

# Chapter 7

## Practice and Practise in University: What Defines Success and How Does Online Assessment Support Achieving This?

C.J. Sangwin

**Abstract** This chapter reviews current assessments in technical subjects such as science, technology, engineering, mathematics, or computer science in higher education. We consider definitions of success in terms of stated policy aims and defacto examination practice. From this background we examine the rapidly changing nature of learning materials from a static physical textbook to online resources with interactive components. Automatic online assessment is a key component of contemporary online learning. The chapter describes some of the current range of tools, from those targeted as very specific subjects to more general methodologies such as comparative judgement. We consider how these new tools will shape students' work and how this might contribute to, or hinder, their success.

**Keywords** Higher education • Online assessment • Learning resources • Assessment design • STEM education

What does it mean to “succeed” in a technical subject such as science, technology, engineering, mathematics, or computer science (STEM) at the undergraduate level? At one level “success” can be defined precisely as doing what you are told to do in high-stakes assessments. However, professional practice may well have other characteristics, the assessment of which are not well served by traditional examinations.

This chapter will review definitions of success in university STEM disciplines looking at published policy statements such as benchmarks and professional society accreditation criteria. I then turn my attention to current assessment practice and what current examinations actually assess. From this background I look at the changing nature of learning resources in science subjects, particularly in mathematics, which include more interactive online activities including online assessment. How are online tools changing the activities students undertake, and the feedback they receive? How do these changing tools match up with the published criteria they seek to serve?

---

C.J. Sangwin (✉)

School of Mathematics, University of Edinburgh, Edinburgh, UK  
e-mail: C.J.Sangwin@ed.ac.uk

Criteria which define success can be found in published subject benchmark statements, which aim to describe the nature and characteristics of university program. In particular they try to describe the standards, in terms of attributes and capabilities, which need to be obtained in order that someone achieves an award of a degree. UK mathematics has, for example, Lawson et al. (2015), and for engineering education (Best et al. 2015; Alpers 2013; Lucas et al. 2014). In computer science the concept of computational thinking is emerging (Wing 2008). To cater for the variation in the background of the incoming students, universities, collectively, offer a very broad range of mathematics and statistics programs. Given this breadth of programs, Lawson et al. (2015) goes no further than specifying subject content as follows “Common ground for all programs includes calculus and linear algebra”. Indeed, what is striking about curriculum documents in all STEM subjects is the lack of emphasis on specific curriculum content. Instead, these standards provide more general guidance and articulate the intended learning outcomes.

Computational thinking is a kind of analytical thinking. It shares with mathematical thinking in the general ways in which we might approach solving a problem. It shares with engineering thinking in the general ways in which we might approach designing and evaluating a large, complex system that operates within the constraints of the real world. It shares with scientific thinking in the general ways in which we might approach understanding computability, intelligence, the mind and human behaviour. (Wing 2008, p. 3717)

Instead of curricula content, all these documents talk in more general terms such as developing habits of mind, and the importance of setting up problems (modelling), mastering techniques for solving particular classes of problems, and the ability to critically discuss whether the solutions to the model fit the real-world problem adequately. For example, engineers should “be skilled at solving problems by applying their numerical, computational, analytical and technical skills, using appropriate tools” (Best et al. 2015, p. 7). The report by Kilpatrick et al. (2001) discussed “five tightly interwoven” threads which make up mathematical proficiency. The first two, *conceptual understanding*, and *procedural fluency*, are relatively well established. For example, Sfard (1991) discussed concepts and concept formation in mathematics in some detail. *Strategic competence* is defined as the ability to formulate, represent and solve problems which arise in real-world situations. *Adaptive reasoning* is defined as the capacity for developing arguments and thinking about whole arguments, including logic, explanation and justification and reflection. *Productive disposition* is confidence in one’s ability and an inclination to see mathematics as sensible and worthwhile. These threads elaborate on earlier frameworks and distinctions, such as that between *relational understanding* and *instrumental understanding* developed by Skemp (1971). While these threads are important to educational research, and influence some research-led teaching, it is not clear that they strongly influence current assessment design, particularly in examinations.

Central to the act of teaching are activities for students to undertake which are likely to produce the desired learning outcomes in students. These activities form the core of formative assessments through which students engage with the subject.

For assessment to be effective both the teacher and the student must accept joint responsibility. The teacher is responsible for structuring the enabling conditions and the student for engaging with them (Biggs et al. 2001). Furthermore, summative assessment aims to select and grade students' performance, and the results de facto indicate whether a student has successfully completed their studies. Benchmark statements also acknowledge the importance and difficulties of assessment. Typically, STEM subjects have a wider distribution of marks than humanities subjects, with some students achieving near-perfect solutions meriting very high marks while other students struggle even to get started on a problem (Lawson et al. 2015, p. 21). In the United Kingdom, a full-time undergraduate program of study is typically made up from individual 10 or 20 credit modules totalling 120 credits per year. A survey of assessments used in university mathematics departments in England and Wales is reported in Iannone and Simpson (2012). Of the 1843 individual modules they examined, over one quarter were assessed entirely by closed-book examination and nearly 70 % used closed-book examinations for at least three quarters of the final mark. It is still the case that success in university mathematics degrees is defined by traditional examination outcomes, and this is likely to also be the case across STEM.

What do these examinations actually ask students to do? To answer this question, Smith et al. (1996) developed a taxonomy of question types, with the goal of using this to construct examinations which assessed a range of skills. Pointon and Sangwin (2003) applied a very similar taxonomy to 486 questions taken from first year university examinations and found that 61 % of the marks for questions required only routine calculation. A further 20 % of the questions required proof, but these proofs tend to be rather well rehearsed. It is not clear the extent to which these are, for the students taking them, a memory test or require a genuine attempt to write a proof. As Smith and Petocz (1993, p. 139) say "Most students do no more than learn proofs by rote, reproducing them as necessary in their examinations, often with mistakes". Note that the examinations analysed by Pointon and Sangwin (2003) did not appear to require students to demonstrate much strategic competence beyond selecting a technique from a well-specified repertoire. Given these assessments appear to me mainly procedural, it is not clear that students taking these courses have a serious opportunity to develop the kind of productive disposition to mathematics which Kilpatrick et al. (2001) envisaged. "Far too often it seems textbooks and examinations would seem content to set only straightforward questions on technique requiring little in the way of a synthesis of ideas and knowledge" (Howson 2013, p. 655).

Across the STEM disciplines, policy documents set out the broad characteristics, such as problem solving, which are cited as valued by the professions. These concepts and higher order competencies are largely based on a foundation of lower level skills. "To be *trained* is to be prepared against surprise. To be *educated* is to be prepared for surprise" (Carse 1987, p. 23). That is to say, training provides specific knowledge and expertise which will be useful in the future. Encountering new situations is often uncomfortable, and training not only avoids this discomfort but enables individuals to respond effectively and efficiently within well-understood

domains. Indeed, the whole purpose of agreed and published engineering standards is to avoid every engineering situation becoming a novel problem-solving exercise. There is safety in working to a proven recipe. In order to practise a skill you need appropriate tasks, and tasks have to be assessed. Therefore such training is valuable, and it is very well served by existing traditional examinations. In this Chap. 1 consider the changing nature of resources in the form of textbooks and newer online assessment systems which are designed to support the associated assessment activities. Many of these resources are designed to help students develop basic skills in mathematics, that is, they are designed as a traditional training in mathematical techniques. In addition, the educated graduate still needs the resources, emotional as well as skills, to respond to a surprising situation outside their training. In due course, I discuss the Moore teaching method in which the primary activity in the class are problems which the students undertake. I then discuss how technology is being used to scale the assessment of students' work in problem-solving classes.

## The Changing Nature of Learning Resources

Until very recently the primary source of practice tasks was printed traditional textbooks. There has been considerable research into mathematics textbooks as artefacts to support teaching and learning, developing theoretical frameworks through which to consider interactions between the teacher, the students, the textbook and mathematics itself (Shield and Dole 2013). A model for how textbooks are used, framed within activity theory, was developed by Rezat (2006) who acknowledged that “the textbook is a historically and culturally formed mediating artefact”. Rezat (2009) he emphasises the use to which textbooks are put by students and concludes with insights into dispositions towards mathematics. “Learning mathematics comprises mainly learning rules, applying rules and worked examples to tasks, and developing proficiency in tasks that are similar to teacher mediated tasks” (Rezat 2006, p. 1267).

Unprecedented changes are taking place to the nature, production, and distribution of textbooks, both at school and in universities. To appreciate the profound speed and scale of this change it is important to understand that, in the United Kingdom at least, mathematics textbooks have been remarkably stable. That is to say, historically there were few books, they were very widely used and they were in print for many years. Many were in print for over half a century, and so it is instructive to retrace history for some considerable time to justify the stability and longevity of mathematics texts. As a starting example take Hutton (1836), which was in print from 1798 until at least 1849. The author, Charles Hutton, wrote a number of very popular and influential textbooks of which Hutton (1836) was his last major work.

Like the *Dictionary* and the *Treatise on Mensuration* it was to have great effect on mathematics education, not only through its many editions, but also for the influence it had on succeeding writers. Countless examples could be given of the nineteenth-century authors who in their writings give credit to Hutton and cite these works as sources for their material. (Howson 2008, p. 69)

Only a couple of years before this the first edition of James Wood's text was also published (Wood 1801). Wood's book was first published in 1795 and remained in print until 1876 (81 years), however during that time the book had numerous editions and revisions. The preface to the 13th edition (p. vi) states that during the period 1801–1848, 32,000 copies were printed in 13 editions. Wood died in 1839 and his book continued to be edited by Thomas Lund who gradually introduced new material and continually revised the text. Two other very popular textbooks, Hall and Knight (1962) and Bonycastle (1836), also have others contributing after the death of one author. This is a contemporary phenomenon, for example Erwin Kreyszig (1922–2008) first published his popular *Advanced Engineering Mathematics* in 1962. Posthumous editions, with new material, continue to be published in his name, for example the 10th Edition of the international student version in May 2011.

Not only were books in print for many years, but there is strong evidence for stability of text and exercises. In particular, there evidence can be found for the influence of one author on another by looking at acknowledged interdependency. Both Hall and Knight (1896, 1962) were in print for over 50 years, and the authors acknowledge previous authors in the preface to their book.

In enumerating the sources from which we have derived assistance in the preparation of this work, there is one book to which it is difficult to say how far we are indebted. Todhunter's *Algebra for Schools and Colleges* has been the recognized English text-book for so long that it is hardly possible that anyone writing a *text-book* on Algebra at the present day should not be largely influenced by it. (Hall and Knight 1896, p. vii)

Todhunter (1897) had five editions between 1858 and 1897 (39 years) and Barrow-Green (2001, p. 189) suggests that the total British sales exceeded 150,000. Todhunter was accused directly by Lund of plagiarism (Barrow-Green 2001, pp. 197–198) and perhaps as a response he acknowledges his sources more fully than many.

The chapters on Surds, Ratio, and Proportion in my Algebra are almost entirely taken from Dr Wood's Algebra. I have frequently used Dr Wood's examples either in my text or in my collections of examples. Moreover, in the statement of rules in the elementary part of my book I have often followed Dr Wood, as, for example, in the Rule for Long Division; the statement of such rules must be almost identical in all works on algebra. (Todhunter 1897, p. vi)

Indeed, there are sections which appear to be copied verbatim. Evidence can also be found for international influence of one algebra textbook on another, for example, Bonycastle (1836) was also influential in the United States. "It is evident that Bonycastle's text was the first popular algebra textbook used in American

Schools. . . . This book to a considerable extent set a pattern for the early algebras to be used in the U.S” (Nietz 1966, p. 48).

Bonnycastle’s algebra is described by Heller (1940) as a watershed in mathematics textbooks because he pioneered the systematic use of exercises. In particular Day (1820), which was abridged and published as Thompson (1848), was an enduring and very popular textbook based on Bonnycastle. Heller (1940) also examined the exercises in a wide range of algebra textbooks and traced heredity in the exercises these books contained.

### *Disruptive Technology*

These textbooks represent an identifiable and continuous chain of history from Bobbycastle in the late 1790s until the 1960s, a period which includes the French Revolution and the start of the space race. While there were certainly other textbooks, the presentation of algebra in the most popular textbooks was remarkably stable during this period. We can be confident that the exercises these books contain was the staple mathematical diet of generations of school students in algebra. In the United Kingdom, this tradition came to an abrupt and identifiable end with the School Mathematics Project, which set out to depart from traditional textbooks.

The Project was based on the work of individual teachers in schools, not of university lecturers or members of committees nor self-professed “educationalists”. And the numbers were huge. In the first decade roughly fifty were involved in the writing and testing of text books; over two thousand had attended the teacher-training conferences; ten times as many would have used or had contact with, the SMP books in classrooms up and down the country. . . . One of the original authors recently wrote “. . . I realise now how idealistic we were. We set out to create exercises where no two questions looked the same so that students were faced with new challenges all the time. This was a reaction to the Durell type texts which had long exercises of very repetitive questions. . . .” (Thwaites 2012, pp. 139–140)

A *disruptive innovation* helps create a new market and value network, and eventually disrupts an existing market and value network by displacing an earlier technology (Bower and Christensen 1995). The SMP was disruptive in the sense of the style of material and in involving large numbers of teachers in the development of the books. The SMP also ran workshops and other events for teachers’ professional development. However, SMP retained a very traditional book format through a respected publisher as a commercial venture. Publication of mathematics textbooks is currently in the process of much more profound disruption. Do current students really want a textbook which is a large physical volume? The convenience of mobile devices for reading and searching has eclipsed the need for large reference works. For example, the *Encyclopaedia Britannica* was in print for 244 years, but ended print production in 2012. Wikipedia is arguably easier to search and access than dozens of large physical static volumes. Indeed, many contemporary textbooks already have a digital version and many have companion activities such as online assessments.

A radical contemporary example is Mooculus, a portmanteau of “MOOC” (Massive open online course) and “calculus”, see <http://mooculus.osu.edu/> (retrieved June 2016). At the heart of this project is a 258-page traditional calculus textbook, presented as a PDF file. What is particularly unusual about this textbook is its completely open nature. The entire typescript of the book, in an editable format (LaTeX), is available for download. The book is licensed under Creative Commons (see <http://creativecommons.org/>, retrieved June 2016). As the name implies, Mooculus is much more than a textbook. The website provides access to an open calculus course, including online video lectures, online assessment exercises, and interactive “explorations”. The explorations are interactive online activities and include graph plotters, step-by-step solvers, and other visualisation tools. There are also opportunities for students to submit edits and changes, although at the time of writing there appear to be few recent edits, suggesting in this case that the opportunity for edits does not necessarily result in large-scale community engagement. That said, the free availability of such books potentially disrupts the commercial business model of publishers, and provides students with free access to high quality books online.

Mooculus shares the software which delivers and assesses online assessments with the very popular Khan Academy, (<http://www.khanacademy.org/>, retrieved June 2016). The Khan Academy offers rather traditional skills-based practice exercises, instructional videos, and a “personalized learning dashboard” through an online website. The dashboard tracks users’ mastery of skills and aims to “empower learners to study at their own pace in and outside of the classroom”. Originally focused on mathematics, Khan Academy now additionally includes work on science, computer programming, history, art history, and economics.

Khan Academy has short instructional videos and exercises as its central feature, rather than a textbook. Indeed, the Khan Academy abandons a linear structure, giving users more choice over which topics to study and when. Users are rewarded with “badges” and “energy points” for completing assessments. Collecting these is undoubtedly motivating for some students, and the popularity of the site is indicative that its materials have fulfilled a perceived need by many of its users. The scoring of energy points brings mathematics closer to an online game.

Computer games are a serious business, and many people of all ages and backgrounds play computer games on a regular basis. Just as with novels, music, and literature, computer games are becoming acknowledged as culturally important activities and experiences. Of course, that does not automatically make computer games high art, but nor does it permit a continuing view of computer games as trivial. Games are big business and they are as diverse as their players. It is therefore not surprising that some educators look to games to promote learning. For example, Devlin (2011) considered the characteristics of an effective educational game in mathematics, and in doing so criticised the design of many contemporary mathematical games. In particular he criticises those who confuse mathematics itself with its representation, for example symbolic or diagrams. He also questions the value of skills-based practice of, for example, multiplication tables or basic algebra; see Cayton-Hodges et al. (2015) for a recent review of mathematical games. An early

example of a mathematical game, explicitly not about skills practices is *L–A Mathematical Adventure*. Released in 1984, this classic text-based adventure game contained a number of mathematical puzzles which had to be solved along the way. Remarkably, it was still available in 2016.

Khan Academy is aimed primarily at school students, whereas the calculus in Mooculus is appropriate for undergraduates. Mooculus and Khan Academy both have large teams of developers, combining subject experts, web developers, and teachers who monitor online discussion. Arguably this has always been the case, with book authors, typesetters, illustrators, and production specialists contributing to traditional book publication. These skills now have to be supplemented by additional expertise needed for interactive technology both for explorations and assessment. The dynamic nature of updates to the websites means that materials evolve over time, rather than having a static publication date.

These developments are mirrored by commercial publishers. A notable example is from Pearson. Their MyMathLab suite of products ties together online assessments with video, interactive materials, and traditional printed books. The online exercises are randomly generated from templates and come in a variety of styles, including multiple choice and algebraic input. There are online tracking tools to help keep students motivated, directing them to the next assessment and informing teachers of what each student has done. Teachers can create assessment regimens for their students online from pre-existing questions tied closely to the published textbooks; see <http://www.mymathlab.com/> (retrieved June 2016). Allen and Seeman (2013) provide a more general survey of the state of online learning in higher education in the United States.

### *Non-linear and Adaptive Learning*

Books are essentially a linear communication format and are often intended to be read in the order in which the author presented the material in the book. Online materials are potentially much more flexible, and the order in which material can be accessed (or made available) does not need to be restricted to a linear format. Adaptive learning systems change the order of presentation to take account of the previous interactions an individual has had. Central to adaptive learning systems is a detailed model of the skill a student is trying to learn. This model has to be expressible in well-defined sub-skills. The cognitive skills for mathematics and computer programming are well suited to this approach, and therefore naturally enough mathematics has been the subject of many projects which seek to automate tuition, see Sleeman and Brown (1982), Appleby et al. (1997) and more recently Heeren and Jeuring (2014). Central to the model are sub-skills which can be isolated. Questions are designed to test knowledge (in various senses) of these skills.

What is problematical is acquiring the procedural knowledge that enables this inert knowledge to become the basis for effective action in the context of use. Production rules



cannot be learned by simply being told. Rather, they are skills that are only acquired by doing. (Anderson et al. 1995, p. 171)

Tutorial software estimates the probability that the student has learned each of the rules in the cognitive model. In some systems the software estimates whether a student has learned an incorrect or “buggy” rule (Burton 1982). For example, in long subtraction of integers some students consistently take the larger digit away from the smaller digit. An answer such as  $654 - 496 = 242$  is consistent with this rule, and carefully designed questions can expose such misconceptions.

One lesson from the efforts to create adaptive learning and online tutoring software is the significant team effort needed to produce a working system. This needs (i) an expert knowledge model, (ii) a student knowledge model, (iii) a tutoring module, and (iv) an interface (Nwana 1990). Given the effort expended on developing these online tutors, it is important to reflect on whether expending a similar amount of effort on the design and development of more traditionally presented learning situations would demonstrate similar gains. Such control studies are rare, and difficult to conduct. What is really generating the learning gains? Is it the careful design, or is there something about being online? This remains to be seen.

What has become clear is that for a typical teacher, in a weekly teaching situation, it is impossible to develop online learning systems of a sufficiently high quality. This has been acknowledged by others, for example “The systems that we developed were inflexible in the way they had to be used and gave teachers no ability to tune the application of the tutors to their own needs and beliefs about instruction” (Anderson et al. 1995, p. 192). For this reason, more agile assessment systems are now in regular use which sometimes lack explicit models both of the student and the cognitive domain being learned. This does not mean they lack sophistication in a number of real senses. Many online assessment systems do have significant domain knowledge encoded, and examples will be given in the next section. This functionality is used to generate very specific formative feedback. In this sense they go well beyond what was possible in the 1990s with the technology available then.

## Online Tools for Assessment

Most online assessment systems are internet based, using a website which manages a student’s identity and tracks their progress through the learning materials. These materials often include online assessments. In some systems questions are provided as a quiz in a fixed linear structure, in others the system builds an internal model of the student’s strengths and weaknesses and the system adapts the subsequent choice of questions (Appleby et al. 1997). At some appropriate point, the student is expected to engage with assessment, and a core part of this requires them to answer

a question. Any online assessment system will have a variety of question types of which multiple choice questions (MCQ) is just one.

MCQ are commonly associated with online assessment. There is general dissatisfaction with multiple choice as an assessment format (Hassmén and Hunt 1994; Hoffmann 1962). The dissatisfaction includes potential problems with guessing, and with reverse engineering questions. For mathematics MCQ are particularly problematic as the relative difficulty of a reversible process is very different in opposite directions. For example, factoring a quadratic is more difficult than expanding out the brackets. When faced with a multiple choice question, the concern is that a strategic student does not answer the question as set, but checks each answer in reverse. This potentially reduces the validity of the question undermining the intentions of the teacher. To test the hypothesis that when faced with a question involving the inverse direction of a reversible mathematical process, a comparative experiment was undertaken to see whether there is evidence that students solve a multiple choice version by verifying the answers presented to them by the direct method, not by undertaking the actual inverse calculation (Sangwin and Jones 2016). This methodology compared students' answers on questions requiring a mathematical expression as an answer with responses to stem-identical multiple choice questions. The findings supported this hypothesis: overall scores were comparatively higher in the multiple choice condition, but this advantage was significantly greater for questions concerning the inverse direction of reversible processes compared to those involving direct processes. For example when asked to factor polynomials the evidence supports the hypothesis that students expand out the answers rather than actually factoring the given expression. To address these problems, a variety of very subject specific-question types have been developed. I consider just two of these, one for assessing students' ability to write fragments of computer code and the other which assesses answers which consist of mathematical expressions, for example an equation.

### *Online Assessment of Coding*

Many students, including the majority taking STEM subjects, learn to program a computer as part of their degree. A basic skill is the ability to write short fragments of code, for example conditional statements, loops, and functions. When writing computer code in software engineering various developers have automated the assessment of fragments of students' code. For a review see Ala-Mutka (2005), with more recent examples in Usener et al. (2012) and Helminen et al. (2013). The students must enter a syntactically valid fragment of source-code which compiles correctly to object-code. Once this is done, the software automates testing the code to establish if it has the correct input–output behaviour. There are also various open sites which enable and encourage students to learn how to write code. For example, Codecademy [sic] <https://www.codecademy.com/> (retrieved March 2016) is an interactive website that offers free coding classes in various programming

languages including Python and PHP. It includes rewards such as badges to motivate users and social elements including a user forum where participants can discuss coding and gain help.

## Assessment of Mathematics

Over the last 25 years there has been a growing community of practice of automatic assessment, in which a student enters an answer which is a mathematical expression, and software establishes the mathematical properties of that answer using computer algebra; see Sangwin (2013, Chap. 8) for a recent review. There are many systems implemented, but most have the following characteristics. Internally there is a question template from which the software generates a random version of the question in a structured mathematical way and automatically generates a full worked solution which reflects this randomisation. The student solves the given problem, perhaps using a pen and paper in the traditional way, or using computer algebra as a tool. Typically the student must enter an algebraic expression into a computer as their answer. Systems vary on precisely how students enter their answer, with the most popular options being a typed linear syntax or a drag and drop equation editor. Once the system has a syntactically valid expression it automatically establishes mathematical properties of this answer using a computer algebra system. On the basis of properties established (or not) the system generates outcomes, including feedback and a score. The system stores data on all attempts at one question, or by one student, for later analysis by the teacher. As a typical example of contemporary assessment software for mathematics I consider STACK, a project which I designed, implemented and maintain. This system uses the computer algebra system Maxima to support the mathematical processes (Sangwin 2013).

A typical assessment situation is shown in Fig. 7.1. A randomly generated question has been displayed and the student has entered their final answer as an

**Fig. 7.1** Example assessment of the final answer using STACK

The screenshot shows a STACK assessment interface. At the top, it says "Find the integral:" followed by the mathematical expression  $\int 3 \cdot x \cdot e^{3x} dx$ . Below this is a text input field containing the student's answer:  $x^2 \cdot e^{(3 \cdot x) + c}$ . A feedback box below the input field states: "Your last answer was interpreted as follows:" followed by the expression  $x^2 \cdot e^{3x} + c$ . A "Check" button is visible below the feedback box. The bottom section of the interface has a yellow background and contains the following text: "Incorrect answer. The derivative of your answer should be equal to the expression that you were asked to integrate, that was:" followed by the expression  $3 \cdot x \cdot e^{3x}$ . Below this, it says: "In fact, the derivative of your answer, with respect to  $x$  is:" followed by the expression  $3 \cdot x^2 \cdot e^{3x} + 2 \cdot x \cdot e^{3x}$ . At the very bottom, it says "so you must have done something wrong!"

algebraic expression. This particular process, symbolic integration, is examined in virtually all traditional calculus courses of which I am aware. Furthermore, this kind of question is also typical of assessment of the kinds of reversible process for which MCQ are so problematic. In this case the expression is judged to be valid, and so the system has assessed this as an answer. Again, in the example shown in Fig. 7.1 the student is incorrect: the answer is not an integral of the expression in the question. The feedback shown in Fig. 7.1 is very specific to the student's actual answer, indeed it has been generated from a computer algebra calculation of the symbolic expression entered by the student. Note also in this example that the teacher has chosen not to display numerical marks to the student at this time. Whether formative feedback, as shown, or a numerical mark are available during or after the quiz is a choice which teachers need to make in each individual situation.

I note that it is not inevitable that technology will be used only to replicate tasks which test simple procedural skills. Technology such as STACK can be used to assess answers for a variety of tasks which are not traditionally set because they require the teacher to undertake a significant computation to establish the properties of the student's answer (Sangwin 2003).

Although online assessment is described as automated the teacher remains responsible. That is to say, when authoring the question the teacher must encode criteria which establish whether or not an expression is correct. The prototype mathematical properties include (i) algebraic equivalence with the correct answer and (ii) that it is written in an appropriate algebraic form, (for example factored). A computer algebra system is readily able to establish such properties, but note that using a CAS is much more sophisticated than using a string match or regular expression. It is also possible to encode criteria which establish if a particular answer appears to arise from a common mistake or misconception. If a student's answer appears to arise from a misconception or satisfies only a subset of the required properties, the teacher is able to encode the award of partial credit or feedback. Potentially automatically generated feedback is specific to the answer and directly related to possible improvement on the task, which is precisely the kind of feedback which research such as Kluger and DeNisi (1996) has suggested is most effective in a formative setting. Partial credit reflects a subjective value judgement, and few colleagues agree on the relative merits of partially correct answers and how many marks they should receive. Whatever decisions are made, because the criteria are objective and specified in advance the assessment is highly reliable.

As open source software which is freely available it is difficult to know how many people actually use STACK. In the year ending 1st April, 2015 STACK was downloaded 10,168 times, but this does not equate to numbers of live servers. To gather data from users, I undertook a survey during May 2015 (Sangwin 2015). There were 40 participants who used STACK and who completed a substantial part of the survey, and STACK is currently being used in eight languages. This collaboration on assessment infrastructure indicates a truly international endeavour. The survey also asked respondents to describe how they used STACK and the responses are shown in Table 7.1. The majority of users indicated both setting

**Table 7.1** Purposes of STACK use

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 %)

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. Although the initial goal for developing such software was formative, it seems inevitable that some mathematics examinations will be conducted entirely online using this type of technology. I am aware of other, similar, software being used for summative examinations in mathematics; for example, Ashton et al. (2006) reported trials of automatic assessments in Scottish secondary school mathematics. When this happens, the attention of students and teachers will be focused much more keenly on CAA as an assessment format.

### ***Problem Solving***

When discussing problem solving a distinction is often made between a *problem* and an *exercise*. A problem is a question for which the process for solving it is unclear. Therefore it is impossible to classify a question as a problem or exercise: it is as much a function of the *particular student* as it is the mathematical processes which lead to a correct solution. As a consequence, one person's problem is another's exercise. This precisely encapsulates the distinction of Carse (1987) quoted above between education and training. Training transforms problems into exercises, but this cannot continue to happen for ever. Hence the question remains: how can genuine problem solving be taught?

There are many attempts to teach students to become more effective at solving genuine problems. For example, the Moore Method (Coppin et al. 2009), is a type of enquiry-based learning (EBL) developed by the influential Texan topologist Robert Lee Moore (1882–1974) for university mathematics courses (Parker 2004). Essentially, a Moore Method class works in the following way.

1. Problems are posed by the lecturer to the whole class.
2. Students solve these independently of each other.
3. Students present their solutions to the class, on the board.
4. Students discuss solutions to decide whether they are correct and complete.

Solutions are not imposed or provided by the lecturer, who chairs discussion before offering their own comments. The essential difference between a Moore

Method course and other problem-based learning approaches is the use of a coherent set of problems on a substantial mainstream curriculum topic, rather than isolated/independent problems, puzzles or investigations. One misconception regarding Moore's Method is that Moore simply stated axioms and theorems and expected students to develop the complete theory. Parker (2004) suggests that Moore actually gave significant help to his students, but that he managed to do so in a way which did not rob them of the intrinsic satisfaction which can be derived from having independently solved a problem. Moore was particularly successful in attracting and encouraging postgraduate students, many of whom adopted his teaching approach. As a result, variations of this method are still used, particularly in the USA, and named after Moore. This particular approach is cited here because it is one with which I have personal experience. After 6 years of running a Moore Method class I am surprised at the stability of the class and the consistency of the outcomes. Indeed, each year I ended up about two problems from the same place with little or no effort on my part to set a pace for the work. Students undergo a personal transformation as they develop their approaches to solving problems. However, this is not an easy process and the following caricatures the cycle of the class.

- Week 1: Anticipation. "What is this class going to be about?"
- Week 2: Excitement and enthusiasm. "Someone is going to take me seriously and this sounds like fun!"
- Week 3: Frustration. "Actually I'm finding these problems a bit difficult!" "So-and-so's presentation was awful. What a waste of time!"
- Weeks 4–5: Despondency, Doldrums and Despair. "I can't do these!"/"They can't do these!"
- Weeks 6–7: Rebuild confidence. "Actually, I can do some of them".
- Weeks 8–9: Adjust expectations. "Problem solving takes time, so how many problems do we expect to do?"
- Weeks 10–11: Collegiate conviviality. "So let's get on with it..."

This class takes a considerable amount of time, and students typically solve only one or two problems per week. This class is explicitly not about covering material efficiently in a traditional way. Every generation of students is likely to need to struggle to develop their own abilities. There is some irony, perhaps, that in order to become proficient and confident in problem solving you need to practise solving problems. Problem-solving classes are also likely to be effective only when the group size is relatively small, that is, of the order of 12–20 students in each group. Furthermore, this is not something which can be done either as a one-off activity or as an extra set of more difficult optional problems. Students need to be immersed in an environment where they are expected to attempt to solve problems themselves, where they need to make partial attempts and where they need to criticise the attempts of their peers. Such classes are much more expensive to run than traditional large lectures. While some colleagues do question whether an institution can afford to run such small classes, my view is that we cannot afford *not* to run them.

Clearly the choice of the problems in a problem-solving class is a key aspect. How does the teacher choose the right problems? Just as with traditional mathematics textbooks, there is a remarkable stability in the problems which have been used with the explicit intention of improving students' problem-solving abilities. For example, a version of the following problem first appeared in Europe in Alcuin of York's *Problems to Sharpen the Young*, written around AD775 (Hadley and Singmaster 1992) and has remained a popular problem ever since (Swetz 2012).

A dog starts in pursuit of a hare at a distance of thirty of his own leaps from her. If he covers as much ground in two leaps as she in three, in how many of his leaps will the hare be caught?

Such questions are a part of mathematical culture and history, and have formed part of the education of many generations. Such problems are enduring cultural artefacts, just as are poems.

It is well known that many students find moving from word problems to setting up an equation which accurately represents it to be very difficult. Clement et al. (1981, p. 288) proposed the following problem to 150 undergraduate students taking calculus: "Write an equation to express the statement '*There are six times as many students as professors at this university*'. Use S for the number of students and P for the number of professors". Of the 150 calculus students, 37 % answered incorrectly and two thirds of incorrect answers were literal translation resulting in the equation  $6S = P$ . See also Fisher (1988) who reported that students continue to perform poorly when attempting it. The majority of students also have difficulty with basic logical reasoning (Wason 1968). There is a growing body of research on how students learn to reason, and the psychological and cognitive basis for problem solving.

Colleagues have varied aspects of Moore Method, with some encouraging students to work as a group, both answering questions and formulating research topics of their own. Such teachers encouraged alternative solutions to be presented and discussed, helping students refine their sense of aesthetics and providing other strategies for subsequent problems. In all forms, a key aspect is that it is the students' responsibility to solve the problems for themselves. And, in all versions *the group* criticises these solutions and ultimately, together with the teacher, decides if a solution is complete and correct. This more social notion of correctness is somewhat at odds with the objective testing of routine problems. How, then can technology be used to assess problem solving such as this?

### ***Comparative Judgement***

The online assessment reported in previous sections has concerned developments which aim to provide very specific assessments in individual subjects. The more sophisticated the tool, the narrower the range of subjects, which can be assessed. At one extreme are multiple choice systems: very general but rather limited. At the

other extreme are very powerful assessment systems for individual subjects which are able to provide specific feedback and complex interactions but are often rather inflexible. Despite the efforts of developers to date, few systems genuinely assess higher order skills. For example, rather than attempting to assess solutions to complex problems, they typically assess answers to more routine problems. As I argued earlier, sustained problem solving is valued and expected by benchmark statements. Traditional examinations, and contemporary online assessment systems focus on skills. This is valuable up to a point, but what about online assessment of problem solving? Assessment of problem solving is also not well served by traditional examinations, but can online technology offer a novel opportunity? All the previous systems attempt to establish objective properties of the student's answer. By contrast, a quite different approach is used in comparative judgement (CJ).

In comparative judgement assessors are presented with pairs of student scripts and asked to decide which of the two students has performed "better" (Pollitt 2012). Ties are not permitted. The outcomes of many such judgements are combined to create a scaled rank order of scripts from "worst" to "best". In psychology the Law of Comparative Judgement (Thurstone 1927), is based on the robust finding that people make much more reliable judgements when comparing one thing with another than they are able to do when trying to make objective judgements of a single item in isolation. Comparative judgement makes use of this law to establish the relative merits of students' work. Comparative judgement appears to be robust even in the absence of precise assessment criteria, such as may happen during problem solving.

Work such as Jones et al. (2014) and Jones and Inglis (2015) investigated the use of CJ for assessment of answers in mathematical problem solving. They found that CJ does offer a mechanism which enables the design and reliable use of more open problems, such as occur in mathematical problem solving. CJ is being used in a wide range of disciplines, including those essay subjects where objective assessment criteria are much harder to specify. CJ is also being used in peer assessment, where students judge one another's work (Jones and Alcock 2014).

For comparative judgement to work, each script has to be used in a sufficient number of judgements, typically of an order of magnitude of 10. Computer technology is able to orchestrate the process of presenting work to a judge in a form suitable for an efficient judgement to take place, such as placing two photographs or text paragraphs side by side on a screen. Computer technology is also ideal for calculating the statistics. Note that the duality between scripts and assessors means that CJ can also be used to rank order the judges themselves, enabling the quality of assessment to be measured, and ineffective judges (who guess perhaps) to be eliminated from the statistics. Although the idea of CJ has been around for nearly a century, only with computer technology has it become really practical as a mainstream assessment format. Note that comparative judgement may be a useful tool for high-stakes assessment, but it is not designed to be able to give detailed or specific feedback, which other bespoke software is designed to do.



## Conclusion

In this Chap. 1 reviewed the definitions of success in university STEM subjects, as articulated in published policy statements. I also considered current assessment practices, in which the traditional timed and unseen examination still predominates. This provided a background against which to consider the changing nature of learning resources, with a move away from static traditional textbooks to more dynamic online resources incorporating sophisticated interactive assessments.

This is a time of rapid change in the nature and availability of resources. At the same time, assessment formats are changing from paper and pencil work to a wide variety of online assessments. Students expect high-quality materials, and they are used to working online. CAA is currently most useful for formative assessment of core skill-based tasks. STACK, as shown in Fig. 7.1, is a typical example here. There appears to be a disconnect between what can be assessed by current technology and tasks which assess the stated goals found in published policy statements. In particular, current technology and current examinations focus on questions which test routine procedural skills at a range of complexity from simple to involved. Published policy statements speak in broader terms, particularly highlighting the important role of problem solving. Current exams often depend on short precise items perhaps to achieve acceptable scoring reliability. Comparative judgement appears to offer one promising solution to the problem of assessing more open-ended problems.

Regardless of the medium used for material—for example printed textbook or online materials—the quality of the curriculum design, the presentation and the assessment will be key in helping students engage with the subject matter: that quality of experience is key in retaining their interest and ensuring their success.

## References

- Ala-Mutka, K. (2005). A survey of automated assessment approaches for programming assignments. *Computer Science Education*, 15(2), 83–102. doi:10.1080/08993400500150747
- Allen, I., & Seeman, J. (2013). *Changing course: Ten years of tracking online education in the United States* (technical reports). Massachusetts, USA: Babson Survey Research Group and Quahog Research Group.
- Alpers, B. (2013). *A framework for mathematics curricula in engineering education: A report of the mathematics working group* (technical reports). Brussels, Belgium: SEFI Mathematics Working Group (ISBN; 978-2-87352-007-6).
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4(2), 167–207. doi:10.1207/s15327809jls04022
- Appleby, J., Samuels, P. C., & Jones, T. T. (1997). DIAGNOSYS—A knowledge-based diagnostic test of basic mathematical skills. *Computers in Education*, 28, 113–131.
- Ashton, H. S., Beevers, C. E., Korabinski, A. A., & Youngson, M. A. (2006). Incorporating partial credit in computer-aided assessment of mathematics in secondary education. *British Journal of Educational Technology*, 27(1), 93–119.

- Barrow-Green, J. (2001). 'The advantage of proceeding from an author of some scientific reputation': Isaac Todhunter and his Mathematics Textbooks. In *Teaching and learning in nineteenth-century Cambridge (History of the University of Cambridge)* (pp. 177–203). Cambridge, UK: Cambridge University Press.
- Best, R., Cleland, D., Cooke, G., Fidler, K., Grieg, A., O'Doherty, D., et al. (2015). *Engineering* (technical reports). Gloucester, UK: The Quality Assurance Agency for Higher Education.
- Biggs, J., Kember, D., & Leung, D. Y. P. (2001). The revised two-factor study process questionnaire: R-spq-2f. *British Journal of Educational Psychology*, *71*, 133–149.
- Bonycastle, J. F. (1836). *An introduction to algebra* (16th ed.). London, UK: Longman.
- Bower, J. L., & Christensen, C. M. (1995). Disruptive technologies: Catching the wave. *Harvard Business Review*, *73*(1), 43–53.
- Burton, R. R. (1982). Diagnosing bugs in a simple procedural skill. In D. Sleeman & J. S. Brown (Eds.), *Intelligent tutoring systems* (pp. 157–183). London, UK: Academic Press.
- Carse, J. P. (1987). *Finite and infinite games* (1st ed.). London, UK: Ballantine Books.
- Cayton-Hodges, G. A., Feng, G., & Pan, X. (2015). Tablet-based math assessment: What can we learn from math apps? *Educational Technology and Society*, *18*(2), 3–20.
- Clement, J., Lochhead, J., & Monk, G. S. (1981). Translation difficulties in learning mathematics. *American Mathematical Monthly*, *88*(4), 286–290.
- Coppin, C. A., Mahavier, W. T., May, E. L., & Parker, E. G. (2009). *The Moore method, a pathway to learner-centered instruction* (Vol. 75). Washington, DC, USA: Mathematical Association of America.
- Day, J. (1820). *An introduction to algebra, being the first part of a course of mathematics* (2nd ed.). New Haven: Howe and Spalding.
- Devlin, K. (2011). *Mathematics education for a new era: Video games as a medium for learning*. Wellesley, USA: A. K. Peters/CRC Press.
- Fisher, K. M. (1988). The students-and-professors problem revisited. *Journal for Research in Mathematics Education*, *3*(19), 260–262.
- Hadley, J., & Singmaster, D. (1992). Problems to sharpen the young. *The Mathematical Gazette*, *76*(475), 102–126.
- Hall, H. S., & Knight, S. R. (1896). *Higher algebra: A sequel to elementary algebra for schools* (4th ed.). London, UK: MacMillan (First published 1887).
- Hall, H. S., & Knight, S. R. (1962). *Elementary algebra for schools, containing a full treatment of graphs* (6th ed.). London, UK: MacMillan (First published 1885).
- Hassmén, P., & Hunt, D. P. (1994). Human self-assessment in multiple choice. *Journal of Educational Measurement*, *31*(2), 149–160.
- Heeren, B., & Jeuring, J. (2014). Feedback services for stepwise exercises. *Science of Computer Programming*, *88*(1), 110–129. doi:[10.1016/j.scico.2014.02.021](https://doi.org/10.1016/j.scico.2014.02.021)
- Heller, H. F. (1940). Concerning the evolution of the topic of factoring in textbooks of elementary algebra published in England and the United States from 1631 to 1890 (PhD). Columbia University.
- Helminen, J., Ihanntola, P., & Karavirta, V. (2013). Recording and analyzing in-browser programming sessions. In *Proceedings of 13th KOLI calling international conference on computing education research* (pp. 13–22). New York, NY, USA: ACM. doi:[10.1145/2526968.2526970](https://doi.org/10.1145/2526968.2526970)
- Homann, B. (1962). *The tyranny of testing*. London, UK: Crowell-Collier.
- Howson, G. (2008). *A history of mathematics education in England*. Cambridge, UK: Cambridge University Press.
- Howson, G. (2013). The development of mathematics textbooks: Historical reflections from a personal perspective. *ZDM*, *45*, 647–658. doi:[10.1007/s11858-013-0511-9](https://doi.org/10.1007/s11858-013-0511-9)
- Hutton, C. (1836). *A course of mathematics: In three volumes: Composed for the use of the royal military academy*. London, UK: Longman.
- Iannone, P., & Simpson, A. (2012). *Mapping university mathematics assessment practices*. Norwich, UK: University of East Anglia.

- Jones, I., & Alcock, L. (2014). Peer assessment without assessment criteria. *Studies in Higher Education*, 39(10), 1774–1787. doi:[10.1080/03075079.2013.821974](https://doi.org/10.1080/03075079.2013.821974)
- Jones, I., & Inglis, M. (2015). The problem of assessing problem solving: Can comparative judgement help? *Educational Studies in Mathematics*, 89(3), 337–355. doi:[10.1007/s10649-015-9607-1](https://doi.org/10.1007/s10649-015-9607-1)
- Jones, I., Swan, M., & Pollitt, A. (2014). Assessing mathematical problem solving using comparative judgement. *International Journal of Science and Mathematics Education*, 13(1), 151–177.
- Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington DC, USA: National Academy Press.
- Kluger, A. N., & DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284. doi:[10.1037/0033-2909.119.2.254](https://doi.org/10.1037/0033-2909.119.2.254)
- Lawson, D., Arrowsmith, D., Bailey, T., Griffiths, J., McAlinden, M., Osbaldestin, A., et al. (2015). *Mathematics, statistics and operational research* (technical reports). Gloucester, UK: The Quality Assurance Agency for Higher Education.
- Lucas, B., Hanson, J., & Claxton, G. (2014). *Thinking like an engineer: Implications for the education system* (technical reports). London, UK: Royal Academy of Engineering (ISBN: 978-1-909327-09-2).
- Nietz, J. A. (1966). *The evolution of American secondary school text-books: Rhetoric & literature, algebra, geometry, natural history (zoology), botany, natural philosophy (physics), chemistry, Latin and Greek, French, German & world history as taught in American Latin grammar school, academies and early high schools before 1900*. Rutland, USA: C. E. Tuttle.
- Nwana, H. S. (1990). Intelligent tutoring systems: An overview. *Artificial Intelligence Review*, 4, 251–277.
- Parker, J. (2004). *R. I. Moore: Mathematician and teacher*. Washington, DC, USA: Mathematical Association of America.
- Pointon, A., & Sangwin, C. J. (2003). An analysis of undergraduate core material in the light of hand held computer algebra systems. *International Journal of Mathematical Education in Science and Technology*, 34(5), 671–686. doi:[10.1080/0020739031000148930](https://doi.org/10.1080/0020739031000148930)
- Pollitt, A. (2012). The method of adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 19(3), 281–300. doi:[10.1080/0969594X.2012.665354](https://doi.org/10.1080/0969594X.2012.665354)
- Rezat, S. (2006). A model of textbook use. In *Proceedings of the annual conference of the international group for the psychology of mathematics education* (30th, Prague, Czech Republic) (Vol. 4, pp. 409–416).
- Rezat, S. (2009). The utilization of mathematics textbooks as instruments for learning. In *Proceedings of CERME6, Lyon France* (Vol. Working Group 7, pp. 1260–1269).
- Sangwin, C. J. (2003). New opportunities for encouraging higher level mathematical learning by creative use of emerging computer aided assessment. *International Journal of Mathematical Education in Science and Technology*, 34(6), 813–829. doi:[10.1080/00207390310001595474](https://doi.org/10.1080/00207390310001595474)
- Sangwin, C. J. (2013). *Computer aided assessment of mathematics*. Oxford, UK: Oxford University Press.
- Sangwin, C. J. (2015). *Who uses STACK? A report on the use of the STACK CAA system* (technical reports). Loughborough, UK: Loughborough University.
- Sangwin, C. J., & Jones, I. (2016). Asymmetry in student achievement on multiple choice and constructed response items in reversible mathematics processes. *Educational Studies in Mathematics*, [10.1007/s10649-016-9725-4](https://doi.org/10.1007/s10649-016-9725-4).
- Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, 22(1), 1–36.
- Shield, M., & Dole, S. (2013). Assessing the potential of mathematics text-books to promote deep learning. *Educational Studies in Mathematics*, 82(2), 183–199. doi:[10.1007/s10649-012-9415-9](https://doi.org/10.1007/s10649-012-9415-9)
- Skemp, R. R. (1971). *The psychology of learning mathematics*. Harmondsworth, UK: Penguin.

- Sleeman, D., & Brown, J. S. (Eds.) (1982). *Intelligent tutoring systems*. London, UK: Academic Press.
- Smith, G., & Petocz, P. (1993). Proofs: Teaching and testing—A tragedy in three acts. *International Journal of Mathematics Education in Science and Technology*, 25(1), 139–155. doi:[10.1080/0020739940250118](https://doi.org/10.1080/0020739940250118)
- Smith, G., Wood, L., Coupland, M., & Stephenson, B. (1996). Constructing mathematical examinations to assess a range of knowledge and skills. *International Journal of Mathematics Education in Science and Technology*, 27(1), 65–77.
- Swetz, F. (2012). *Mathematical expeditions: Exploring word problems across the ages*. Baltimore, USA: The Johns Hopkins University Press.
- Thompson, J. B. (1848). *Elements of algebra: An abridgement of Day's algebra* (10th ed.). New Haven, USA: Durrie and Peck, Loomis and Peck.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(4), 273–286.
- Thwaites, B. (2012). *The school mathematics project, 1961–1970: A decade of innovation and its sequel*. Cambridge, UK: Cambridge University Press.
- Todhunter, I. (1897). *Algebra for the use of colleges and schools* (5th ed.). London, UK: MacMillan and Co.
- Usener, C. A., Majchrzak, T. A., & Kuchen, H. (2012). E-assessment and software testing. *Interactive Technology and Smart Education*, 9(1), 45–56.
- Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20, 273–281.
- Wing, J. M. (2008). Computational thinking and thinking about computing. *Philosophical Transactions of the Royal Society A*, 366, 3717–3725. doi:[10.1098/rsta.2008.0118](https://doi.org/10.1098/rsta.2008.0118)
- Wood, J. (1801). *The elements of algebra: Designed for the use of students in the university* (3rd ed.). Cambridge, UK: Cambridge.