

EFFECTS OF A DIGITAL INTERVENTION ON THE DEVELOPMENT OF ALGEBRAIC EXPERTISE¹

Christian Bokhove

Paul Drijvers

Freudenthal Institute, Utrecht University

c.bokhove@uu.nl

p.drijvers@uu.nl

In this article we report on the effects of a digital intervention on the development of algebraic expertise of 17-18 year old students in the Netherlands. The question to be answered was whether the intervention would be effective and what factors influenced the outcome. With notions of formative assessment and symbol sense as guiding theoretical concepts, the intervention's design principles included the concepts of crises, formative scenarios and feedback. The intervention aimed to improve algebraic expertise and was deployed in fifteen grade 12 mathematics classes in nine schools. Data included results from pre- and posttests, scores, questionnaires and log files of the students' digital work, and responses to a student survey. Results from the effect study, analyzed with multilevel models, showed that the intervention was effective in improving algebraic expertise. Factors that significantly contributed to the posttest score were pretest results, the amount of time invested in digital self-tests and attitude towards mathematics. The intervention's success was not significantly influenced by other variables. We conclude that these types of intervention have a potential for the acquisition of versatile algebraic expertise.

algebraic expertise, assessment, digital intervention

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 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. 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 seems to focus on procedural skills rather than on conceptual understanding. 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.

THEORETICAL FRAMEWORK

As the study aims to design an online 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.

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. Hattie and Timperley (2007) distinguish four feedback types and investigate the effectiveness of (combinations of) feedback types. In this study, these different faces of feedback form the basis for the intervention's main design principles.

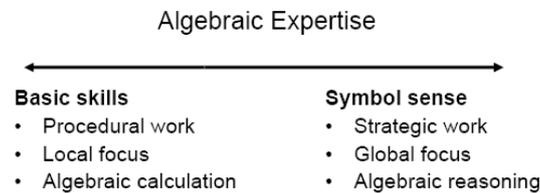


Figure 1. Algebraic expertise as a dimension (Drijvers, Goddijn, & Kindt, 2010)

The main aspect of *algebra didactics* is the view on 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 (Arcavi, 1994, 2005) involves strategic work with a global focus and emphasis on algebraic reasoning. In this study, the online activities are meant to offer both procedural and symbol sense opportunities.

This framework is helpful to better articulate the study's 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?

METHODS

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, 2010). 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, which we report on here, concerned a large-scale classroom experiment (N=324).

Four categories of tasks were included, characterized by the following examples:

1. Solving equations with common factors:
Solve $(x^2 - 7x + 12) \cdot (8x - 11) = (x^2 - 7x + 12) \cdot (3x + 14)$
2. Covering up sub-expressions (Wenger, 1987):
Solve for v : $v \cdot \sqrt{u} = 1 + 2v \cdot \sqrt{1 + u}$
3. Resisting visual salience in powers of sub-expressions:
Solve $(x - 3)^2 + 4 = 40$

4. Recognizing ‘hidden’ factors (Tempelaar, 2007):

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

The intervention² consists of a pen-and-paper pretest, four digital modules d1-d4, each covering one of the above item categories, digital diagnostic module d5, digital test module d6 and a pen-and-paper posttest (Figure 2, upper part). Cronbach’s alpha values show that all parts of the intervention are reliable. Special attention is paid to the feedback, and its timing and fading in particular.

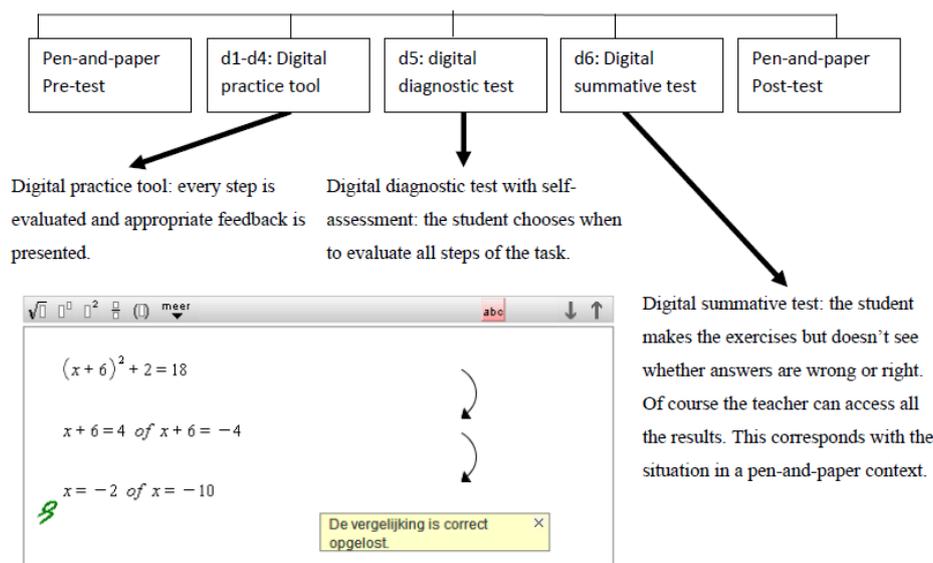


Figure 2. Outline of the formative scenario underpinning the intervention

Participants were 324 12th grade 17-18 year old students from fifteen classes from nine Dutch secondary schools. 43% of the participants were female and 57% were male. The total time needed to complete the module was about six hours work, excluding pre- and posttests. Teachers were supported by mailings and by the project website www.algebrametinzicht.nl.

Data includes the students’ general characteristics, including exam results, pre- and posttest results, scores and log files of the digital work, pre questionnaire which probed attitudes and behaviors concerning mathematics, post questionnaires addressing the students’ motivation and the way they perceived the intervention in the classroom. In this article only a few variables from the student questionnaire are included in the analysis.

Two analyses were carried out to answer the research questions. For the first questions concerning effect, we used t-tests. For determining predictors explaining the effect, we used multilevel analysis (Snijders & Bosker, 1999), using MLwiN 2.22 as software tool.

RESULTS

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. 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 ($M=-1.493$, $SE=2.339$), $t(285)=-20.602$, $p<.001$, $r=.773$, which is a large effect, too. The scores did not show a ceiling effect.

To investigate differences between the different types of items, Table 1 shows the fraction of correct answers for all students, all parts of the digital module and for each of the four item categories. The improvements are significant across all four categories.

Table 1. 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*** (3)
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 (4)
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.

Table 2 shows the results of the multilevel analysis. These results imply that students with an average pretest score, who spend an average amount of time in d5 and d6, 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 pretest score, the posttest 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, and for every hour spent in part d6 an additional 7.282 points ($SE=4.631$, $p<.05$). 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. These attitudes could potentially account for significant point differences in test scores between individuals.

CONCLUSION

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.

Table 2. The multilevel model

	Null model	Model 1a	Model 1b	Model 2a	Model 2b	Model 4 ⁽¹⁾	Model 5
<i>Fixed</i>							
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)***
<i>Random</i>							
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) **
<i>Model fit</i>							
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
<i>Df</i>		2	1	4	2	1	2
<i>P</i>		<i>p</i> <0.001	<i>p</i> <0.001	<i>p</i> <0.01	<i>p</i> <0.001	<i>p</i> <0.001	<i>p</i> <0.001
Reference		M0	M0	M1b	M1b	M2b	M4
N	286	286	286	286	286	286	286

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.

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, 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.

NOTES

1. An extended version of this paper has published in Computers and Education.
2. The module (in Dutch) is available through <http://www.fi.uu.nl/dwo/voho>

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.
- 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.
- 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.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- 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.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81-112.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23-31.
- 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.

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

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

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

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

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.