

Digital Tools for Algebra Education: Criteria and Evaluation

Christian Bokhove · Paul Drijvers

Published online: 30 April 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract In the Netherlands, as in many other countries, the algebraic expertise of students graduating from secondary education is an issue. The use of digital tools for algebra education is expected to change epistemologies, activity structures and student achievement. Therefore, a study was set up to investigate in what way the use of ICT in upper secondary education might enhance the algebraic expertise of students. One of the first decisions to be made concerned the choice of appropriate digital tools. This paper describes the process of designing and using an instrument for evaluating digital tools. The conceptual framework guiding this process includes notions on symbol sense, instrumental genesis and formative assessment. The evaluation instrument is designed through a Delphi method and provides a blueprint of tool features that are relevant for the purpose of this study. The results show that such an evaluation instrument is valuable both for choosing appropriate digital tools and for making concrete the aims and expectations that researchers have on the issue of integrating technology in algebra education. The final instrument is presented and illustrated through examples implemented in different digital algebra tools.

Keywords ICT tools · Algebraic skills · Evaluation instrument · Tool use · Assessment

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,

C. Bokhove (✉) · P. Drijvers
Freudenthal Institute for Science and Mathematics Education, Utrecht University, Aïdadreef 12,
3561 GE Utrecht, The Netherlands
e-mail: c.bokhove@fi.uu.nl

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 and Chick 2000, p. 216).

The above quote raises the issue of technology in algebra education. Educational technology plays a twofold 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 and 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 paper.

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 et al. 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 and 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, i.e. 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 paper 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 and Trouche 2005) showed that instrumental genesis is a time-consuming and lengthy process. Adapted form of Chevallard’s (1999) framework, Kieran and 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, i.e. 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 and Simpson, 2004; Hattie and Timperley 2007; Nicol and MacFarlane-Dick 2006). The feedback can be direct (Bokhove, Koolstra, Heck and 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 and 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 paper 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 (Hearshaw et al. 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 and 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 et al. 2004; Nicaud et al. 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 and MacFarlane-Dick 2006) and the 11 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 and Boldy 2007), a research study on digital assessment (Jonker 2007), the Freudenthal

Institute's mathematics wiki¹ 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 five 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 two out of seven 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, Schellens, Valcke and 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's² 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.

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

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

4 Results

4.1 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 ten 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 1 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 1.

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 2 shows the top five of important criteria with their accompanying weight means, as well as the bottom five.

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.

4.2 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

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 |

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 |

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

| Item category | Tool | | | | | | |
|------------------|--------|--------|----------|--------|--------|------------|---------|
| | WIMS | STACK | MAPLE TA | DME | WIRIS | ACTIVEMATH | 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 |

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 2 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 3 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 1) 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.

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

5.1 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, 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.

This criterion is exemplified in Fig. 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. 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 Fig. 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 five 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.

5.2 Tool Criteria Exemplified

Evaluation criterion ten 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 and Drijvers 2006), as well as to the notion of instrumental genesis described in the conceptual framework. The example in

³ <http://www.fi.uu.nl/~christianb>.

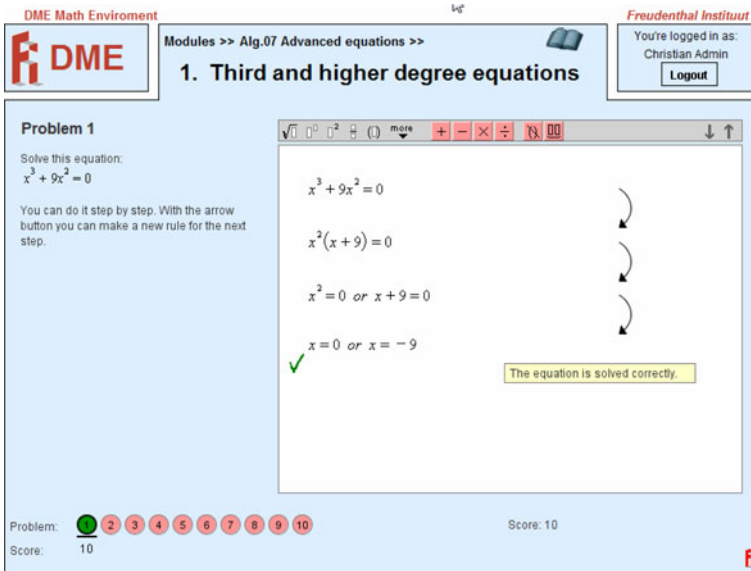


Fig. 1 Stepwise strategy in the DME

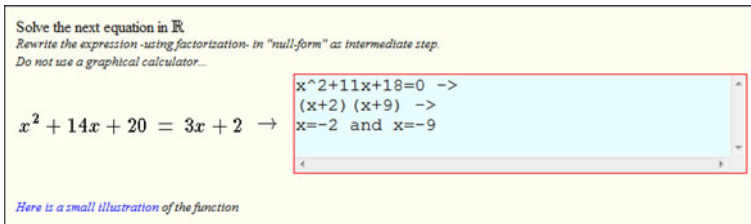


Fig. 2 Steps in Wims

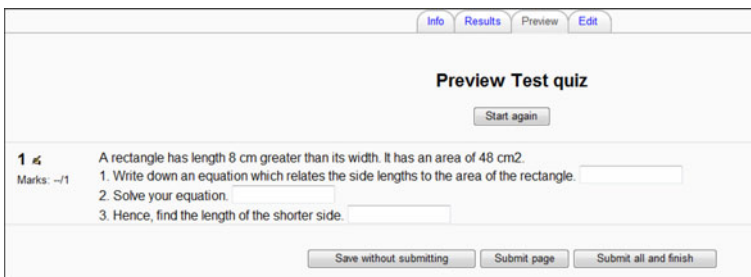


Fig. 3 Example of a multi-component question that can be authored in STACK and Moodle

Fig. 4 shows WIRIS providing an intuitive interface with notations that resemble conventional mathematics representations.

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

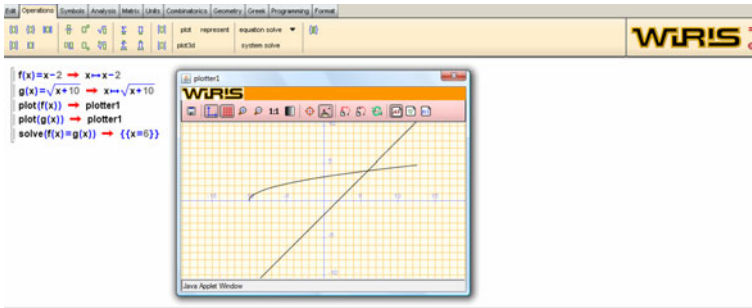


Fig. 4 Wiris’ graphical user interface

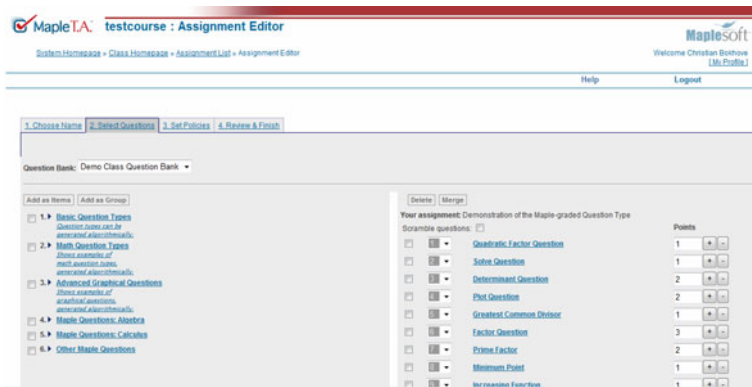


Fig. 5 Question banks in Maple TA

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 i.e. contend that digital algebra tools need to provide easy-to-use question management facilities.

A relevant example is provided in Fig. 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.

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, Fig. 6 shows that Webwork—in use at many universities in the United States—comes with a very large database of questions at university level.

5.3 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).

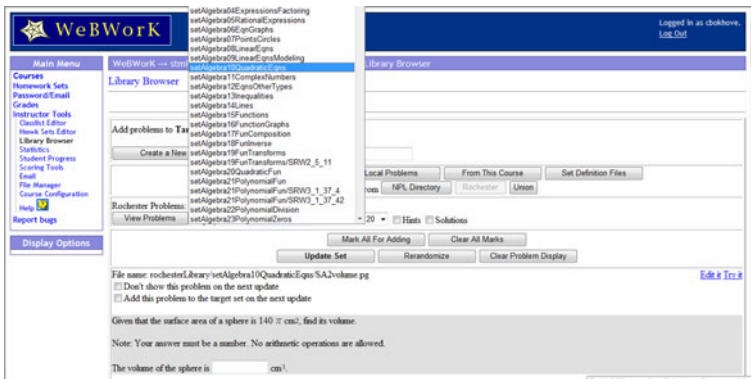


Fig. 6 A massive amount of readily available content in Webwork

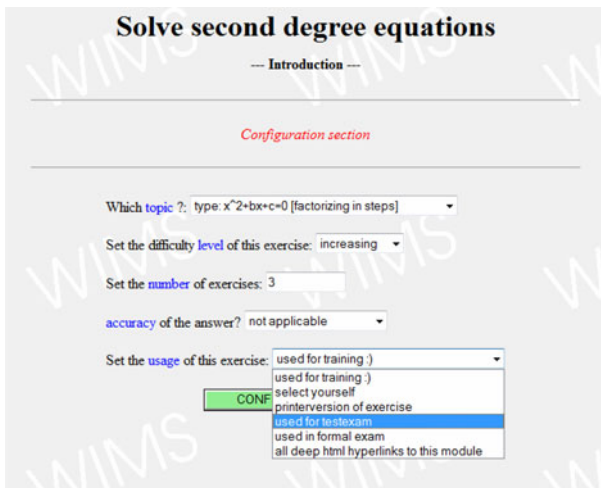


Fig. 7 Assessment modes in Wims

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, Fig. 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).

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

Main Page | Search | Notes | My Profile | Tools | Print | Logout | Help

Learning context adaptivity 17/17

ActiveMath Content Examples > Test Pages > Learning context adaptivity

The learningcontext sensitive AND

The learning context of a user might influence the appearance of the formulas. This behaviour is now realized and can be tested here. The following logical expression A and B will appear in textual notation if the user is higher_education. Otherwise the hat symbol is used. You can check that by logging in and setting the user preferences to learning context "school, higher education".
 $X \wedge Y$

Subformula semantics

This is an example formula. It applies the operator dubb_plus_minus to the three variables x,y and z. its semantics are add the first two and subtract the third. The subformulas "+" and "-" refer to the arith operators. (brackets still refer to dubb_plus_minus) $(x+y)-z$

Fig. 8 Adaptive content in Activemath

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.

6 Conclusions and Discussion

In this paper 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.

Acknowledgments We thank Platform Bèta Techniek and the Dudoc programme for funding this project UU4, as well as the respondents in the modified Delphi process, and Chris Sangwin in particular, for their participation in the rating process. We also thank Jan van Maanen for his valuable comments on this article.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix 1

See Table 4.

Table 4 List of criteria, weights and rankings

| Criteria | Average Weight | Rank |
|---|----------------|------|
| <i>1. Algebra category</i> (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 |
| <i>2. Tool category</i> (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 |
| <i>3. Assessment category</i> (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 |
| <i>4. General category</i> (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 2

See Table 5.

Table 5 Tool data sheet

| Name | Date | Version | URLs | Features, Characteristics |
|--|------------|----------------------|---|--|
| WIMS | 12-01-2009 | 3.57 (3.64 released) | http://wims.unice.fr/ For Dutch content: http://wims.math.leidenuniv.nl/wims | 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 |
| STACK | 12-01-2009 | 2.0 | http://www.stack.bham.ac.uk/ | 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 |
| Maple TA | 12-01-2009 | 4.0 | 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 | 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 |
| Digital Mathematical Environment (DME) | 12-01-2009 | Version 12-01-2009 | Several tests were made in a special environment for secondary and higher education topics: http://www.fi.uu.nl/dwo/voho | 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. |
| Wiris | 12-01-2009 | 2 | http://www.mathsformore.com/ , http://www.wirisonline.net/ | 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 |
| Activemath | 12-01-2009 | 1.0 | http://www.activemath.org/ | 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 |
| WebWork | 12-01-2009 | 2.4.1 | http://webwork.maa.org/moodle/ | 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 |

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), 245–274.
- Bedaux, J., & Boldy, M. (2007). *Ontwerpversie Aanbevelingen Toetsstandaarden Wiskunde*. Retrieved October 5, from 2007 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.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–144.
- Bliss, J., & Ogborn, J. (1989). Tools for exploratory learning. *Journal of Computer Assisted Learning*, 5, 37–50.
- Bokhove, C., Koolstra, G., Heck, A., & Boon, P. (2006). *Using SCORM to monitor student performance: Experiences from secondary school practice*. Retrieved March 28, 2008, from <http://mathstore.ac.uk/articles/maths-caa-series/apr2006/index.shtml>.
- Chevallard, Y. (1999). L'analyse des pratiques enseignantes en théorie anthropologique du didactique. *Recherches en Didactique des Mathématiques*, 19, 221–266.
- 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.
- 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. *Nieuw Archief voor Wiskunde*, 5(7), 198–203.
- Engineering Council (2000). *Measuring the mathematics problem*. Retrieved January 10, 2010, from <http://www.engc.org.uk/ecukdocuments/internet/document%20library/Measuring%20the%20Mathematic%20Problems.pdf>.
- Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, 1, 3–31.
- 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.
- 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 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.
- Hoch, M., & Dreyfus, T. (2004, July). *Structure sense in high school algebra: the effect of brackets*. Paper presented at the PME28, Cape Town, South Africa.
- Jonker, V. (2007). *Digitale authenticiteit toetsen van wiskunde*. Retrieved March 28, 2008, from http://www.fi.uu.nl/wisweb/kloof/digitaaltoetsen/documents/20070329_klo_authentiek.pdf.
- 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: National Academy Press.
- 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.
- National Council of Teachers of Mathematics. (2008). *The Role of Technology in the Teaching and Learning of Mathematics*. Retrieved October 24, 2008, from <http://www.nctm.org/about/content.aspx?id=14233>.
- National Mathematics Advisory Panel. (2007). *Preliminary Report*, from <http://www.ed.gov/about/bdscomm/list/mathpanel/pre-report.pdf>.

- 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.
- 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.
- 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*, 42(2), 215–224.
- Tempelaar, D. (2007). Onderwijzen of bijspijkeren? *Nieuw Archief voor Wiskunde*, 8(1), 55–59.
- 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 June 2, 2008, from <http://tlc.uwaterloo.ca/is301a/groups/sample/feedback.html>.
- Van Streun, A. (2000). Representations in applying functions. *International Journal of Mathematical Education in Science and Technology*, 31(5), 703–725.
- Verillon, P., & Rabardel, P. (1995). Cognition and artifacts: A contribution to the study of thought[t] in relation to instrumented activity. *European Journal of Psychology in Education*, 9, 77–101.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Zorn, P. (2002). *Algebra, computer algebra and mathematical thinking*. Retrieved July 4, 2009, from <http://www.stolaf.edu/people/zorn/cretepaper.pdf>.