

Who uses STACK? A survey of users of the STACK CAA system, May 2015

C. J. Sangwin

Mathematics Education Centre, Loughborough University, Leicestershire, LE11 3TU

C.J.Sangwin@lboro.ac.uk

June 2015

1 Introduction

STACK is online software which implements a question type designed to assess mathematics and is described in more detail in Section 2 below. This document contains the results of a survey of users of STACK. STACK is open source software which is freely available. In the year ending 1st April 2015 STACK was downloaded 10168 times from the Moodle plugin database. However, it is difficult to know how many people actually use STACK and this does not equate to numbers of live servers. Since users are at liberty to download and use the software the developers have no intrinsic mechanism for tracking usage, which a commercial licence might provide. The goals of the survey are (i) to provide a minimum benchmark for the number of STACK users, (ii) understand how STACK is being used in extent and purpose of use, (iii) provide users with a mechanism to suggest changes or new features.

2 Background to STACK

Valid assessment is the cornerstone of effective teaching. Automatic assessment is commonly associated with multiple choice questions (MCQ) or similar provided response question types. Such question types are referred to as *objective* because the outcome is independent of any bias by the assessor. MCQs have been criticized for many years: [4] claims they “*favour the nimble-witted, quick-reading candidates who form fast superficial judgements*” and “*penalize the student who has depth, subtlety and critical acumen*”. Authors such as [3] and [6] provide evidence that the MCQ format itself has inherent gender bias, although we have been unable to replicate this effect.

For mathematics MCQs are particularly problematic as the relative difficulty of a reversible process, e.g. integration compared to differentiation, is markedly altered in different directions. As [12] says

The strategic student does not answer the question as set, but checks each answer in reverse. Indeed, it might be argued that it is not just the strategic, but the *sensible* student, with an understanding of the relative difficulties of these processes, who takes this approach. This distortion subverts the intention of the teacher in setting the question, so that we are not assessing the skill we wish to assess.

During the 2014-2015 academic cycle Ian Jones and I undertook a comparative experiment to test the hypothesis that when faced with a question involving the inverse direction of a reversible mathematical process, students solve a multiple choice version by verifying the answers presented to them by the direct method, not by undertaking the actual inverse calculation. This experiment compared students' answers

on STACK questions with responses to stem-identical multiple choice questions. The findings supported this hypothesis: overall scores were higher in the multiple choice condition compared to the constructed response condition provided by STACK, but this advantage was significantly greater for questions concerning the inverse direction of reversible processes compared to those involving direct processes. E.g. when asked to factor polynomials the evidence supports the hypothesis that student expand out the answers rather than actually factoring the given expression. This reduces the validity of MCQ for assessing such reversible processes and justifies the rationale for the design and development of STACK. While the term ‘objective test’ is often taken to be synonymous with MCQs there are many situations, particularly in mathematics, where the properties of an answer provided by a student can be established objectively and automatically. For these reasons, where practical in mathematics, automatic assessments which accept a mathematical expression from the student are to be strongly preferred.

Motivated by the need to assess answers to students’ work, the STACK online assessment system uses computer algebra to support the assessment process. It also

- generates random versions of questions in a structured mathematical way;
- accepts answers from students which contain mathematical content, rather than MCQs;
- establishes the mathematical properties of those answers;
- generates outcomes (including feedback) which fulfill the purposes of formative and summative assessment;
- stores data on all attempts at one question, or by one student, for analysis by the teacher.

STACK uses the computer algebra system Maxima to support the mathematical processes, and Moodle to provide a context in which the activity takes place. A demonstration server is available at <http://stack.bham.ac.uk/moodle>.

The answer to a STACK question consists of a mathematical expression and students have to enter their answer into a machine. To answer a STACK question a student must use a typed linear syntax, e.g. $2x^3$ can be entered as $2*x^3$. I believe that undergraduate students reading mathematics, science and engineering should learn how to type simple mathematical expressions into a machine accurately using a relatively strict syntax. This issue is discussed in more detail in [14].

Once the student has entered a valid expression the computer algebra system Maxima is used to establish mathematical properties relevant to the question. The prototype tests establish that the student’s answer is (i) algebraically *equivalent* to the correct answer and (ii) in the appropriate *form*, (e.g. factored). That said, the use of computer algebra is a long way from string matching, or the use of regular expressions. It is the relevant properties of expressions which matters and answers are often non-unique. Where the student’s answer does not satisfy all the properties the teacher is able to encode feedback. Potentially this is specific to the answer and directly related to possible improvement on the task. This is precisely the kind of feedback which research such as [5] has suggested is most effective. Unusually for CAA, STACK may include and display results of computer algebra calculations within such feedback which can be as detailed as appropriate to the situation. This is a particular distinguishing feature of STACK.

STACK is an optional question type for Moodle which was designed and developed by Chris Sangwin at the University of Birmingham starting in 2004. During 2011—2012 version 3 was written in collaboration with Tim Hunt at the Open University, with contributions from Matti Harjula at Alto University, Helsinki and Matti Pauna at the University of Helsinki. More details of the goals and design are given in [12]. It has also recently been incorporated into the ILIAS learning environment, see Section 8.

The screenshot shows a STACK question interface. On the left, a sidebar indicates 'Question 1' is 'Correct' with a mark of '1.00 out of 1.00'. The main area contains the question: 'Solve the following inequality: $x^2 - 2 \cdot x - 3 \leq 0$.' Below this, it instructs the user to 'Write your answer as a logical sequence of inequalities, such as $5 \geq x$ and/or $x \leq 13$.' The user's input is shown as 'Solution : -1<=x and 3>=x'. A feedback message states 'Your last answer was interpreted as follows: $-1 \leq x \wedge 3 \geq x$ '. A 'Check' button is visible. Below the feedback, a green bar indicates 'Correct answer, well done.' and 'Marks for this submission: 1.00/1.00.' The solution steps are: 'To solve the following inequality: $x^2 - 2 \cdot x - 3 \leq 0$. Factorise to give $(x + 1)(x - 3) \leq 0$. The product will be ≤ 0 if $x + 1 \geq 0$ and $x - 3 \leq 0$. This solution can be entered as $x \geq -1$ and $x \leq 3$.'

Figure 1: A typical STACK question

3 Survey design

The survey of STACK users consists of two parts. The first is an online questionnaire which was publicised by directly contacting people who had corresponded with the developers about STACK in the past, and also by posting a request to participate to the Moodle “Mathematics tools” forum and a number of CAA email lists. The questionnaire was available for approximately three weeks, and for completeness is given in Appendix A.

The second part of the survey involved asking colleagues known to make substantial use to write short free-form case studies describing their use and experiences. These case studies may not be representative of small scale users, e.g. colleagues using STACK for one or two classes. They do illustrate the larger scale uses to which STACK is being put.

Note that for the purposes of this report students were not surveyed directly. This has been done in the past, almost annually by the author. Negotiating access to students at other institutions requires care, and interpreting their responses must be done in the context of how STACK is used for assessment as an overall part of their course. E.g. in the past students have voiced negative reactions to “STACK” because this was the primary formative assessment experience. Closer analysis was needed to separate the reaction some students have to challenging mathematical assessments from the details of STACK as an online system.

4 Results

There were 55 responses to the survey, all of whom agreed to participate. Given that the survey was selectively publicised to teachers, this is considered to be a significant response rate. Note also that STACK is designed with a client-server architecture, and so respondents represent teachers or system support staff. 15 respondents had not used STACK, and these people were directed online from Q2 to Q14. Their responses are collected in a separate case study in Section 9.

There were 40 participants who used STACK and who completed the substantial part of the survey. Percentages below are calculated on the basis of STACK users, not survey responses. STACK is currently being used in the languages shown in Table 1. Note that since some users reported more than one language the total exceeds the 40 participants. Also, some centres have more than one person responding to this survey. Nevertheless, STACK is being used in 8 languages.

Language	Number of responses	Years of use	Number of responses
		0	1
English	22	1	19
Finnish	8	2	6
German	5	3	5
Swedish	4	4	1
Portuguese	3	5	3
Spanish	3	6	1
Korean	1	7	1
Japanese	1	8	2
		9	1

Table 1: Language (left) and duration of use (right)

Type of use	Number (%) of respondents
Formative quizzes for registered students.	30 (75%)
Summative quizzes which contribute to a course mark.	30 (75%)
Online timed examinations.	8 (20%)
Open access practice site.	10 (25%)
MOOC or other open structured course.	4 (10%)

Table 2: Purposes of STACK use

Question 4 asked the participants “4. For how many academic cycles (years) have you used STACK?” Responses are also shown in Table 1. There are a small number of loyal users who have been using STACK for nearly a decade, indeed since version 1 of the software. Until recently STACK was retaining about one new user per year. Now there are also a large group of new users who have used STACK for one academic cycle. This is encouraging, but only a follow up survey will establish how many continue to use STACK in the medium to longer term.

Purpose, extent and frequency of use

Question 5 asked respondents to describe how they used STACK, responses enabled any number of uses to be selected from a list. Responses are shown in Table 2. 24 users indicated both setting formative quizzes for registered students and summative quizzes which contribute to a course mark. This corresponds with the design purpose of STACK. Eight people make use of STACK for online timed examinations. This is a change from previously reported use. Four respondents chose “other”, two of whom were using STACK only for evaluation or development at this stage. One respondent uses STACK for analysing physics lab data and the last uses STACK for “homework”.

Question 6 asked for the extent of use, e.g. approx number of students, frequency of quizzes. This is open ended, and more difficult to analyse. However, a range of different types of users can be found.

The first group of 9 people are using STACK for relatively small numbers of students, ranging from 20-90. The second group of 15 have 100-250 students, and are typically teaching large traditional university lecture courses. For this group the frequency of quizzes ranged from 2 per week to 2 per year, but one quiz every week or fortnight was the norm. The third group of 5 had much larger courses, ranging from 250 to “several hundred” students. One user has 1500 students and sets two quizzes a month.

What is clear from this data is (i) there is no typical use, and (ii) respondents were very likely to use STACK in more than one context, both in very small and in very large classes.

Q7 sought to gather how useful people found the sample materials. STACK is distributed with a “demon-

stration course” containing over 550 questions covering algebra, calculus, and “core pure” mathematics methods. Most of these questions generate random versions for the student and contain full worked solutions. 13 participants had not used these materials and a number of others were unaware they existed. The materials come assembled into quizzes and are distributed with the STACK source code as a single backed-up Moodle course. Below is a direct URL.

https://github.com/maths/moodle-qtype_stack/blob/master/samplequestions/STACK-demo.mbz

7 participants used the sample materials as a way to learn how to author STACK questions. 9 participants made extensive use of these materials, and valued them greatly.

They are extremely useful. I can't imagine making questions without using them as a starting point.

Yes, regularly. These are indispensable for reference when writing questions. If I find something similar enough, I will take a sample question and adapt it for my own students.

At least one translated some of the examples and “*adopted the ideas behind them for new exercises*”. The question of how to create questions which can be used in more than one language has been raised before and is a possible future project.

Likes and dislikes

The next two questions asked about likes and dislikes. The mathematical nature of the answer processing is seen as the strongest feature and was cited by many respondents. This includes the ability to generate random questions, evaluate mathematical answers and give answer-specific feedback.

Using random variables, questions can be attempted repeatedly without students repeating the same calculations. Some students learn more effectively this way - they jump straight into the question, get it wrong, read the feedback, then try it again until they get it right. Power of using Maxima, high level of control over all aspects of question.

Some respondents cited the value of practice when learning mathematics. “*STACK makes students practice a lot. Students like working with it and especially the feedback.*”

Frees the teacher from tedious, mechanical marking, and counteracts problems with students copying the solutions from each other. The students appreciate immediate feedback on correctness, and feedback that helps them solve the problem correctly.

While some people found it easy to integrate to online courses utilizing Moodle, this is clearly a barrier for people who use another learning environment. Open source code is valued by a number of users.

New features requested

A number of respondents asked either for features which already exist, or for better documentation. These issues are clearly related!

The following features, listed approximately with the most popular first, were requested which are already on our list for future development.

- Better editing of response trees, perhaps with a text-based editing format to enable copy and paste from one question to another.
- Better support for graphics, including dynamic graphic input such as that provided by GeoGebra. Note prototype code for this has been developed.

- The ability for authors to load Maxima packages. As frustrating as this is for question authors this is not a feature which will be provided in the near future as there are serious security concerns about which commands teachers can and cannot use.
- Support for physical units. This is an important library I would like to add.
- A wider variety of input types e.g. including multiple choice and a virtual keyboard (like DragMath).

Three respondents asked for STACK to be easier to install, and to be faster. This is a positive change from previous surveys, and it does depend on the respondent's level of autonomy (and support) on the server they use.

A number of users asked for better support for drawing in HTML content from the web. This has already been addressed and code is currently being tested for STACK 3.4. See Section 10.1 below on Question Blocks. One respondent complained that STACK can be inflexible, and they would like to be able to change the structure (i.e. the number of input / validation/ feedback cells) of a question depending on the responses given by the student. The current Question Blocks code does not support this, but it is a planned extension to include this behaviour once we are satisfied the new features are robustly engineered.

Some respondents asked for "adaptive testing" where one question is triggered by the current response, or a detailed student model. Adaptive tests should be constructed at the "quiz level" whereas STACK is a "question type". Since the STACK code base is not really at the appropriate level this is not a feature I plan to add soon. I would encourage people interested in adaptive testing to look at previous research in this area of mathematics education, such as [15] and [1].

A number of respondents asked for the ability to test student's work "step by step". Much of this work is referred to as *reasoning by equivalence* and I have prototype code to automate assessment in this way. Serious consideration has been given already to how the design of this might work. There are important interface problems to overcome. There are also mathematical difficulties, notably how domains are tracked, which (all) current CAS do not deal with in an ideal way. Some of the research underlying the design is reported in [13]. Maxima needs more comprehensive libraries for dealing with systems of inequalities and intervals of real numbers for this to work in a very robust manner, and this is a priority for future development.

The following features were requested by individuals,

- Assignment pdf export.
- Support for templates so that options don't have to keep being set for each new question. [A text-based authoring system might partially and better support this request.]
- The ability to control output without putting everything into a string to suppress Maxima simplification.

One respondent said

Frankly, I do not need more features. For me it's ok.

Intriguingly there were no feature requests which significantly change the design of STACK which are novel.

Question 11 asked *Have you modified STACK, or been involved in its development? If so how?* One respondent had contributed major new sections to part of the code, and four other people had made minor modifications, some of which remain local to their site. Five respondents had translated STACK to their own language and a further three people had posted bug reports. The remaining 30 respondents have not altered the code or reported bugs.

Bug reports, suggestions and feature requests can be posted online in two places. The first is the Moodle "Mathematics Tools" forum, <https://moodle.org/mod/forum/view.php?id=752> and the second is directly in the code repository on GitHub <https://github.com/math5/>

5 Case study: Loughborough University, UK

STACK has been in use at Loughborough University since October 2013. In these two academic cycles the use has steadily grown.

During the summer of 2014 existing paper-based diagnostic tests were automated as STACK questions, in the process moving away from a reliance on MCQ. Two tests were created, one assumed students had studied GCE Advanced Level mathematics and the other assumed only GCSE. The move away from a paper based format has the following effects:

- Students (potentially) had more flexibility in undertaking the test in their own time and place, although this brings with it the possibilities of plagiarism, impersonation and collusion.
- Students got an immediate result, whereas paper based tests needed to be processed on optical forms which was expensive in terms of staff time, and introduced a delay.
- Students had access to specific feedback and a worked solution after the due date.
- Staff members could access statistics and individual performance online.

The tests were used with four groups of engineering students (total of approx 800 students) and overall the test worked very well at a technical level. Analysis of the results pointed to some improvements in the questions, and some areas on which we need to focus more teaching for these students, e.g. partial fractions.

The course Mathematics for Chemical Engineers made use of STACK quizzes for (i) open access practice and (ii) five summative online quizzes each contributing 4% to the module mark. This is a 20 credit module running over two semesters in year 1. The cohort consisted of 107 students. Five open access practice quizzes were made available during semester one with a significant increase in participation compared with the last session. The practice materials contained random questions, formative feedback during the attempt and model answers at the end. 20% of the formal assessment for this module was given for results to five CAA tests. Students were given a week during which to sit the assessment. Each assessment was time-limited to 90 minutes and all quizzes were a mixture of STACK and MCQs. These ran with acceptable mark distributions and good participation levels, as in previous years.

The HELM project (<http://helm.lboro.ac.uk/>) was a major curriculum development project undertaken by a consortium of five English universities which ran from October 2002 to September 2005. The aim was to enhance the mathematical education of engineering undergraduates by the provision of a range of flexible learning resources, which included computer based assessments tied to 50 workbooks of approximately 50 pages each. These workbooks are widely used at Loughborough and engineering mathematics courses are often assembled using a selection of workbooks as required. A decade on these online assessments need updating and revising. Over the summer 2014 three the books were chosen,

- HELM 9 - Vectors
- HELM 10 - Complex numbers
- HELM 12 - Applications of Differentiation

STACK questions were developed and used with 175 students taking MAA104, Mathematics for Aeronautical and Automotive Engineering. A more substantial longer term commitment will be needed to provide updated companion quizzes for the other HELM workbooks.

6 Case study: The Open University, UK

Contributed by Tim W. Lowe, *Department of Mathematics and Statistics, The Open University, Walton Hall, Milton Keynes, MK7 6AA, United Kingdom*

The Open University (OU) is the UK's leading distance learning university with students studying from home, mainly on a part-time basis.

STACK has been used in a number of taught mathematics modules that have been introduced by the OU since 2014. This report concentrates on its use in MST124 and MST125 (*Essential mathematics 1 and 2*), two first level modules (each being equivalent to one-quarter of a years full-time study) which introduce and teach calculus and other related topics. STACK is used in these modules to generate and mark randomised questions with both numerical and algebraic answers which test a range of different techniques. STACK is also being used in M303 (*Further pure mathematics*), a third level module (equivalent to one-half year of full-time study) which teaches topics including number theory; rings, fields and groups; and metric spaces. Here, STACK is used to test, for example, number theory and finding the minimal polynomial of an element in a field extension.

MST124 and MST125 have similar structures. Each is offered to students twice per year (beginning in either February or October). Approximately 4000 students study MST124 each year, and 1000 study MST125. Each module contains 12 study units, each of which (except for one MST125 unit) has an associated formative Practice Quiz which students can take as many times as they wish to practise their mathematics. All except for one of these quizzes are formed from randomised STACK questions. Each quiz contains approximately 11 questions on average, and the average number of student attempts at these quizzes for the module presentations beginning in October 2014 are approximately 1500 for MST124 and 430 for MST125. Students are given three attempts at answering a question, with small hints being given after an incorrect answer, and a final full worked solution to the particular variant of the question asked. In addition, each module contains four summative STACK assignments, where only one attempt at each question is permitted. The marks for these assignments contribute 25% of continuous assessment mark for the module, the remainder being for the written assignments.

The high numbers of students on these modules lead to the University recording one million student STACK question attempts between January 2014 and March 2015.

Student feedback on the role of the practice quizzes has been positive, as exemplified by the following quotation from a module online forum.

I do value the practice quizzes because you do get practice, the feedback explanations are very good and detailed.

[I] have done two runs of the practice quiz which has made a big difference for me and will become something I will be doing much more as helps so much.

I have been doing the practice quiz every morning now for 2 weeks ... I managed 100% on one of the quizzes and over 90% now on all of them so my confidence level is going up!

Students use the quizzes both to consolidate learning and as a revision tool. Many students attempt individual quizzes multiple times. Students seem to read the worked solutions in detail, with questions being raised in online forums if parts of a solution cannot be understood. There was concern in initial presentations of the modules about the use of STACK in summative assessments, mainly arising from students losing marks when a correct answer was input using the wrong syntax. This situation is improving, both as a result of improvements to questions and Potential Response Tree marking schemes, and due to improved syntax checking at a system level, thanks to the STACK developers.

7 Case study: Aalto, Finland

Contributed by Antti Rasila, *Aalto University, Helsinki, Finland.*

Aalto University, located at the Helsinki Metropolitan area, is the leading university in Finland specialising on science, arts, technology and business. Aalto University was established in 2010, by a merger of three universities: the Helsinki School of Economics, Helsinki University of Technology (TKK) and the University of Art and Design Helsinki.

The computer aided assessment system STACK was introduced at the former Helsinki University of Technology in 2006, and extensively modified for the local requirements (see [2, 9, 12]). For discussion and motivations on our work at Aalto University, see [9, 10, 11]. New features added to the Aalto version of STACK (branched from version 1.x of the main trunk) included user input and display modifications such as multiple input fields in one exercise, two dimensional (i.e., matrix-valued) input fields, and more browser independent rendering of mathematical formulas and answer dialog elements (see [2]).

Until recently, mainly this local experimental branch of STACK has been used at Aalto University, but our aim has always been to merge the useful features back to the trunk. Furthermore, we have continued to develop new features to the trunk also. Examples of more recent development are allowing output of variables in Maxima syntax, and the conditional and iterative generation of question text. The latter includes introducing `if` and `foreach` expressions to STACK syntax, the possibility of re-defining of variables inside the question text, and content generation with external tools using parameters from the question or student's answer to it (see [11]). The development branch of Aalto is expected to be fully merged into the main trunk of the STACK codebase by the next release, version 4.0. The plan is to phase out our version of STACK 1.x from the production use by the autumn semester, 2015. During the academic year 2014–15, the number of students who used versions 1.x and 3.x were 1629 and 1156, respectively. Note that here is some overlap, as the same student may have taken several courses and used the both versions.

During the academic year 2014–2015 exercise assignments using STACK exercises have been implemented in practically all BSc level courses of engineering mathematics at Aalto. The volumes of the major courses in engineering mathematics are shown in Table 3. These courses are compulsory mathematics for the first and second year students majoring in various fields of engineering and science. These courses could be classified as blended learning, and they consist of lectures, classroom exercises, as well as STACK exercise assignments. The courses last six weeks, and students typically are assigned up to ten STACK exercises each week, so that STACK replaces some of the traditional classroom exercises. STACK is also used as a form of continuous assessment since the final grades are a result of the mid-term exams and the points earned in the exercises [7, 11]. STACK assignments are also used in teaching more advanced topics such as complex analysis and discrete mathematics, and for teaching other mathematical subjects such as logistics and, in particular, physics, but topics other than mathematics are not included in the above statistics.

This report is based the paper [11], and on joint work with M. Harjula, L. Havola, H. Majander, J. Malinen, H. Tiitu, and many others.

Note also that Aalto is coordinating a Finnish national project that includes all seven Finnish universities with MSc program in engineering: Aalto University, Åbo Akademi, Lappeenranta University of Technology, Tampere University of Technology, University of Oulu, University of Turku and University of Vaasa. All these universities are either using STACK already, or starting to use it by autumn semester 2015.

8 Case study: Mathematical Assessment in ILIAS Using STACK

Contributed by Matthias Kunkel, *ILIAS open source e-Learning.*

ILIAS is an open source Learning Management System, originally developed at University of Cologne and running today at many German and Swiss universities. Since 2014, STACK has been used in ILIAS as

Topic	Version	Courses	Exercises	Students	Marks
Matrix algebra	1	7	122	796	16506
	3	1	21	40	679
Single variable calculus	1	4	51	721	11427
Multivariable calculus	1	4	48	705	15536
Vector analysis	3	3	42	226	2748
Probability and statistics	1	2	39	563	12691
	3	2	84	452	14294
Total		23	407	3503	73881

Table 3: The usage of STACK at the large BSc level engineering mathematics courses at Aalto University during the academic year 2014–15. STACK versions 1 and 3 have been used as shown. From the paper [11].

a question type for assessments in Mathematics in which the user’s input is evaluated using a Computer Algebra System (CAS).

It has been possible to implement E-assessments as well as e-exams with the ILIAS integrated testing tool for many years. But the previously available question types were not satisfactory for use in mathematical education. The main reason is that these question types were not able to verify the learner’s answer and to decide if the input is right or wrong from the mathematical point of view. In the past, students’ answers could only be compared to predefined solutions but not calculated.

To change this, several lecturers and tutors in Mathematics within the German ILIAS community formed a special interest group in 2012 and started to evaluate possibilities for connecting ILIAS to a CAS. They had a look at several existing tools, commercial ones as well as open source solutions. Finally, the group recommended the ILIAS development team should use STACK and Maxima (CAS) for mathematical assessments. The main reasons for this decision were the features of STACK and how it interacts with Maxima, as well as the fact that all components are open source and could be easily connected to ILIAS and used without additional costs.

On the conceptual level, the ILIAS team decided to make STACK a specific test question type within the existing ILIAS test tool. The look & feel of STACK questions is similar to other ILIAS question types and questions can easily be combined with other question types to create extensive self-assessments or e-exams. The STACK question type has been implemented as a plugin to reduce dependencies between the ILIAS and the STACK development cycle.

Due to restricted funding and development resources, a first version of the STACK question was coded for ILIAS in 2014 that only allowed for the import of existing STACK questions and to run them in an ILIAS test. The questions themselves still had to be created in Moodle, exported in Moodle XML format and finally imported in ILIAS. In January 2015, version 2 of the plugin was published. It offers an integrated authoring interface to create STACK questions with all its options within ILIAS itself. The next step on the roadmap is the implementation of a test question export from ILIAS into Moodle XML format. This will allow the re-use of ILIAS made questions in Moodle. The aim of this development is to establish a platform-independent exchange of mathematical test questions among German universities - no matter if they use Moodle or ILIAS as LMS.

One advantage of the current implementation of STACK in ILIAS as a question type is that it is easy to set up assessments consisting of different question types. Such tests in Mathematics combine questions for calculating (STACK) with multiple choice, gap questions or image map questions. Authors of test therefore have more didactic options for creating and designing tests compared to a system that only offers one question type. Additionally, STACK questions in ILIAS can re-use features that come with every test question type in ILIAS easily.

At German universities, STACK questions are often used in self-assessments for first-year students to evaluate their mathematical competency and to offer them training to improve their calculation skills. Due to the option of random selection of questions in tests, ILIAS can automatically build new combinations of test questions in each test run and always present new tests to the learners. While users quickly get bored when having to answer the same questions again and again, a randomly built test with changing questions keeps the learner's attention much longer — and therefore improves the training effect.

But today, STACK is also used at universities to offer online exams. Due to the Bologna process and its introduction of Bachelor and Master programmes in the German university system, the number of exams has increased in almost every faculty. While paper-based exams need a certain time to be evaluated and rated, online exams are evaluating given answers automatically. This is a significant reduction of workload for the lecturer. This fact made online exams very popular at German universities in recent years. Thanks to STACK, such tests can also now be offered in subjects related to Mathematics.

While a technical solution to run tests in Mathematics is now available for ILIAS, lecturers and teachers are more and more interested in sharing STACK questions among each other and to improve the quality of the questions. Currently, a simple sharing platform has been set up in the ILIAS community as a pilot service. It enables teachers to offer and describe STACK questions and to download them. In the future, STACK question sharing will be extended and offered for other users, too, including those from universities using Moodle. All questions will be published under a Creative Commons license and should be available for anyone. As most of the questions do not contain much text, it should be relatively easy to translate them to other languages and re-use them outside Germany, for example by the British STACK community.

9 Case study: non-users

The survey was open to people who had not used STACK, and publicity encouraged people who had not used STACK or who had decided not to continue to use STACK to participate. From a developer's perspective feedback from this group can be particularly difficult to collect. 15 respondents had not used STACK. By answering "no" to the second question, (Do you make use of the STACK CAA system?) the survey tool would jump to Q14, asking for experience with other CAA.

One responded said STACK was too advanced for their teaching needs. A number of people had did not have Moodle, e.g. *"School decided not to go forward with Moodle as a VLE"*, or are unable to install an optional plugin on a hosted service.

A number of people use something else, e.g. Dewis and Numbas, through Blackboard.

All the systems have their distinctive advantages are drawbacks: Dewis is complicated to code and a bit basic of interface, but server-side and very thorough on data retention; Numbas has a good platform and is simple to use for setting up assignments, but it locks away much of its power when it comes to writing new questions; and STACK's link with Maxima and mathematical rigour are attractive, but its narrow platform support is an obvious restriction for many. Numbas is great for colleagues who don't want to write their own stuff but just want a simple, off-the-shelf solution. Its interface is really friendly, and I can easily see how other CAA systems could learn from the way that it works.

One person wanted more materials.

We stopped using STACK because it doesn't have a problem library. We use WeBWorK for many, many years and it is impossible to import our huge problem library (written in perl) to STACK. So we expanded webwork but we "dumped" STACK

STACK is distributed with a comprehensive demonstration course of materials for calculus methods and algebra, but it lacks the large library of WeBWorK.

Only one person cited technical problems for not using STACK:

I used STACK for a semester or two. We had a lot of lag when we were running STACK on our server but we could not determine the root of the issue or the fix so the competition seemed like an obvious choice. We now use MapleTA which had very similar issues, the difference was we were paying a team of MapleTA people to fix the many issues that kept appearing and so didn't have to use our own time. I believed STACK was a better choice in product but I was outvoted by the rest of my discipline.

10 Future plans for STACK

Future plans are listed online

https://github.com/mathsmoodle-qtype_stack/blob/master/doc/en/Developer/Future_plans.md

Comments, discussion and contributions are very welcome. One feature which will be released in the next version, and which is a substantial change in behaviour, is described below. The code for this is currently being tested.

10.1 Question blocks

Question blocks add flexibility to STACK questions by adding functional structures to the text based fields of the question, known as “CAS_{text}”. In the past, CAS statements could be evaluated and the mathematics inserted into otherwise static text. This was used to insert random values, feedback and worked solutions.

There are two important changes

1. The value of an expression can be inserted into the text, in Maxima syntax, not just the \LaTeX displayed form.
2. Conditional “blocks” statements can be evaluated, and blocks of text included or excluded on the basis of these blocks.

For maximum flexibility, blocks can be nested and conditionally evaluated. A body of CAS_{text} is then repeatedly processed until all blocks have been interpreted. This is a core part of CAS_{text} and so applied to all appropriate parts of the question. Note that currently the parameters to blocks in the question body may not depend on the student's answers. This means that you cannot reveal an input block based on student input. We may expand the capabilities in the future to remove this restriction once we are satisfied these new features have proved their reliability in practice.

11 Conclusions

In the last two years there have been substantial increases in the number of STACK sites and the numbers of users, both staff and students. The current code based has proved reliable and robust for groups of up to 3000 students, e.g. at the United Kingdom Open University. Colleagues reported favorable reactions from students, although this report did not seek student reactions directly. STACK is being used internationally, and to underpin a number of substantial online learning projects. The features of STACK currently enable a wide range of mathematical assessments, and the current list of planned new features would substantially add to this range and depth. Further development work will be needed to implement these over the next few years.

Online assessment is now seen as a mainstream part of teaching mathematics at university, and there are a number of different online systems to support this. STACK's strength is the mathematical sophistication provided by a full-features CAS. The ILIAS community have demonstrated the ability to connect the STACK question type to learning systems beyond Moodle. Together with the growing list of Moodle users prospects for maintaining and improving STACK are healthy.

References

- [1] J. Appleby, P. C. Samuels, and T. T. Jones. DIAGNOSYS – a knowledge-based diagnostic test of basic mathematical skills. *Computers in Education*, 28:113–131, 1997.
- [2] M. Harjula. Mathematics exercise system with automatic assessment. Master's thesis, Helsinki University of Technology, 2008.
- [3] P. Hassmén and D. P. Hunt. Human self-assessment in multiple choice. *Journal of Educational Measurement*, 31(2):149–160, 1994.
- [4] B. Hoffmann. *The tyranny of testing*. Crowell-Collier, 1962.
- [5] A. N. Kluger and A. DeNisi. Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2):254–284, 1996.
- [6] G. C. Leder, G. Rowley, and C. Brew. *International comparisons in mathematics: The state of the art*, chapter Gender differences in mathematics achievement: Here today and gone tomorrow?, pages 213–224. Falmer Press, 1999.
- [7] H. Majander and A. Rasila. *Tutkimus suuntaamassa 2010-luvun matemaattisten aineiden opetusta*, chapter Experiences of continuous formative assessment in engineering mathematics, pages 197–214. Tampereen yliopistopaino Oy - Juvenes Print, 2011.
- [8] R. C. Paiva, M. S. Ferreira, A. G. Mendes, and A. M. J. Eusébio. Interactive and multimedia contents associated with a system for computer-aided assessment. *Journal of Educational Computing Research*, 52(2):224–256, 2015.
- [9] A. Rasila, M. Harjula, and K. Zenger. Automatic assessment of mathematics exercises: Experiences and future prospects. In *ReflekTori 2007: Symposium of Engineering Education*, pages 70–80. Helsinki University of Technology, Finland, Teaching and Learning Development Unit, <http://www.dipoli.tkk.fi/ok>, 2007.
- [10] A. Rasila, L. Havola, Majander H., and J. Malinen. Automatic assessment in engineering mathematics: evaluation of the impact. In *ReflekTori 2010: Symposium of Engineering Education*. Aalto University, Finland, Teaching and Learning Development Unit, <http://www.dipoli.tkk.fi/ok>, 2010.
- [11] A. Rasila, J. Malinen, and H. Tiitu. Automatic assement and conceptual understanding. *Teaching Mathematics and its Applications*, 2015.
- [12] C. J. Sangwin. *Computer Aided Assessment of Mathematics*. Oxford University Press, 2013.
- [13] C. J. Sangwin. An audited elementary alegbra. *The Mathematical Gazette*, July 2015.
- [14] C. J. Sangwin and P. Ramsden. Linear syntax for communicating elementary mathematics. *Journal of Symbolic Computation*, 42(9):902–934, 2007.
- [15] D. Sleeman and J. S. Brown, editors. *Intelligent Tutoring Systems*. Academic Press, 1982.

A Survey questions

For future reference, these are the exact survey questions.

1. You are being invited to take part in a research study. The purpose and details of this study have been explained to me in the information above. I understand that this study is designed to further scientific knowledge and that all procedures have been approved by the Loughborough University Ethical Advisory Committee. I have read and understood the information above and this consent form. I have had an opportunity to ask questions about my participation. I understand that I am under no obligation to take part in the study. I understand that I have the right to withdraw from this study at any stage for any reason, and that I will not be required to explain my reasons for withdrawing. I understand that unless I chose otherwise all the information I provide will be treated in strict confidence and will be kept anonymous and confidential to the researchers unless (under the statutory obligations of the agencies which the researchers are working with), it is judged that confidentiality will have to be breached for the safety of the participant or others. I agree to participate in this study.

2. Do you make use of the STACK CAA system?

3. Which language(s) do your students use (E.g. English, German)?

4. For how many academic cycles (years) have you used STACK?

5. Please select any of the following which apply to your use:

5.a. If you selected Other, please specify:

6. Please describe the extent of *your* use, e.g. approx number of students, frequency of quizzes.

7. Have you made use of the sample materials? If so please describe your extent of use.

8. What do you most like about STACK?

9. What do you like least? What difficulties have you or your students had?

10. What features would you most like us to add?

11. Have you modified STACK, or been involved in its development? If so how?

12. Have you published any papers reporting STACK use? If so, please provide references or links.

13. Do you have any example questions you would be willing to contribute to the demonstration site? (Under a Creative Commons Attribution-ShareAlike licence.) If so, please supply an email address on the next page.

14. If you don't use STACK, then why don't you use STACK? Do you use or have you used another computer aided assessment system for mathematics? If so what is it? Please compare your experiences.

15. If you are willing to be contacted by me, please supply an email address or other contact information.