Institution: Maxwell Institute of Mathematical Sciences		
Unit of Assessment: UoA 10 – Mathematical Sciences		
Title of case study: STACK- Online aided assessment of reasoning		
Period when the underpinning research was undertaken: July 2015-Dec 2020		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s): Prof. Christopher Sangwin	**Role(s) (e.g. job title):** Professor of Technology Enhanced Science Education	**Period(s) employed by submitting HEI:** July 2015-
Period when the claimed impact occurred: 2014-2020		
Is this case study continued from a case study submitted in 2014? No		

1. Summary of the impact

STACK is an online assessment software for mathematics and other STEM disciplines. At December 2020, STACK is used across UK HE within over 1300 registered learning management systems, including the Open University. Translated into 20 languages, STACK is used by over 30 universities in Germany, by every university in Finland, and in Japan, Israel, Kenya and the USA. STACK is used for international development and commercial school textbooks. Through all of these instances, it has improved the learning of hundreds of thousands students, it has supported institutions to pivot to online learning during the COVID-19 pandemic, and it saves tens of thousands of hours of human marking annually. The underpinning research has directly influenced the development of many similar/competitor products.

2. Underpinning research

Assessment and feedback are essential for effective learning. Bloom's work on "Mastery learning" in the 1980s found that the gap between tutors and traditional large-group teaching could be closed by regular testing with carefully designed mastery tests. Using online assessment provides an opportunity to implement Bloom's ideas, however, online assessment has predominantly relied on multiple choice tests. Multiple choice questions are particularly problematic in mathematics because, in addition to elimination and guessing, students can reverse-engineer many questions. Research with STACK found that when faced with 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, see [3.1].

Development of STACK is based on the observation that authentic and valid assessment requires the student to provide an answer which contains substantial content, rather than using a multiple choice (or similar) question. Students must input mathematical expressions, and complete mathematical arguments; the software establishes objective properties of those arguments and provides outcomes including feedback and statistics. This technology is essential in developing fully online courses, thereby changing teaching structures, which move away from large groups to enable students to progress at an individual pace.

Professor Sangwin initiated the STACK project and platform from scratch, and subsequently worked in partnership with other institutions such as the Open University and Loughborough University to develop different aspects of STACK. His leadership has enabled the research from other universities to have a global reach. Since 2015, The University of Edinburgh has

been the home of the STACK project. We hosted the second international STACK conference in April 2019.

STACK focuses on using computer algebra to provide teachers with very carefully designed tools to enable online assessment. Key research at The Maxwell Institute includes [3.2] which expanded the computer algebra features to include dimensional numerical quantities, based on SI and fine-tuned to the needs of assessments. Further work carried out at the Maxwell Institute [3.3] has expanded the functionality to include algebraic line-by-line reasoning. Reasoning by equivalence is where an equation is manipulated to generate a new and equivalent equation, typically ending when the equation is solved. Professor Sangwin's research [3.4] found that approximately a third of the method marks are awarded for reasoning by equivalence in final high school mathematics examination questions. There are many other forms of reasoning but reasoning by equivalence forms a stepping stone to advanced proof [3.5]. In STACK a student's whole argument then becomes a single mathematical object (just as an equation can be treated as a single object) which is to be subjected to formal verification. The research reported in [3.2, 3.3, 3.4] led directly to significantly extended functionality.

3. References to the research

[3.1] C. J. Sangwin and I. Jones. Asymmetry in student achievement on multiple choice and constructed response items in reversible mathematics processes. Educational Studies in Mathematics, 94:205–222, 2016. https://doi.org/10.1007/s10649-016-9725-4

[3.2] C. J. Sangwin and M. Harjula. Online assessment of dimensional numerical answers using STACK in science. European Journal of Physics, 38, 2017. https://doi.org/10.1088/1361-6404/aa5e9d

[3.3] C. J. Sangwin. Reasoning by Equivalence: The Potential Contribution of an Automatic Proof Checker. In: Hanna G., Reid D., de Villiers M. (eds) Proof Technology in Mathematics Research and Teaching. Mathematics Education in the Digital Era, vol 14. Springer, Cham. 2019. https://doi.org/10.1007/978-3-030-28483-1_15

[3.4] C. J. Sangwin and N. Kocher. Automation of mathematics examinations. Computers and Education, 94:215–227, 2015. https://doi.org/10.1016/j.compedu.2015.11.014

[3.5] C. J. Sangwin and Bickerton, R. Practical Online Assessment of Mathematical Proof. Accepted, International Journal of Mathematical Education in Science and Technology, (2020) http://arxiv.org/abs/2006.01581

4. Details of the impact

STACK is a contemporary assessment software for mathematics. It is installed on the learning management system of over 1300 registered websites with groups of up to 1500 students [5.1]. Extrapolating, we estimate 1.2 million students worldwide are benefitting from this methodology. Having been translated into 20 languages, STACK is used in over 30 universities in Germany, every university in Finland and many other countries including Japan, Israel, Kenya, and the USA [5.2, 5.3, 5.4]. The abacus consortium, established to share STACK and other mathematics assessments, currently has 38 partners [5.5], further extending the reach of STACK.

In comparison with other online assessment software, STACK offers unique additional functionality that changes the way that both learners and teachers engage with learning and assessments. Students enter mathematical expressions and STACK will assess the properties of these expressions; teachers author their own questions; STACK can generate questions with random variables, to reduce copying of answers; STACK allows teachers to give partial marks and tailored feedback depending on the different mathematical properties of the students' answers. STACK supports multipart questions, enabling teachers to write

step-by-step questions; STACK automates the "question testing" process enabling robust questions with long-term support. This means that there is almost double the assessment per student using half the staff time, increasing the efficiency of staff while improving student learning [5.3]. The research described in [3.4] has led directly to uptake for physics textbooks by the commercial textbook publisher, Physics Curriculum & Instruction, and more generally wider use in STEM subjects. The publishers write "*Instructors using our system have stated that students have greater success with solving physics problems, and gain deeper understanding into problem-solving strategies, when compared to traditional homework environments of working alone without computer aided assistance"* [5.6].

Combining online assessment with human marking saves institutions thousands of hours of work each year: the University of Edinburgh estimates the School of Mathematics saves over 8,300 hours of work annually, through 7,500 accounts, equating to over GBP200,000 of savings year on year. This is replicated for STACK users across the world, for instance *"STACK is used for both summative and formative assessment. Over 1.2 million STACK questions were answered by Open University students in the 2019/20 academic year. This saves a significant amount of staff time each year. Marking this volume of questions, even if only taking 5 seconds per question would take approximately 1 year of staff time each year"* [5.7].

As a result, STACK enables institutions to redeploy human resource and maintain quality despite growing student numbers. The University of Edinburgh consolidated online assessment in most year 1 and 2 mathematics and general science courses with the support from a dedicated learning technologist [5.8]. The University of Durham established a similar post in 2020 [5.9], and University College London appointed two people to 3-year posts to create STACK questions. This has changed the way courses in mathematics and science are delivered across these institutions. The Head of Department, Mathematics, UCL writes: *"The impact of this change of assessment on our year 1 students, most of whom will end up studying the whole year remotely, is immense. They get much more detailed feedback than they would have had in a normal year; and this is vital when we're exactly at the time when the support structures (such as informal contact with tutors and peers) that would have formed the academic safety net have been damaged by the pandemic"* [5.10].

The novel work described in [3.3] extends assessment capabilities to entire mathematical arguments in a well-defined class by assessment of students' working itself. Feedback is provided to increase students' understanding of mathematics, improve students' experience, and raise their competence. The Open University comments "*Whilst it is difficult to isolate the effect of STACK ….. modules within Mathematics and Statistics enjoy some of the highest student satisfaction ratings within the University. The use of STACK questions within modules is often praised by students in unsolicited comments within end-of-module feedback"* [5.7].

In 2020, to mitigate the effects of COVID-19 on teaching, the Maxwell Institute ran 12 practical workshops on question authoring, with over 350 academic staff attendees from across the world, and presented 15 online talks about STACK, the underlying research, and how to use this technology in teaching [5.11]. These "*seminars and conference talks have guided and challenged colleagues' thoughts on the nature of assessment and how to assess different mathematical topics, particularly for distance-learning students. The most recent example of this was Chris Sangwin's online workshop on the assessment of mathematical proof, given in September 2020"* [5.7]. As a further response to COVID-19, STACK was used in December 2020 to replace university examinations, in Edinburgh reducing the marking time in one module alone from 35 person days to 22 person days with a combination of automatically marked STACK assessments and human marking of more complex proofs.

In addition, the research underpinning STACK has influenced the design of other successful online assessment systems, particularly the open source NUMBAS system and "Learning Algebra" from Haese Mathematics, [5.12].

5. Sources to corroborate the impact

[5.1] Data on number of sites using STACK (increasing from 134 in August 2013 to 1352 in December 2020), number of languages into which STACK has been translated, and numbers of student users https://moodle.org/plugins/stats.php?plugin=qtype_stack.

[5.2] 11 published case-studies of worldwide STACK use – https://www.maths.ed.ac.uk/~csangwin/stack/2019-cate-case-studies.pdf or https://stack-assessment.org/CaseStudies/.

[5.3] An overview of the STACK assessment system can be found at https://stack-assessment.org/. The STACK demonstration site can be viewed at https://stack-demo.maths.ed.ac.uk/demo/ and the underlying source code at https://github.com/maths.

[5.4] Letter from Director, IDEMS International Community Interest Company.

[5.5] Collective of 35 international universities devoted to using STACK in teaching https://abacus.aalto.fi and letter of support from the Abacus coordinator.

[5.6] Commercial use in School Physics textbooks, e.g. https://stack-assessment.org/CaseStudies/2019/PhysicsCurriculum/ and letter of support from the publisher, Physics Curriculum and Instruction.

[5.7] Letter from Director of Teaching, at the School of Mathematics and Statistics and the Head of School of Physical Sciences at the Open University outlining the centrality of STACK in their mathematics provision.

[5.8] C. J. Sangwin and K. Zerva. Developing online learning materials to support undergraduate education at the University of Edinburgh. Mathematics Today, 212-215, 2020.

[5.9] Letter from Head of Department of Mathematical Sciences, Durham University detailing their use of STACK for e-assessment.

[5.10] Letter from Head of Department of Mathematics, UCL, indicating the impact of STACK in teaching, particularly during the pandemic.

[5.11] Agenda from meetings and conferences (e.g. international STACK users conference, Germany, Nov 2018).

[5.12] Letter from Haese Mathematics publishers describing the significant influence published research about STACK has had on the design of their proprietary “Learning Algebra” software.