Rejuvenating the HELM Workbooks as Online STACK Quizzes in 2020

https://doi.org/10.3991/ijet.v17i23.35923

Konstantina Zerva[1(✉)], Christopher Sangwin[1], Ian Jones[2], David Quinn[1]
[1]School of Mathematics, University of Edinburgh, Edinburgh, UK
[2]Mathematics Education Centre, Loughborough University, Loughborough, UK
k.zerva@ed.ac.uk

Abstract—HELM (Helping Engineers Learn Mathematics) was a three-year project undertaken by a consortium of five English universities between 2002 and 2005. HELM released a range of workbooks and online resources, designed to enhance the mathematical education of engineering undergraduates, which were well received, and widely used. All that remained in 2020 from HELM were PDFs of "student workbooks" which could not be edited. Since 2018, the University of Edinburgh has developed and taught fully online courses for university mathematics in which all materials form part of coherently organised digital exercises and expositions. The work reported in this paper combines the two approaches: we take a similar approach to the existing successful Edinburgh course and rejuvenate the static HELM materials by creating online resources designed to support the mathematical education of engineering undergraduates. The paper describes the rationale, educational design, practical execution and preliminary analysis of the workbooks' usage at the University of Edinburgh.

Keywords—mathematical education of engineering undergraduates, online assessment, curriculum development

1 Introduction

As a response to COVID-19 during the summer of 2020, the University of Edinburgh and Loughborough University collaborated to develop online workbooks for engineering and STEM mathematics based on the previous HELM (Helping Engineers Learn Mathematics) project workbooks. HELM materials were the outcome of a three-year curriculum development project undertaken by a consortium of five English universities led by Loughborough University between 2002 and 2005. As of April 2020 only the PDF files remained publicly available from the original project. We sought out, and obtained permission to use, the original source files for HELM.

In the summer of 2018 we developed a fully online course "*Fundamentals of Algebra and Calculus*" (FAC) through the STACK (System for Teaching and Assessment using a Computer algebra Kernel) online assessment system at the University of Edinburgh [1], [2]. This course has been tried and tested, and proved highly successful. In the summer of 2020 we developed another fully online course "*Introduction to*

Mathematics and its Applications" (IMA) and used many of the techniques developed for FAC in a wide variety of other courses as a response to COVID-19. Based on our positive experience of running and evaluating FAC, we decided to rejuvenate the HELM materials in the style of FAC to better support engineering students.

In Sections 2 and 3 of this paper we describe FAC and the HELM project respectively. Then Section 3 provides some basic information about the STACK assessment system, and the Moodle learning environment. Section 4 describes our approach to the material development during the summer of 2020, Section 5 details how we have made use of these materials during 2020–21 and provides some comment on preliminary evaluation. Section 6 describes our intended next steps.

2 Background: FAC online course

The University of Edinburgh course "*Fundamentals of Algebra and Calculus*" (FAC) is designed to prepare students for study of mathematics at the University of Edinburgh, where incoming students have a wide range of mathematical backgrounds [1], [2]. Fundamentals of Algebra and Calculus covers approximately the material taught as part of the Scottish Qualifications Authority (SQA) *Advanced Higher Mathematics* and English Advanced GCE ("A-level") in *Further Mathematics*. The course is optional for incoming students, has run since 2018, and is taught completely online. The central idea behind the implementation of FAC is to take the book and put the academic content inside a sequence of interactive online quizzes, an idea we discuss in detail in [3]. That is, there are short blocks of text, worked examples, interactive diagrams, and short focused video clips embedded within the sequence of exercises and questions. Students work through the quizzes, and success on the course is judged by their success on the quizzes. To be clear: there are no textbooks, no printed notes or PDF documents; no regular live lecture events; no problem/exercise sheets to hand in (i.e. homework); and no separate final written examination. The metaphor we followed is to *put the book inside the quiz*.

The focus is firmly on exercises and questions which direct students' attention and require their activity. Most questions in the course (apart of the tests at the end of each week) give immediate feedback on students' attempts. In that way students become aware of what they had learned and they can immediately act on the feedback by retrying the question [4], [5], [6]. The development of FAC was informed by previous educational research on mastery learning [7], interleaving of topics [8], [9] and by using specific techniques such as sequences of faded worked examples [10]. The topics are arranged into 10 weekly units, each of which has a common structure:

- a "Getting Started" section, which motivates the week's topic and reviews pre-requisite content (e.g. differentiation facts when starting integration),
- four sections of content, each of which is designed to take around 2 hours to complete (roughly equivalent to one traditional lecture plus associated practice),
- a 90-minute "Practice Quiz" with a mix of questions on the work of the week; this can be taken an unlimited number of times, with full feedback provided on each attempt, and
- a 90-minute "Final Test" which is similar in style to the Practice Quiz, but only allows a single attempt.

All of these are implemented as quizzes in the Moodle virtual learning environment, containing STACK (and other) questions specifically designed for assessment of mathematics. The Moodle quiz environment and STACK questions are described in detail in Section 3.

A student's grade on the course is determined by combining the results of the 10 weekly Final Tests (together worth 80% of the grade) with a final 2–hour synoptic test covering topics from the whole course (worth the remaining 20%). The course was well–received by students in its first delivery in the 2018/19 academic year, and their overall performance on the course was strong: the mean course result was 66%, while 47% of students achieved an A grade and only 5/109 students failed. See [11] for more details.

Following the success of FAC, we created another course "*Introductory Mathematics with Applications*" (IMA) during the summer of 2020. IMA focuses on non-specialist students willing to consolidate and build their previous mathematics education to better prepare for a wide variety of degrees. Put more clearly, IMA was designed for students *not* taking a mathematics (or other STEM) degree, but still interested in taking an optional mathematics course in year 1 at university. This course was first introduced in the academic year 2020–21.

3 The HELM project

The mathematical education of engineers and scientists has been seriously considered as a distinct area of mathematics education for some time, e.g. see [12]. The engineering community have developed curricular guidance, so that different degree providers can prepare engineering graduates with a comparable mathematics education. The work of the European SEFI group has been particularly influential here, e.g. [13], [14] and [15].

HELM (Helping Engineers Learn Mathematics) materials were the outcome of a three–year curriculum development project undertaken by a consortium of English universities led by Loughborough University. It was funded by a £250,000 grant from the Higher Education Funding Council for England under the Fund for the Development of Teaching and Learning during the period from October 2002 to September 2005. The HELM learning resources were produced primarily by teams of writers and developers at five universities: Hull, Loughborough, Manchester, Reading and Sunderland. The outcomes of this project included workbooks, interactive learning segments, a computer–aided assessment regime which was originally intended to 'drive the student learning' and a report on possible modes of usage of the materials: see [16].

A further grant of £25,000 was awarded to disseminate the HELM learning and assessment resources across higher education institutions and the HELMet ("HELM educational transfer") project, ran from October 2005 to July 2006 and involved the Loughborough HELM team working with six universities who were using (and adapting) HELM materials: Leicester, Newcastle–upon–Tyne, Nottingham, Oxford Brookes, Portsmouth, and Salford.

3.1 HELM workbooks

Originally HELM had online support, but all that remained in 2020 were the PDF files of 50 Workbooks which are available publicly at https://www.lboro.ac.uk/departments/mlsc/student-resources/helm-workbooks/. These files comprise:

- 46 Student Workbooks, written specifically with the typical engineering student in mind. These are the core of the academic content of HELM.
- A Workbook containing an introduction to dimensional analysis, supplementary mathematical topics and physics case studies.
- A Workbook containing Engineering Case Studies ranging over many engineering disciplines.
- A Students' Guide.
- A Tutor's Guide.

Anyone wanting to know more details about the original HELM project, and who contributed to HELM, should consult the HELM Tutor's Guide.

The Student Workbooks, the academic core of the HELM project, include (a) worked examples, (b) tasks for students to undertake with space for students to attempt the questions, and, often, intermediate results provided to guide them through problems in stages, and (c) exercises where normally only the answer is given.

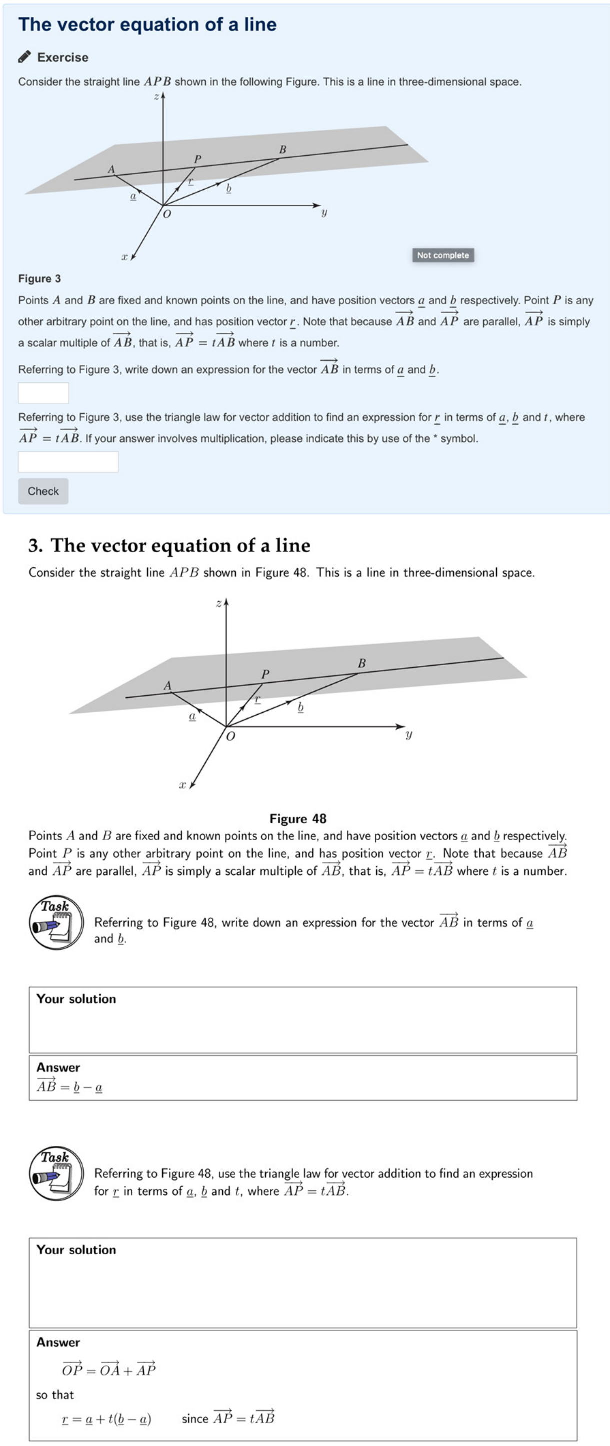
These Workbooks are subdivided into smaller manageable Sections, with each workbook containing 2–7 sections (190 sections in total). Each Section is available as a single PDF and was designed to be a self–contained topic. The length of the Sections ranges from 4 to 34 pages, representing a few hours' work for students. For the current project, each Section or PDF file, was converted into a single online 'quiz' (in Moodle terminology) containing the written content, including worked examples, and appropriate interactive content built in STACK and other question formats. An example of a subsection in PDF and interactive formats is shown in Figure 1.

4 The STACK online assessment system

The STACK online assessment system (System for Teaching and Assessment using a Computer algebra Kernel) is a highly sophisticated assessment software for mathematics, science and related disciplines. STACK uses the Computer Algebra System (CAS) Maxima to support the assessment process, a topic discussed in detail in [17]. STACK is open source software which forms a question type in the Moodle and ILIAS (Integriertes Lern-, Informations – und Arbeitskooperations-System [German for "Integrated Learning, Information and Work Cooperation System"]) learning management systems, which is very widely used internationally and which has been in continuous development and use since 2005.

STACK has the following key features:

- STACK uses the CAS to generate questions with random variables, ensuring different students see different variants of the question.



The vector equation of a line

Exercise

Consider the straight line APB shown in the following Figure. This is a line in three-dimensional space.

Figure 3

Points A and B are fixed and known points on the line, and have position vectors $\underline{a}$ and $\underline{b}$ respectively. Point P is any other arbitrary point on the line, and has position vector $\underline{r}$. Note that because $\overrightarrow{AB}$ and $\overrightarrow{AP}$ are parallel, $\overrightarrow{AP}$ is simply a scalar multiple of $\overrightarrow{AB}$, that is, $\overrightarrow{AP} = t\overrightarrow{AB}$ where t is a number.

Referring to Figure 3, write down an expression for the vector $\overrightarrow{AB}$ in terms of $\underline{a}$ and $\underline{b}$.

Referring to Figure 3, use the triangle law for vector addition to find an expression for $\underline{r}$ in terms of $\underline{a}$, $\underline{b}$ and t, where $\overrightarrow{AP} = t\overrightarrow{AB}$. If your answer involves multiplication, please indicate this by use of the * symbol.

Check

3. The vector equation of a line

Consider the straight line APB shown in Figure 48. This is a line in three-dimensional space.

Figure 48

Points A and B are fixed and known points on the line, and have position vectors $\underline{a}$ and $\underline{b}$ respectively. Point P is any other arbitrary point on the line, and has position vector $\underline{r}$. Note that because $\overrightarrow{AB}$ and $\overrightarrow{AP}$ are parallel, $\overrightarrow{AP}$ is simply a scalar multiple of $\overrightarrow{AB}$, that is, $\overrightarrow{AP} = t\overrightarrow{AB}$ where t is a number.

Task

Referring to Figure 48, write down an expression for the vector $\overrightarrow{AB}$ in terms of $\underline{a}$ and $\underline{b}$.

Your solution

Answer

$\overrightarrow{AB} = \underline{b} - \underline{a}$

Task

Referring to Figure 48, use the triangle law for vector addition to find an expression for $\underline{r}$ in terms of $\underline{a}$, $\underline{b}$ and t, where $\overrightarrow{AP} = t\overrightarrow{AB}$.

Your solution

Answer

$$\overrightarrow{OP} = \overrightarrow{OA} + \overrightarrow{AP}$$

so that

$$\underline{r} = \underline{a} + t(\underline{b} - \underline{a}) \qquad \text{since } \overrightarrow{AP} = t\overrightarrow{AB}$$

Fig. 1. HELM subsection example: interactive (top) and PDF (bottom) versions

- Students provide the final answer in the form of a mathematical expression, e.g. an equation, rather than responding to multiple choice questions. See [18] for a discussion of problems with multiple choice in mathematics.
- STACK establishes the mathematical properties of students' answers and generates outcomes including partial credit when an expression only satisfies some of the required properties.
- STACK can generate plots dynamically within any part of the question, including feedback in the form of a plot of the student's expression.
- STACK supports working with significant figures and scientific units, see [19].
- Students can work line by line reasoning by equivalence until they have a final answer in the correct form, see [20].

More information about STACK can be found on the documentation site: https://stack-assessment.org/ (Jan 2022).

STACK gives teachers the freedom to write their own questions while keeping questions authoring as simple as possible. That said, authoring STACK questions is a complex task. STACK grades answers through algorithms knows as "potential response trees" (PRTs), where teachers can give partial marks and tailor feedback depending on the different mathematical properties of the student's answer. For example, the teacher may give partial marks for an answer that is equivalent to the correct one but not written in the conventional algebraic form, or to the correct level of numerical accuracy.

A unique characteristic of STACK is that separates "*validity*" from "*correctness*". When a student types an answer, it is interpreted by the CAS and a "validation box" is shown displaying how the student's answer is interpreted. This gives the student a chance to fix any syntax errors before their answer is marked (e.g. typing sin$x$ instead of sin(x)). The validity/correctness separation helps students understand what type of answer is expected and this significantly reduces the extent to which students are penalised on a technicality.

In many situations we want the correct answers to have specific properties or follow a specific pattern. The teacher should be able to encapsulate the algebraic and pedagogic content of all questions of a certain class. Generally, this is achieved by randomising the answer, and reverse-engineering using this answer back to build the question seen by a student. For example, if we want to ask a student to solve a quadratic equation it is sensible to randomise the roots and use the roots to construct the quadratic, rather than randomise the coefficient a, b, c in $ax^2 + bx + c = 0$. In this way we make sure that all random variants of a question have the same proprieties, e.g. they all have integer roots, or perhaps all have a double root, etc. Care needs to be taken to ensure that each variant of the question has the same difficulty and that we do not set a mathematically impossible problems. Academic judgement, and constrained random generation, are both essential to create high-quality learning materials.

It is important to test questions to ensure they work correctly in advance of releasing them to students. Testing questions is difficult and time consuming, but important to ensure quality control. STACK automates testing by allowing authors to write "Question Tests" in a similar way as unit testing is handled in software development. The question author may establish with confidence that the "Potential Response Trees" are processing the student's answer as expected and that the marks are allocated correctly.

Indeed, STACK is distinctive in enabling authors to unit-test their materials in advance, and this has proved highly valuable in practice, both before the first use and for longer-term maintenance year to year. As a minimum, the teacher can automatically check that the pre-specified "teacher's answer" is actually awarded full marks by the potential response trees.

To help ensure questions work correctly teachers are strongly advised to pre-generate and "deploy" variants of a question. When a student attempts the question they will be given a random selection from the deployed variants, rather than a version randomly generated on the fly. Deploying variants aids quality control in two ways. Firstly, the teacher can run all test cases on all the deployed variants in advance. Automatic testing of variants significantly reduces the occurrence of impossible random questions. Secondly, the teacher can check each deployed variant is of comparable difficulty by reviewing the variants and exercising their academic judgement. Using the combination of test cases and deployed variants it is unlikely a student will be given a random variant which does not work correctly.

STACK gives the teacher the ability to write a full worked solution for each question, which follows the randomisation. The question author may need to define more question variables than are needed to specify the question in order to give a sufficiently detailed worked solution. This may require to spend more time on creating the question but it will normally be worth the effort as, in our experience, students find that a step-by-step solution is a particularly useful recourse for their learning. Access to the worked solution is normally given to the student once they have used all their permitted attempts at the question, once they have given a correct answer, or once a "due date" has passed. The quiz environment is flexible providing a teacher with a wide range of options suitable for purely formative teaching through to formal timed examinations.

4.1 Moodle and the Moodle quiz

Moodle is an online learning management system (LMS), widely used in a variety of learning situations including universities. The first version of Moodle was released in August 2002, and it is currently distributed under the GNU General Public License. The Moodle Project is led and coordinated an Australian company with approximately 50 developers which is financially supported by a network of partner companies worldwide, and a larger group of open-source developers.

The Moodle quiz is a core module which enables teachers to select a list of questions for students, and manage policies such as dates and the availability of worked-solutions. The quiz has a wide range of features making is suitable for a very wide range of use, including formative assessment or formal online examinations.

Moodle, an acronym in which the 'M' stands for Modular, has been successful in-part because it is possible to adapt and customise Moodle in many ways. In particular, the plugin mechanism allows for third-party features to be included and Moodle currently has hundreds of plugins which extend all aspects of the core functionality. The STACK question type, qtype_stack, is one such plugin which extends the functionality of the quiz sub-system for mathematics.

STACK is also available for the ILIAS learning system which is widely used in Germany.

5 Translating HELM workbooks into interactive workbooks

As a response to COVID-19 during the summer of 2020, the University of Edinburgh and Loughborough University collaborated to provide better materials to support the teaching of engineering students during a rapid emergency response for the forthcoming 2020–21 academic year. We needed reliable and comprehensive materials, designed with engineering students in mind, over which we had intellectual property rights permitting us to put the content online. In this context HELM, at this point only available as static PDF, was a natural choice as a starting point. For example, many design decisions (notation, terminology, content order) had been made during the original project. The materials were collaboratively developed to a high standard, were revised and mistakes removed regularly. The HELM materials remain in use some 15 years later.

To translate the HELM workbooks into interactive online workbooks we recruited 18 PhD students inters (Digital Creation Assistants) to help with the digitisation of the content. At the University of Edinburgh the digitisation of HELM workbooks was part of ASID project (Assist, Support, Implement, Deliver) [21], which had as main goal to facilitate the transition to hybrid teaching in the School of Mathematics. At Loughborough University the digitisation was part of the Centre for Mathematical Cognition's impact-generating activities.

Initially, each intern was assigned 2 HELM sections to ensure they started their work with a common task and to help them understand how to create online materials. They were asked to create one quiz per HELM section, containing all the worked examples, and with appropriate STACK and other Moodle questions (Multiple Choice, True/False, Drag and Drop, Matching). The process of converting each HELM section (PDF file) into an online quiz consisted of two main parts. First, the content (text) was converted to HTML. Second, the STACK and other format questions were created.

Most of the HELM PDF files were available in HTML. To convert the remainder we used pandoc (https://pandoc.org/) which creates clean HTML and retains the LaTeX maths environments. The .tex files included some antiquated code, and one student developed a script to automate the conversion to contemporary .tex code. The figures that were included on the .tex files were converted to Scalable Vector Graphics (SVG) files. SVG format was preferred amongst other images format mainly because it is a scalable vector format that can be resized to any dimension without losing any quality, and is likely to be supported for the longer term. In a learning context, an SVG file is an excellent option because it creates really sharp images which are good for people with vision impairments.

As was mentioned earlier, each HELM section includes simple tasks for students to undertake and exercises (normally at the end of each section) for students to work through. These were implemented using STACK and other Moodle question types, where appropriate. We decided that, as part of this project, all STACK questions needed to have:

- basic randomisation (unless impossible),
- question tests and a set of pre-deployed variants (for quality control),
- and full worked solutions, reflecting the particular random variant.

Of course, it is inappropriate to direct translate all paper-based questions into online questions without taking into consideration the educational construct they are testing. In some cases we needed to modify the existing questions in an appropriate way.

Quizzes may contain interactive components, such as graphs and diagrams in which the user can "drag" components to better understand the underlying relationship. However, we only included other software in our quizzes where the licence is compatible with "CC BY". ("Non-commercial licences" such as GeoGebra are incompatible and so cannot be included.). We mostly decided to use JSXGraph for interactive graphics, for consistency and maintenance.

All interns received training on STACK. We organised a practical workshop to introduce them to STACK, author their first STACK question and get familiar with the environment. STACK is, by design, a sophisticated tool and therefore unfortunately has a steep learning curve: most users need around a month to become proficient and independent question authors. To make the most of the limited time available, we identified similar STACK questions from existing courses and shared these examples with the interns. In most instances the interns simply needed to edit/modify these questions; only rarely were they required to write a question from scratch. We had frequent meetings where we discussed issues and resolved any enquires they had. A local STACK expert worked closely with the interns and had one-to-one sessions with them when necessary.

To coordinate the progress of each HELM section we used a ticketing system on TEAMS, a board where we could monitor the status of each quiz (in progress, ready for checking, completed). When the interns had finished a whole section (content and STACK questions), the section was reviewed by the local STACK expert, for STACK related issues, and by a more experienced intern for style consistency. The interns received feedback on their conversion and then did the relevant changes before moving the task onto the next board. When both style and STACK were accepted, the HELM section was checked by 2 undergraduate students for typographical errors and any smaller issues. At the end each HELM section, which was now fully translated into a Moodle quiz, was reviewed by the course organiser to verify that the educational content is maintained.

5.1 Style guide

To ensure the consistency of the layout between all authors we created a style guide, which largely follows the style of our previous FAC course. The guide outlines the agreed style these materials will adopt and contains information about the way we are naming the quizzes and the questions in them, the quiz settings and the content appearance (key points, examples, matrix inputs). See Figures 2 and 3 for some examples of style we used in the workbooks. Many of the mathematical design decisions, e.g. a consistent notation of derivatives, integrals and scientific units, had been made by the original HELM project. These decisions are a huge help when coordinating a large team of authors, and made it easier for us, but we had to recognise and respect this previous work.

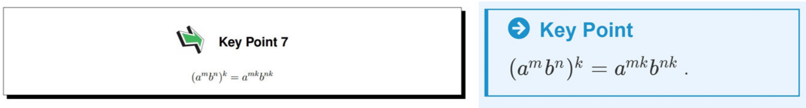


Fig. 2. Style of a "Key point": PDF (left) and interactive (right) versions

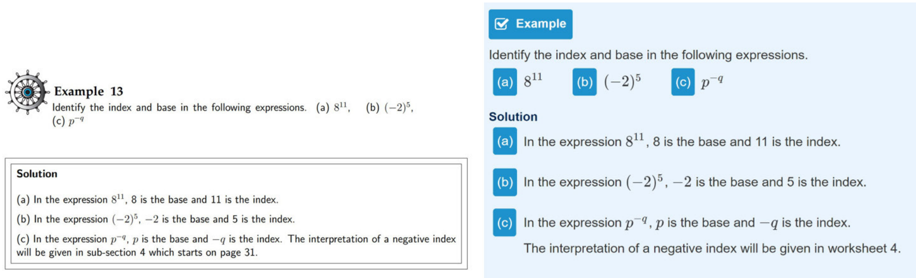


Fig. 3. Style of an "Example"—multipart: PDF (left) and interactive (right) versions

5.2 Workhours

Creating quality online learning materials requires a team effort and many work hours. We estimate that each intern could complete 2 HELM sections (content and STACK questions) during a working week. For our courses at the University of Edinburgh we required 48 HELM sections for semester 1 and 32 HELM sections for semester 2. Lecturers at Loughborough required a further 11 sections that were not already needed for the Edinburgh courses. Together this required around 1600 hours only for the content creation.

After being created, each Section was reviewed to verify that it was consisted with the style guide. This process took around 50 hours and it was necessary for the first few Sections (the interns became familiar with style guide after creating 2-3 sections each). Review of STACK question required more time and we estimate that it took on average 2 hours/section, so around 180 hours in total.

In addition 16 of the undergraduate students helped with testing and small edits in the workbooks. Each section needed a full day for testing, review and updates, but we manage to test only the sections that were created during summertime and needed for semester 1. Roughly speaking the process took 500 hours. Also the course organiser was doing the final review, and this required around 100 hours.

Summing all together, transforming almost half of the HELM workbooks to interactive quizzes required more than 2400 hours which is equivalent to 300 person-days of effort. This is a substantial investment, and we hope the results rejuvenate the materials and provide a starting point ongoing projects.

5.3 Challenges

Translating HELM resources into interactive online quizzes was a challenging process that required a lot of effort and coordination.

Regarding the general creation of the workbooks, we had many hands working in parallel (16 interns in Edinburgh all new to STACK) but only a few heads reviewing the work (3 STACK experts and 1 course organiser). The speed of work varied significantly, some of the interns were working part-time and most were later involved in other ASID tasks apart from HELM. As a result, the workbooks were completed out of sequence and we were inevitably giving repetitive feedback to different authors. It would be ideal to receive the HELM sections (now transformed into Moodle quizzes) in the order they appear on HELM because during the reviewing process we would be able to understand the course materials and how students will engage with them and to ensure at least local consistency in how the different authors presented the STACK questions. The workbooks could only be reviewed in sequence towards the end of the summer.

The main inconsistency was the technical quality of STACK questions. As we mentioned before, STACK has a steep learning curve, and STACK users need to pay attention to various aspects while creating a question, e.g. have proper randomisation, write a fully worked solution, add test cases, and deploy variants.

At the coding level, improper randomisation was a significant issue during the first review of the workbooks and during the first year of using them. Some of the issues (impossible cases) were easy to spot by running the background testing on questions as this would flag such errors. Consistency of random variables and not using reverse engineering in specific questions, were more difficult to spot and the only way to notice it was by reviewing the deployed variants. It is difficult to record and communicate why the initial author decided to delete a few deployed variants from a question and if in the future more variants are generated then it is likely some "bad" variants would reappear. Also, if you are sharing questions with colleagues, they won't know if the questions have bad randomisation.

The main issue we encountered in STACK questions was a lack of a fully worked solution. The original HELM source provides answers to all questions but typically this is the final answer, not a step-by-step solution. We know that fully worked solutions are a particularly valuable resource for the students. Inevitably, not all interns wrote full solutions and so these questions did not have the detailed model solutions we wanted to provide. Writing high quality worked solutions takes time and requires the author to have a clear understanding of the (i) the mathematics, (ii) what students might need highlighting. It is not enough to only rely on the answer the software calculates. Unfortunately, the only way to check the quality of worked solutions is by reviewing each of them: a task that requires a lot of time.

Most interns did not add any test cases on the first workbooks they created or some of the test cases they created were failing. Most of the interns did not deploy any variants and the majority of the deployed ones didn't give any useful information about the question because the question notes were inadequate (just a list of numbers, not a meaningful note about the question and the answer). These elements are important for managing the materials and are minor issues because they can be detected automatically when running the bulk test of all questions.

Our main suggestion for people who want to undertake a similar project is to hire fewer people for a longer period of time. Success depends on the number of expert staff, who will act as managers/coaches on the project. We believe that a manager can efficiently manage 3 interns who are new to the technology to be used.

6 Use of HELM online workbooks at the University of Edinburgh

This section contains a preliminary evaluation of the use of the HELM online workbooks at the University of Edinburgh during 2020–21.

6.1 Use of workbooks

The School of Mathematics at the University of Edinburgh has four courses specifically designed for engineers and chemists: Engineering Mathematics 1a (EM1a), Engineering Mathematics 1b (EM1b), Mathematics for the Natural Sciences 1a (MNS1a) and Mathematics for the Natural Sciences 1b (MNS1b), which run over 2 semesters. These courses cover the standard first year Engineering Mathematics (functions, linear algebra, calculus, 1st order ODEs).

Before 2020, the weekly structure of the courses was the following:

- Three 1-hour lectures, partially flipped.
- One 1-hour workshop/tutorial.
- 3 formative online (STACK) quizzes, with unlimited attempts, one linked to each lecture.
- 1 summative online (STACK) quiz covering the week's topic.
- A written assessment during even weeks.

During academic the year 2020–21 all our teaching moved online. We had around 500 students in year 1 Engineering and Chemistry which was a greater number than previous years. The weekly structure of the courses for 2020–21 was:

- 1 HELM workbook/week (3–4 sections). They served as the first introduction to each topic.
- Short video lectures to expand on theory and go through more complicated examples.
- One 1-hour online workshop/tutorial.
- 1 summative online (STACK) quiz.
- A written assessment on even weeks.

The courses continued to use the same textbook in 2020–21 as they had in previous years.

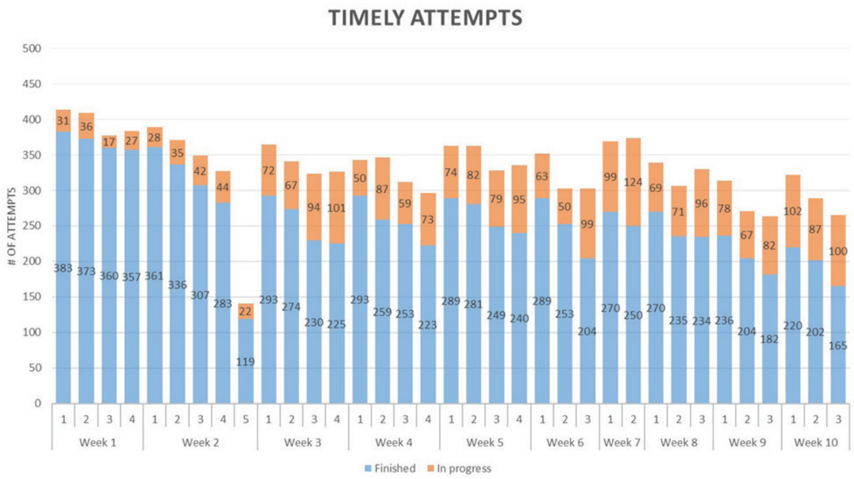
The HELM online workbooks were used as learning resources, and although STACK was set to award a score this did not contribute to students' final mark on the course. We hoped that the score awarded in the workbooks offered an additional incentive for students to engage and do well in the activity while also allowing students to track their progress. The students had unlimited attempts on the quizzes and they could come back on them for further practice throughout the semester and the revision period. The quizzes were set on "adaptive mode", meaning the students could immediately check their answer and retry with a small 10% penalty. A "Check" button appears in each question, and when clicked the students received immediate feedback on their response. Unsuccessful students who read and act on the feedback can review preceding content again if needed before continuing.

The students' evaluation of our courses indicated the HELM workbooks were considered to be the most valuable resource. They enjoyed the challenge that was presented at the workbooks, citing easy tasks at the beginning of each section and more challenging questions at the end. The students appreciated the existence of the "Check" button (the adaptive mode), that they were able to get the live feedback on the attempt. The element of enjoyment has been reported by [22] and shows the need to combine utility with enjoyment in a successful formative online assessment. Enjoyment is also related with the gamification in learning and as it has been reported by [23] students who use game–based pedagogical strategies have high levels of motivation.

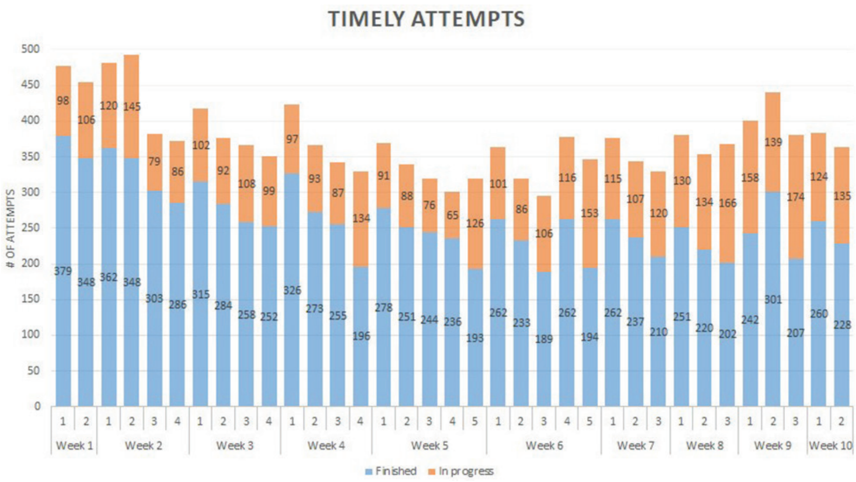
6.2 Preliminary evaluation

Students behaviour regarding the timing of interacting with HELM resources. Each week's materials were released the Friday before the beginning of that week, and the deadline for the summative assessment was the following Monday. Most weeks have either 3 or 4 quizzes. This gave the students 10 days (including both weekends) to interact with each week's content. We intended the students to have their first interaction with the workbooks during this period before accessing the other elements of the course and specifically before attempting the summative assessment.

We did some analysis of the overall attempts on each quiz to find how many of them happened in this 10–days period. We call these attempts *timely attempts* and they are depicted in Figures 4a and b for semester 1 and semester 2 respectively. Note that for semester 1 the EM1a and MNS1a courses had slightly different content, and they were 2 separate courses on the Moodle server. The analysis was made on the EM1a course because it had the most students, 362 fully matriculated students for EM1a in comparison to 130 fully matriculated students for MNS1a. In semester 2 both courses have the same content and had one course on Moodle, so the analysis happened for all the matriculated students on both courses (364 for EM1b and 124 for MNS1b).



(a) Number of students attempting the quizzes timely in semester 1 (EM1a only)



(b) Number of students attempting the quizzes timely in semester 2 (both EM1b andMNS1b)

Fig. 4. Timely attempts

Almost all students engaged with the materials within the first 10 days of their release (timely attempts). Since completion of the quizzes didn't offer any marks, and wasn't necessary for accessing each weeks assessment, a number of students didn't finish/submit these quizzes. We can see that the number of attempts decreases as the semester moves on, as we would expect. Also, we notice that there is a small drop in the timely attempts as the semester progresses. The first and second quiz tend to have more attempts than the third and fourth and we suggest this is because these topics are covered at the beginning of each week and students have more time to engage with them before taking the summative assessment. The fifth workbook of Week 2 on Semester 1 appears as an outlier here as this was brought forward from week 3 at the last minute and many students will have worked on this in line with the initial schedule.

Students behaviour regarding the timing of HELM quizzes vs assessed quiz. We expected students to engage with the formative HELM quizzes first, before trying the summative assessment. We know that some students go straight to the summative assessments, and only then look at the formative materials. Evidence for this behaviour comes from other courses in the School of Mathematics which have "reading quizzes" to replace one of the traditional lecture. In the past, we pre-assigned reading from the textbook and so could not see this kind of strategic behaviour.

For this analysis, we only looked at 3 weeks within semester 1 and 2. Each semester has 11 teaching weeks and any week between weeks 3–9 was an ideal candidate. We avoided weeks 1, 2 because some students may change courses and also avoided weeks 10, 11 because there are various deadlines towards the end of the semester and students tend to be less engaged. For each student's attempt on a certain week we looked on the "Started time" of the formative quizzes and compared them with the "Started time" of the summative quiz, and checked if a student opened the summative assessment before the formative quizzes. Our findings are shown on Table 1.

Table 1. The number of students who opened the summative assessment before the formative assessment.

Weeks	Assessment Only	Opened Assessment before WQs	Assessment and WQ Much Later
S1–W3	15	29	9
S1–W6	27	42	6
S1–W8	20	48	1
S2–W3	48	39	11
S2–W6	46	86	19
S2–W8	47	99	23

We found that every week there was a small number of students who only engage with the summative assessment and do not attempt the (formative) HELM quizzes. For semester 1 this was between 4 % to 7.5 %. For semester 2 this number was higher and around 10%. This is not particularly surprising. We found that during semester 1 between 8% to 13% of the students looked at the summative assessment and then looked at the assessed quizzes. During semester 2 this number was higher, between 8% to 20%. It is worth mentioning that during semester 2 there is a small, but not negligible, number of students who engage only with the summative assessment during that teaching week and with the HELM quizzes later on during the semester, mainly during the revision period.

We believe that the difference in students' engagement between semester 1 and 2 is a reflection of the quality of the workbooks and students may have relied on other sources, such as the textbook. The development of some of the HELM sections, which were used during semester 2, happened during semester 1 and less time was available to us to develop the semester 2 materials. So, the semester 2 workbooks were less complete in terms of written solutions and robust checking. Improving quality of HELM materials was our goal during the summer of 2021.

We have previously mentioned in [24] that developing online educational resources, especially online assessments, ideally needs three phases.

1. Creating the questions prior the teaching period.
2. Monitoring the questions during the period they are used ensuring any ambiguities and technical problems are resolved immediately.
3. Revising the questions in the light of their use during the previous teaching period, identifying any weakness and adding better feedback.

The STACK software generates detailed statistics on what students actually type in. In many situations our prior professional experience can suggest which feedback will actually be needed by students, and we can add this to the questions in advance. Often students surprise us with solutions we did not expect, and only the phase 3 revision will reveal such answers, and allow us to improve the feedback in light of what the students actually did.

7 Next steps

The original HELM project was a major collaborative effort which has earned respect for the quality of the original materials, and remained popular for many years. Our short-term goal was to collaboratively create online workbooks in a style similar to that of other successful courses at the University of Edinburgh, see [3]. The medium-term plan is to make a general release in early 2022, once materials have been tested for at least one academic cycle. Partners have agreed to release the resulting online versions of each HELM workbook they write, including all the online questions. We agreed to use the Creative Commons Attribution licence (CC-BY, https://creativecommons.org/) and make the source of the questions available with the STACK software itself. We will post updated on the further development of HELM materials on https://stack-assessment.org/ in the future.

The ultimate goal is to complete any workbooks for which there is a genuine ongoing demand. We appreciate that the mathematical needs of future engineers is remarkable stable, (e.g. see [14], [15]) and so there is likely to be enduring demand for much of the original HELM materials. We also appreciate there will need to be changes to the workbooks, and completely new workbooks. In particular, statistics and data science are a much more prominent part of engineering education than when HELM was conceived. We hope that future projects will contribute new workbooks to complement the existing HELM materials, rather than starting entirely from scratch. There is a natural framework, style guide and constituency for such material.

We are also mindful that the education of engineers has common ground internationally, and anticipate that HELM could be valuable in non-English speaking countries. STACK support multi-lingual materials, and HELM might prove to be a valuable resource if/when translated into other languages.

8 Conclusions

The project of transforming the HELM workbooks into interactive online workbooks showed that creating quality learning materials requires a team effort. First we need a robust, user-friendly and feature-rich online platform, like Moodle together with support for highly interactive mathematical questions, such as STACK. Second we need well-designed educational materials, such as the HELM workbooks. Third we need a tested delivery format like the one we designed for FAC, which promotes student engagement with the content and promotes mastery learning. And forth we need a consistent style guide, robust quality control processes and ultimately a high standard of delivery. As mentioned by [25] the quality and design of the assessment tasks are directly related to students' learning and empowerment.

The use of the workbooks in our teaching during the academic year 2020–21 showed that our students valued them as an important educational resource. Our students appreciated the interactivity of the workbooks, in particular they appreciated the immediate feedback on their attempts.

The preliminary evaluation of the workbooks that we used at the University of Edinburgh during the academic year 2020–21 gave us a better insight of the resources and helped us to identify "weak" questions which we update during the summer of 2021.

9 Acknowledgments

We would like to thank all the Edinburgh and Loughborough interns who worked on this project.

Interns from Loughborough University: Curtis Hooper, Damien Crossland.

Interns from the University of Edinburgh (2020): Anna Mkrthchyan, Nivedita Viswanatha, Miles Savva, Rueben Wheeler, Yvonne Calo, Luis Varga Mieles, Charlotte Summers, Danny Walker, Albert Sola Vilalta, Ivona Gjeroksa, Ruth Reynolds, Tom Byrne, Trang Nguyen, Mildred Aduamoah, Will Reynolds, Kai Hugtenburg. Malthe Sporing for helping creating the style guide and reviewing workbooks for style.

Interns from the University of Edinburgh (2021): Ilyas Nicholson, Adrián Doña Mateo, Alex Andrews.

10 Disclosure

The authors report there are no competing interests to declare. This work was funded by the authors' institutions: The University of Edinburgh and Loughborough University. The source code of the materials developed as part of this project will be released as part of the STACK project: see https://github.com/maths/moodle-qtype_stack

11 References

[1] R. Gratwick, G. Kinnear, and A. K. Wood, "An online course promoting wider access to university mathematics," *Proceedings of the British Society for Research into Learning Mathematics*, vol. 40, no. 1, pp. 1–6, 2020. [Online]. Available: https://bsrlm.org.uk/wp-content/uploads/2020/05/BSRLM-CP-40-1-04.pdf

[2] G. Kinnear, A. K. Wood, and R. Gratwick, "Designing and evaluating an online course to support transition to university mathematics," *International Journal of Mathematical Education in Science and Technology*, vol. 0, no. 0, pp. 1–24, 2021. https://doi.org/10.1080/0020739X.2021.1962554

[3] C. J. Sangwin and G. Kinnear, "Coherently organised digital exercises and expositions," *PRIMUS*, 2021. https://doi.org/10.31219/osf.io/jhngw

[4] D. R. Sadler, "Formative assessment: revisiting the territory," *Assessment in Education: Principles, Policy & Practice*, vol. 5, no. 1, pp. 77–84, 1998. https://doi.org/10.1080/0969595980050104

[5] P. Dawson, M. Henderson, P. Mahoney, M. Phillips, T. Ryan, D. Boud, and E. Molloy, "What makes for effective feedback: staff and student perspectives," *Assessment & Evaluation in Higher Education*, vol. 44, no. 1, pp. 25–36, 2019. https://doi.org/10.1080/02602938.2018.1467877

[6] M. Weurlander, M. Söderberg, M. Scheja, H. Håkan Hult, and A. Wernerson, "Exploring formative assessment as a tool for learning: students' experiences of different methods of formative assessment," *Assessment & Evaluation in Higher Education*, vol. 37, no. 6, pp. 747–760, 2012. https://doi.org/10.1080/02602938.2011.572153

[7] B. S. Bloom, "The 2 sigma problem," *Educational Researcher*, vol. 13, no. 6, pp. 4–16, 1984. https://doi.org/10.3102/0013189X013006004

[8] N. J. Cepeda, E. Vul, D. Rohrer, J. T. Wixted, and H. Pashler, "Spacing effects in learning a temporal ridgeline of optimal retention," *Psychological Science*, vol. 19, no. 11, pp. 1095–1102, 2008. https://doi.org/10.1111/j.1467-9280.2008.02209.x

[9] R. F. Hopkins, K. B. Lyle, J. L. Hieb, and P. A. S. Ralston, "Spaced retrieval practice increases college students' short- and long-term retention of mathematics knowledge," *Educational Psychology Review*, vol. 28, no. 4, pp. 853–873, 2016. https://doi.org/10.1007/s10648-015-9349-8

[10] P. Kirschner, J. Sweller, and R. Clark, "Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching," *Educational Psychologist*, vol. 41, no. 2, pp. 75–86, 2006. https://doi.org/10.1207/s15326985ep4102_1

[11] G. Kinnear, "Delivering an online course using STACK," in *Contributions to the 1st International STACK conference 2018 in Fürth, Germany*. Zenodo, 2019. [Online]. Available: https://doi.org/10.5281/zenodo.2565969

[12] L. Mustoe, *Maths for and science engineering*. LTSN MathsTEAM Project, 2003.

[13] M. D. J. Barry and N. C. Steele, "A core curriculum in mathematics for the european engineer," Société Européenne pour la Formation des Ingenieurs (SEFI), Plymouth, UK, Tech. Rep. 92.1, 1992.

[14] L. Mustoe and D. Lawson, "Mathematics for the European Engineer: A curriculum for the twenty-first century," SEFI Mathematics Working Group, Tech. Rep., 3 2002, iSBN 2-87352-045-0.

[15] B. Alpers, "A Framework for Mathematics Curricula in Engineering Education: A report of the mathematics working group," SEFI Mathematics Working Group, Brussels, Belgium, Tech. Rep., 2013, iSBN 978-2-87352-007-6.

[16] D. David Green, M. Harrison, A. Palipana, D. Pidcock, and J. Ward, "The helm mathematics learning and assessment regime for engineering students," in *Proceedings of International Conference on Engineering Education*, 10, 2004.
[17] C. J. Sangwin, *Computer Aided Assessment of Mathematics*. Oxford, UK: Oxford University Press, 2013. https://doi.org/10.1093/acprof:oso/9780199660353.001.0001
[18] 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*, vol. 94, pp. 205–222, 2017. https://doi.org/10.1007/s10649-016-9725-4
[19] C. J. Sangwin and M. Harjula, "Online assessment of dimensional numerical answers using STACK in science," *European Journal of Physics*, vol. 38, no. 3, pp. 1–12, 2017, 035701. https://doi.org/10.1088/1361-6404/aa5e9d
[20] C. J. Sangwin, Proof Technology in Mathematics Research and Teaching, Mathematics education in the digital era. Cham, Switzerland: Springer International, 2019, vol. 14, ch. Reasoning by equivalence: the potential contribution of an automatic proof checker, pp. 313–330. https://doi.org/10.1007/978-3-030-28483-1_15
[21] C. Sheard, C. Desvages, and D. Jordan, "The ASID project: A student-centred outlook on the development of hybrid teaching and learning," University of Edinburgh, Teaching Matters Blog, 2021.
[22] E. Acosta-Gonzaga and N. R. Walet, "The role of attitudinal factors in mathematical on-line assessments: a study of undergraduate stem students," *Assessment & Evaluation in Higher Education*, vol. 43, no. 5, pp. 710–726, 2018. https://doi.org/10.1080/02602938.2017.1401976
[23] S. A. Zabala-Vargas, L. H. García-Mora, E. Arciniegas-Hernandez, J. I. Reina-Medrano, B. de Benito-Crosetti, and A. Darder-Mésquida, "Strengthening motivation in the mathematical engineering teaching processes – a proposal from gamification and game-based learning," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 6, pp. 4–19, 2021. [Online]. Available: https://doi.org/10.3991/ijet.v16i06.16163
[24] C. J. Sangwin and K. Zerva, "Developing online learning materials to support undergraduate education at the University of Edinburgh," *Mathematics Today*, vol. 56, no. 3, pp. 98–101, 2020.
[25] M. S. Ibarra-Sáiz, G. Rodríguez-Gómez, and D. Boud, "The quality of assessment tasks as a determinant of learning," *Assessment & Evaluation in Higher Education*, vol. 46, no. 6, pp. 943–955, 2021. https://doi.org/10.1080/02602938.2020.1828268

12 Authors

Konstantina Zerva studied Physics at University of Ioannina, Greece and I received her PhD in Experimental Nuclear Physics in 2013 from the same University. After that, she worked as a Teaching Assistant at the School of Physics and Astronomy and at the School of Mathematics at the University of Edinburgh. Since 2017 she has been working as Facilitator in Technology Enhanced Mathematical Education at the School of Mathematics at the University of Edinburgh. Her main role is to design and develop online assessments for the undergraduate Mathematics courses using the STACK assessment system and also to give practical support to colleagues about the best use of e-assessment. She has a growing interest in scholarship of learning and teaching and how technology and e-learning can be used to support learners and educators. In 2019 she was awarded a fellowship in the Higher Education Academy.

Christopher Sangwin joined the University of Edinburgh in 2015 as Professor of Technology Enhanced Science Education. His learning and teaching interests include (i) automatic assessment of mathematics using computer algebra, in particular the development of the STACK system, and (ii) problem solving using student-centred approaches. Chris Sangwin is the author of a number of books, including "Computer Aided Assessment of Mathematics'' and the popular science book "How Round is Your Circle'', which illustrates and investigates many links between mathematics and engineering using physical models. He's also co-authored a number of school textbooks published by Haese Mathematics.

Ian Jones is a Reader in Educational Assessment at the Mathematics Education Centre, Loughborough University, UK. He obtained his PhD in Mathematics Education at the University of Warwick and worked as a Research Fellow at the University of Nottingham before moving to Loughborough University as a Senior Research Fellow funded by a Royal Society Fellowship. His primary research interests are in assessing conceptual understanding, applying comparative judgement methods to address education research questions, and children's development of algebraic thinking.

David Quinn joined the University of Edinburgh in 2015 and has been teaching engineering and mathematics students in higher education in Scotland since 2012. He achieved fellowship of Higher Education Academy in 2019. David is currently the organiser of four non-specialist mathematics courses offered by the School of Mathematics.

Article submitted 2022-09-10. Resubmitted 2022-10-20. Final acceptance 2022-10-24. Final version published as submitted by the authors.