tools die wij relevant achten voor het doel van het onderzoek. Het instrument geeft inzicht in deze eigenschappen en weerspiegelt tevens de focus van het onderzoek. Daarnaast wordt geconcludeerd dat de Digitale Wiskunde Omgeving (DWO) over het geheel genomen het beste scoort en derhalve het meest geschikt is voor dit onderzoek, ook omdat het onderzoek zich richt op zowel vaardigheden als inzicht. Een sleuteleigenschap van de DWO is dat het hierin mogelijk is om stapsgewijze oplossingsstrategieën te gebruiken. Daarnaast is de DWO eenvoudig in het gebruik, wordt het gehele oplossingsproces van de leerling opgeslagen, en worden er mogelijkheden voor formatieve toetsing geboden doordat diverse strategische modi, feedback en de optie tot zelf-toetsing ter beschikking staan. Deze resultaten onderbouwen de toolkeuze en geven antwoord op de voor de hand liggende vraag: waarom *dit* tool en welke eigenschappen zijn hier het belangrijkst? Met behulp van de DWO is vervolgens een prototypische online interventie voor de ontwikkeling van algebraïsche expertise ontworpen.

Hoofdstuk 3 rapporteert over de eerste onderzoekcyclus bestaande uit één-op-één hardop-denksessies die gehouden zijn met vijf 6 vwo-leerlingen uit de Natuur-profielen met wiskunde B in het pakket. Het ontwerpproces en de één-op-één sessies laten zien dat de concepten symbol sense, gestalt en visual salience ook relevant zijn bij de inzet van di-gitale activiteiten. Uit de observaties blijkt dat leerlingen die de digitale omgeving gebruiken zowel symbol sense gedrag laten zien alsook het ontbreken van dergelijk gedrag. De begrippen gestalt view en visual salience zijn tevens nuttig bij het analyseren van leerling werk. Hoewel leerlingen werken in een digitale omgeving en niet met pen en papier, komen deze resultaten overeen met bevindingen uit pen-en-papiercontexten (Arcavi, 1994; Wenger, 1987). Bij het oplossen van algebraopgaven in een digitale omgeving kunnen leerlingen elke strategie gebruiken en daarmee aspecten van gestalt en visual salience tonen en verder ontwikkelen. De digitale omgeving faciliteert deze ontwikkeling door de wiskundige interface en feedbackmogelijkheden, die moeilijker te realiseren zouden zijn in een pen-en-papieromgeving. Door de theorie rond visual salience uit te breiden naar patronen wordt ook een verband gelegd tussen enerzijds het toepassen van standaard algoritmes en het herkennen van patronen.

Hoofdstuk 4 gaat over hoe de kwalitatieve observaties, beschreven in het vorige hoofdstuk, zijn gebruikt om de feedback in de interventie te verbeteren. De feedbackmogelijkheid was ook één van de criteria voor algebratools zoals genoemd in hoofdstuk 2. Na bestudering van het leerlinggedrag concluderen we dat we simpelweg aan leerlingen zelf kunnen vragen welke feedback de digitale interventie zou verbeteren. Een kansrijke methodologie is: (i) zet de digitale interventie in en vraag leerlingen om ‘hardop te denken’; (ii) groepeer alle reacties; (iii) maak vervolgens een ontwerp voor de feedback; (iv) implementeer het feedback ontwerp in de omgeving; (v) test de interventie met behulp van het feedback design, en maak een volgende iteratie. Op basis van de gewenste eigenschappen uit hoofdstuk 2, de digitale activiteiten over symbol sense (hoofdstuk 3), en de methode om feedback te verbeteren (hoofdstuk 4), is de interventie op iteratieve wijze bijgesteld. In een tweede cyclus is de gereviseerde interventie in de praktijk getest in een 6 vwo-wiskunde B-klas van 31 leerlingen, waarna we de laatste iteratie baseren op enkele centrale ontwerpprincipes.

In hoofdstuk 5 lichten we deze ontwerpprincipes toe, alsook de manier waarop deze principes werkten in de derde en laatste cyclus van het onderzoek. De drie gehanteerde ontwerpprincipes richten zich op feedback en bestaan uit: timing en fading met behulp van formatieve scenario’s, crises en variatie in feedback. De eerste, timing en fading, houdt in dat de online digitale interventie begint in een oefenmodus met veel feedback, daarna een diagnostische modus gebruikt waarbij feedback alleen aan het *einde* van een taak verschijnt, en tenslotte eindigt met een laatste, summatieve digitale toets waarbij geen feedback wordt gegeven, behoudens de eindscore. Het tweede principe, crises, betreft het opzettelijk aanbrengen van moeilijke, niet-standaard items in de serie opgaven, die alleen kunnen worden opgelost als de leerling de structuur van een expressie echt door heeft. Meteen na een crisis-item ontvangen leerlingen verschillende soorten feedback, zoals films, hints en uitgewerkte voorbeelden. De variatie in type feedback is het derde ontwerp-principe. Dit principe leidde tot twee aparte experimentele condities voor de studie. De eerste conditie betreft het enkel aanbieden van taakgerelateerde feedback op alle tussen-stappen van de oplossing. De tweede, zelfregulerende conditie biedt extra knoppen voor hints, volgende stappen en uitgewerkte oplossingen. Samenvattend is de conclusie dat het gebruik van feedback in een online algebra-interventie een significant effect kan hebben. Verder verschaffen de ontwerpprincipes een brede blik op het potentieel van feedback in digitale omgevingen. Hoewel het verminderen van de hoeveelheid feedback gedurende de interventie –zoals ook bedoeld is met het formatieve scenario- ‘pijn’ doet omdat leerlingen op eigen benen moet staan, is er wel een leereffect. Met andere woorden: het ontwerpprin-cipe ten aanzien van formatieve scenario’s verbetert de verwerving van algebraïsche ex-pertise. De resultaten van het principe ten aanzien van crises zijn daarentegen gemengd. Op de fractie van correct beantwoorde vragen lijken crisis-items geen effect te hebben. Dit zou veroorzaakt kunnen zijn door het feit dat leerlingen net zo lang door kunnen werken aan een item tot ze het correcte antwoord gevonden hebben. Kijken we daarom naar het aantal pogingen na een crisis opgave, dan blijkt dit in drie van de vier series taken af te nemen. In de serie waarin dit niet gebeurde, zou dit veroorzaakt kunnen zijn door het ge-bruik van worteltekens en logaritmes. Of het daadwerkelijk de crises zijn die de expertise van een leerling verbeteren, blijft echter uit deze resultaten onduidelijk. De resultaten over de twee feedbackcondities geven een gemiddeld effect voor de tweede feedbackconditie met zowel taak gerelateerde als zelfregulerende feedback. Samen met een gedetailleerdere blik op drie casussen is het duidelijk dat beide op een formatieve wijze kunnen worden in-gezet *voor* het leren van algebra. Leerlingen kunnen de feedback gebruiken om moeilijk-heden te overwinnen en te controleren of ze op het pad van de juiste oplossing zitten of niet.

In hoofdstuk 6 beschrijven we de algemene effecten van de online interventie in de der-de en laatste cyclus op negen scholen (11 klassen, N=324 leerlingen). Voor de analyse is multilevel analyse gebruikt. De interventie heeft een gemiddelde duur van ongeveer vijf uur en heeft een groot effect op de verbetering van algebraïsche expertise, waarbij de post-test score (M=78.71, SE=15.175) significant hoger uitvalt dan de pretest score (M=51.55, SE=21.094), $t(286)=-22.589$, $p<.001$, $r=.801$. Het vergelijken van pre- en posttest scores op het gebied van symbol sense laat zien dat ook de symbol sense score op de posttest (M=1.462, SE=1.504) significant hoger is dan de pretest score (M=-1.493, SE=2.339, $t(285)=-20.602$, $p<.001$, $r=.773$). Volgens Cohen (1992) is dit een groot effect. Dit duidt er op dat er niet alleen een verbetering is in de prestaties, maar ook een verbetering in het herkennen van patronen en in het gevoel voor symbolen. Een grote verbetering is ook zichtbaar in de prestaties op de vier in hoofdstuk 3 beschreven itemcategorieën, waarbij de categorie die gebaseerd is op Wenger (1987), waarvan een voorbeeld "los op voor *v*: $v \cdot \sqrt{v} = 1 + 2v \cdot \sqrt{1+u}$" de grootste verbetering laat zien. Een andere bevinding is dat voorkennis, tijd besteed aan digitale zelftoets en eindtoets, en algemene houding ten aanzien van wiskunde de beste voorspellers zijn voor een hoge posttest score. Het feit dat de hoeveelheid aan digitale zelftoets en eindtoets bestede tijd betere voorspellers zijn dan tijd besteed aan de oefenmodus suggereert dat leerlingen het beste kunnen leren 'op eigen benen te staan' als de hoeveelheid feedback vermindert in de aanloop naar summatieve toetsing. De twee andere significante voorspellers, pretest score en houding ten aanzien van wiskunde, liggen meer voor de hand. Wat op voorhand echter nog niet duidelijk was, is of deze variabelen ook invloed zouden hebben op de posttest score bij gebruik van een *online* omgeving voor algebraïsche *expertise*. Het feit dat er geen voorspellende variabelen betreffende de houding ten opzichte van ICT en wiskunde waren, zou er op kunnen duiden dat de conventionele pen-en-papieraanpak en de digitale aanpak op één lijn liggen (Kieran & Drijvers, 2006); immers, die houding had geen invloed op de posttest score.. Ook de variabelen betreffende de kwaliteit van de school (geoperationaliseerd door trend examencijfers), totale oefentijd en of leerlingen meer thuis of op school werkten, bleken geen voorspellers van de posttest score.

In hoofdstuk 7 monden de diverse hoofdstukken uit in conclusies, aanbevelingen, discussiepunten en suggesties voor vervolgonderzoek. We beschrijven de conclusies door weer terug te komen op de centrale onderzoeksvraag en de verschillende onderdelen daarvan:

> *Op welke wijze.* Er zijn meerdere manieren om een leeromgeving te ontwerpen. Deze studie laat zien dat een raamwerk dat gebaseerd is op formatieve toetsing, met feedback als leidend principe, succesvol kan zijn.
> *Gebruik van ICT.* We hebben gezien dat we ICT, zeker in de vorm van een online omgeving, kunnen gebruiken voor het leren van algebra. Hierbij moeten we rekening houden met de voor- en nadelen van een tool en met het doel waarvoor we dit inzetten. Bovendien, vanwege instrumentatie moeten leerlingeigenschappen worden meegenomen bij het onderzoek. Zo was er enerzijds in onze studie geen verschil in sekse voor wat betreft effect, anderzijds maakte voorkennis van de leerling wel een verschil: de op de pretest beter scorende leerlingen hadden minder baat bij de interventie.
> *Verwerven.* Bij het ontwikkelen van algebraïsche expertise hebben we gezien dat het gebruik van opzettelijk gecreëerde crises het benodigde aantal pogingen lijkt te verkleinen.
> *Oefenen.* Een didactisch scenario dat de leerlingen in staat stelt om te oefenen aan het begin van een interventie en gebruik maakt van timing en fading, bevor-

dert dat leerlingen leren 'op eigen benen staan' en niet afhankelijk zijn van de gebruikte software.
Toetsen. Een didactisch scenario heeft ook summatieve toetsing nodig. Het combineren van formatieve en summatieve toetsing in één scenario werkt, in het bijzonder als dit gepaard gaat met aangepaste feedback.
Algebraïsche expertise. De resultaten van summatieve toetsing tonen aan dat de interventie een groot effect heeft op beide kanten van de dimensie algebraïsche expertise, basisvaardigheden en symbol sense.

Wat voegt deze studie toe aan de al bestaande kennis? Om dit te beschrijven, komen we terug op de drie pijlers van het theoretische raamwerk: ICT tool use, algebraïsche expertise, toetsing en feedback. De eerste pijler speelde voornamelijk een rol in de eerste voorbereidende cyclus van de studie: door expliciet tool use eigenschappen toe te voegen aan het evaluatieinstrument, bijvoorbeeld gebruiksgemak, verzekerden we ons ervan dat de uit te kiezen tool het beste bij ons doelen zou passen. De tweede pijler betreft algebraïsche expertise. De dimensie basisvaardigheden – symbol sense bleek nuttig voor het identificeren van items en itemcategorieën en voor het interpreteren van leerlingwerk. Zowel de uitbreiding van de theorie met pattern salience als de vier door ons vastgestelde categorieën dragen bij aan de operationalisering van het begrip symbol sense in relatie tot digitale activiteiten. Het begrip formatieve toetsing, de derde pijler, hielp ons bij het ontwerpen van de interventie, en de drie ontwerpprincipes over feedback in het bijzonder. Ook deze drie principes vormen een operationalisering van het concept formatieve toetsing ten behoeve van het ontwerp van feedback binnen online activiteiten.
De gehele studie leidt tot de volgende aanbevelingen voor ontwerpers en docenten:

1. Gebruik feedback van leerlingen en docenten om een ontwerp van digitale activiteiten te verbeteren.
2. Maak volledig gebruik van de mogelijkheid om feedback te laten genereren door een digitale omgeving.
3. Kies een toepasselijke tool en weet zeker wat een tool wél en niet kan.
4. Integreer activiteiten in een pedagogisch en didactisch scenario.
5. Zet niet alleen in op vaardigheden of alleen op inzicht, aangezien beide met elkaar verweven zijn als verschillende aspecten van algebraïsche expertise.
6. Gebruik 'ongewone' opgaven om het leren te verbeteren, bijvoorbeeld door expres een crisis-item toe te voegen.
7. Beschouw het gebruik van online omgevingen als een prikkel om na te denken over de verhouding tussen huiswerk en schoolwerk.
8. Maak gebruik van differentiatie, bijvoorbeeld omdat goede leerlingen minder baat hebben bij de interventie.

Eén van de discussiepunten in het laatste hoofdstuk is de relatie tussen het verwerven van vaardigheden en het verwerven van inzicht: waar eindigt een vaardigheid en waar begint begrip, en is het wel vruchtbaar om dit onderscheid te hanteren? We denken dat een integrale benadering, waarbij algebraïsche vaardigheden en symbol sense gezien worden als één continue dimensie, een eind kan maken aan discussies over oorzaak en gevolg. Een ander punt betreffende de focus op maar een klein deelgebied van de algebra betreft de vraag of de bevindingen ook op een breder gebied van de algebra, of zelfs op andere wiskundige onderwerpen, van toepassing zijn. We denken dat we degelijke algemene principes hebben neergelegd die ook van toepassing kunnen zijn op andere vakgebieden, en we raden derhalve meer onderzoek hiernaar aan. Nog een discussiepunt betreft de methodologie van de studie: we willen bij een onderzoek zo veel mogelijk controle over de onderzoekssituatie hebben, maar het feit dat leerlingen ook thuis aan het werk kunnen, maakt dat die controle lastig te bewerkstelligen is. We zullen voorlopig genoegen moeten nemen en gebruik moeten maken van logbestanden en vragenlijsten voor leerlingen. Een laatste punt betreft 'didactiek van de technologische tekortkoming': duidelijke minpunten van digitale (algebra) tools moeten niet worden 'goedgepraat' door middel van een didactische uitleg van het minpunt.

Samenvattend, nader onderzoek is nodig om deze bevindingen te repliceren en om ons begrip van de inzet van ICT voor algebra te verbeteren. Hierbij moeten we zeker rekening houden met nieuwe ontwikkelingen op het gebied van software en hardware, waarbij in het bijzonder de ontwikkeling van tablettechnologie er kansrijk uitziet.

.

Acknowledgments

This thesis marks the end of four years of research into the use of ICT for algebra. It has been a wonderful experience, albeit also a very tiresome one, especially because of the 0.6 fte. allocated per year. The research was carried out besides my ‘regular’ job at St. Michael College in Zaandam. I thank my school for allowing me to do this research.

I had a wonderful relation with my daily supervisor Paul Drijvers and I cannot express enough gratitude for that. I especially loved the way in which we complemented each other with me being the rogue agent without structure, and you being the structured scientist. I have learned a lot from you and I sincerely hope we can continue to work together in the field of ICT and mathematics. I’m sure we will.

Of course, supervision of my research was strengthened by another driving force behind the Institute, supervisor Jan van Maanen. Many thanks for your critical and encouraging remarks.

I thank the DUDOC Program and all the people involved in it (Tjeerd Plomp, Marie-Christine Knippels, Gjalt Prins, all fellow-Dudoccers, etc etc) for the opportunity to carry out this study. I thank all the (anonymous) reviewers who have made sure that all the articles that were submitted and/or accepted improved during the review process, and also all the colleagues whom I’ve met at conferences and who have sharpened my thoughts on so many things in the study. Also a big thank you to Nathalie Kuijpers, Alison Clark-Wilson and Miranda Theodore for proof-reading some chapters.

Many thanks to Alexander Rozbitzsch for his comments on multilevel analysis. Professionally, I would like to thank Nathalie Kuijpers and Betty Heijman for their work in Framemaker and in finishing the layout of the document.

There are many tools that I have used during the study. I thank all the designers and programmers. In some chapters I give the credit that is due. In particular I would like to thank Peter Boon for designing such a wonderful digital environment for mathematics. I thank my colleagues at the FIsme for their inspiration, but also all the people who were curious and interested in my research. This motivated me a great deal. Although in the end datamining was not used for the data analysis in the research, the meetings with Diederik Roijers on experimental methods in education were inspiring, as well as the work of Claudia Corriveau when she visited the institute. The research would not have been possible without hundreds of students, dozens of teachers, expert researchers and many schools. I can´t thank them all here but my thanks go out to all those who participated.

Finally I would like to thank my wife Miriam, and my children Kieran, Skye, Fionn, Bran and Alec. I know they sometimes had to cope with my absence, physically and/or mentally, and all the low and high points. Without their support I would not have finished this document. I can finally finish all the baseboards!

Curriculum Vitae

Christian Bokhove was born on September 16th, 1975 in Amsterdam, the Netherlands. After attending secondary school in Enkhuizen he obtained his B.Ed and M.Ed in mathematics education in the years that followed. Since 1998 he is a maths teacher at St. Michael College, a secondary school in Zaandam, the Netherlands. He also taught computer science and was head of ICT at that same school. In these years, he participated in many projects at the Freudenthal Institute for Science and Mathematics Education (FIsme) involving mathematics and ICT, such as Wisweb, WELP and Galois. In 2007 he became a Ph.D. student at the FIsme under the wing of the DUDOC Program, a program that aimed at allowing teachers to perform educational research in academics. He has given numerous presentations for education practitioners and at national and international conferences.