Preprint – to appear in Proceedings of 13th International Congress on Mathematical Education, Hamburg, 24-31 July 2016

DEVELOPMENT OF STACK ASSESSMENTS TO UNDERPIN MASTERY LEARNING

A. Rasila and C.J.Sangwin
Aalto University, Finland. University of Edinburgh, United Kingdom

This paper reports significant developments in computer aided assessment (CAA) which avoid the well-known problems with multiple choice testing. We describe our implementation of this CAA to underpin a major national project which aims to teach both mastery and problem solving skills in undergraduate mathematics.

MATHEMATICAL PROFICIENCY AND MASTER LEARING OF SKILLS

Mathematical proficiency is a set of interrelated skills, e.g. Kilpatrick, Swafford, & Findell, (2001) identified conceptual understanding, procedural fluency, strategic competence, adaptive reasoning and productive disposition as five important strands. Different areas of mathematical proficiency require different learning strategies, e.g. conceptual and procedural abilities are typically learned though conscious practice of exercises. To develop logical thinking skills normally require significant self-reflection and discussions with a human teacher. Similarly, advanced problem solving skills usually require a significant amount of human-to-human interaction. These skills form a loose hierarchy: weak basic conceptual and procedural skills seriously hinder a student's ability to formulate and solve mathematical problems. Weak problem solving skills do not necessarily stop a student from high achievement with procedural tasks. We separate mathematical skills (loosely) into two groups: mastery and problem solving skills. The essential distinction is that mastery skills are rarely the end goal, rather they form part of a subsequent wider task. Problem solving skills can often only be evaluated in terms of better–worse rather than right–wrong.

The starting point of our endeavor is the remarkable observation of Bloom (1984) that students taught by an individual tutor achieve test scores which are two standard deviations better than students who attend traditional classroom teaching. Mastery Learning (ML) is an educational philosophy proposed by Bloom as a partial solution to the problem of finding resources for individual tutorials. In ML students are regularly tested by using formative tests and students are required to demonstrate a correct answer to 90% of the test problems, i.e. demonstrate “mastery”. When a student falls short of mastery further teaching and testing is repeated, several times if necessary. In traditional settings such extensive testing is impractical, but with automatic online assessment this practical barrier is removed. However, mastery learning can lead into surface-oriented learning strategies, especially if formative testing is mainly based on multiple choice questions.

In this paper we report significant developments in computer aided assessment (CAA) which avoid the well-known problems with multiple choice testing, and our implementation of such online assessment to underpin a major national project which aims to teach both mastery and problem solving skills for university mathematics.

CONTEMPORARY ONLINE ASSESSMENT

CAA is now a well-established part of many mathematics courses at university. CAA has moved well beyond multiple choice questions (MCQ) in mathematics: a recent survey was published as (Sangwin, 2013, Chapter 8). Most online assessment engines for mathematics now have the following characteristics. Under the control of both teacher and student, the software selects a question for a particular student, e.g. a fixed list, a random selection from a pool, or an adaptive system chooses a question based on an internal student model. Often a random version of a question is generated using computer algebra from a question template in a structured mathematical way, and this includes a full worked solution which reflects this randomisation. The student solves the given problem, perhaps using a pen and paper in the traditional way, or using computer algebra as a tool. Systems vary on precisely how students enter their answer, with the most popular options being a typed linear syntax or a drag and drop equation editor. The software automatically establishes the mathematical properties of that answer using a computer algebra system (CAS). A prototype property is algebraic equivalence with the teacher's answer, and another property is that the student's answer is written in a particular form, e.g. factored. On the basis of these properties outcomes, such as formative feedback, can be generated. The system stores data on all attempts at one question, or by one student, for later analysis by the teacher.

Each of these phases can be (and indeed is) subject to specific research studies. Together they constitute a complex amalgam. For example, how students learn mathematical notation, and interact with software such as CAS or CAA is an important topic. In evaluating the effectiveness of any tool there is a subtle interplay between the features the software offers, the intentions of the teacher in deploying them, and the use to which they are put by students. Our practical research follows the design research paradigm.

STACK is a typical example of contemporary assessment software for mathematics, with most of the features described above. In the year ending 1st April 2015 STACK was downloaded 10168 times and data from a survey during May 2015 showed that STACK is currently being used in eight languages, with groups of students ranging from small groups of 20-90 students up to teachers with cohorts of up to 1500 students. STACK is being used for open access practice, for formative quizzes, for quizzes which contribute to the final grade and also for timed summative examinations.

THE FINNISH ABACUS MATERIAL BANK PROJECT

Abacus, based at Aalto University Finland, is a major national project to use computer aided assessment to implement mastery learning, based on problems implemented predominantly using STACK. In addition to these "mastery" problems, the students are also given wider problem solving tasks and projects, which are solved in groups under the supervision of a teaching assistant. The project is motivated by the data from previous usage of STACK at Aalto University, which demonstrates that the students who start working with STACK-based assignments, on average, reach the mastery level by training with the system, even when there are no material rewards for it. The theory of ML also explains our observation that working with STACK improves examination scores, in particular, for lower performing students, see Rasila, Havola, Majander, & Malinen, (2010).

A serious obstacle to the use of STACK has been the lack of available learning materials and associated support services. For this reason, many teachers opt for commercial solutions, despite their relatively high prices and other shortcomings such as lack of support for local language and curriculum. More commonly, teachers simply decide not to use computer-based materials and learning environments at all. In order to address these obstacles in higher education, the Finnish universities have founded a mathematics e-learning cooperative material pool Abacus, which ultimately aims at establishing a national and even international standard for open source e-learning software in mathematics and STEM education. The initial partners of the material pool were the seven Finnish universities with MSc programs in engineering, namely, Aalto University, Åbo Akademi University, Lappeenranta University of Technology, Tampere University of Technology, University of Oulu, University of Vaasa and University of Turku. Since the project inception Metropolia University of Applied Sciences and University of Helsinki have joined the pool. Negotiations with several partners are ongoing, and the project is actively looking for further international partners. The project involves training a substantial number of staff, and the development and deployment of comprehensive learning materials nationally in Finland.

STACK was originally developed for mathematics, but it is also very useful for learning other science and engineering topics involving mathematical expressions. At Aalto University, STACK has been used in Physics, Statistics, Logistics and Control Engineering. There has even been a project to develop a STACK-based system for medical dose calculations for nursing students. As our experience and confidence develops, we identified certain technical shortcomings, such as a lack of support for SI standard measure units. This has been recently solved by development work done at Aalto University as part of ongoing design research cycles. See also Rasila, Malinen, & Tiitu (2015). More substantial developments include expanding STACK to adaptive testing, in which the system develops a model of the learner, and on the basis of this selects different questions. We also intend to develop game-like situations which develop mastery, and which are underpinned by STACK's mathematical sophistication.

There are also deeper and more philosophical issues. E.g. the main obstacle to solving mathematical problems in applied science is not in solving the problem using mathematical techniques, but rather in seeing where the mathematical problem is, i.e. modelling. When the problem is stated already formulated as a mathematical problem assignment the deepest part of the assignment is has already been solved, but the same shortcoming arises in traditional pen-and-paper exercises assignments.

One area identified by our research as a central activity which bridges skill-based tasks and problem solving is reasoning by equivalence. Reasoning by equivalence is a formal symbolic procedure where an algebraic expression is manipulated to generate a new and equivalent expression, e.g. a term within an algebraic expression is identified and then replaced by an equivalent term. Reasoning in this way we generate a new problem having the same solutions, and we continue until a "solved" form is reached. A recent survey of the extent to which final high school mathematics examination questions could be automatically assessed found that approximately a third of the method marks are awarded for reasoning by equivalence. There are other forms of reasoning, e.g. calculus operations to find extreme values, estimation and implication arising from inequalities, but reasoning by equivalence is of central importance. Furthermore, reasoning by equivalence forms the basis for formal proof, e.g. proof by induction and some proofs in real analysis. Hence it is a stepping stone to genuine problem solving. See Sangwin (2015) for a selection of examples, and a discussion of the issues involved.

When undergraduate students were asked to solve equations such as (x+5)/(x−7)−5=(4x−40)/(13−x), they typically reason by equivalence working line by line. Most students need many lines of working, for this example typically about a dozen. This observation justifies the need for a reasoning by equivalence interface, not just relying on the final answer. But, students do not pay attention to domains of definition or explicitly indicate which steps guarantee equivalence of adjacent lines and which do not. This is problematic because elementary algebra contains a number of subtle "traps", including division by zero, or gaining/loosing solutions by squaring/square rooting both sides of an equation, see (Sangwin, 2015).

Reasoning by equivalence software already exists, for example, MathXpert of Beeson (1989). However, here a student indicates what they would like to do and a CAS undertakes the calculation for them. We have already implemented a basic equivalence reasoning engine for STACK, but this merely replicates current practice, as do Heeren and Jeuring (2014). Indeed, there is much to be criticised in the presentation of logical reasoning in elementary algebra textbooks during step by step solving. The design of an automatic assessment system provides a unique opportunity to undertake design research in this area from a novel angle, and the tools we develop now will influence the way future students and teachers understand algebraic reasoning.

CONCLUSION

STACK is a suitable platform for wide range of applications to underpin mastery learning of traditional procedural skills. The existing features allow us to support mastery learning without relying on multiple choice questions. The recently implemented developments significantly expand the range of assessments which can be automated in a way which respects the intentions of the question author and teacher, and hence open a door to novel and innovative teaching strategies. These are being implemented in an ambitious project on a national scale in Finland.

References

Beeson, M. (1989). Logic and computation in Mathpert: An expert system for learning mathematics. In Kaltofen, E. and Watt, S. M. (Ed.). *Computers and mathematics '89.* (pp. 202–214). New York/Heidelberg/Berlin: Springer-Verlag.

Bloom, B. S. (1984). The 2 sigma problem. *Educational Researcher*, 13(6), 4–16.

Heeren, B., & Jeuring, J. (2014). Feedback services for stepwise exercises. *Science of Computer Programming*, 88(1), 110–129.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: helping children learn mathematics*. Washinton D.C: National Academy Press.

Rasila, A., Havola, L., Majander, H. & Malinen, J. (2010). Automatic assessment in engineering mathematics: evaluation of the impact. In *Reflektori 2010: Symposium of engineering education. Teaching and Learning Development Unit,* 37–45.

Rasila, A., Malinen, J., & Tiitu, H. (2015). Automatic assessment and conceptual understanding. *Teaching Mathematics and its Applications*. 34(3), 149–159.

Sangwin, C. J. (2013). *Computer aided assessment of mathematics*. Oxford: Oxford University Press.

Sangwin, C. J. (2015). An audited elementary algebra. *The Mathematical Gazette*, 99(545), 290–297.