

Quantum: tests worth teaching to

Tim Oates, Cambridge Assessment
Robert Coe, Durham Centre for Evaluation and Monitoring (CEM)
Simon Peyton Jones, Computing At School and Microsoft Research
Tim Scratcherd, Naace
Simon Woodhead, Diagnostic Questions

V11 March 2016

Summary

Project Quantum will help computing teachers check their students' understanding, and support their progress, by providing free access to an online assessment system. The assessments will be formative, automatically marked, of high quality, and will support teaching by guiding content, measuring progress, and identifying misconceptions.

The new computing curriculum in England establishes computer science as an entirely new foundational school subject, alongside maths and natural science, but lacking the depth of experience in teaching and assessment they enjoy. Computing teachers need help, and they need it soon. Our new platform will give them access to high-quality but low-stakes formative assessment for the new computing programmes of study, and will do so in months not years.

Teachers will be able to direct pupils to specific quizzes and their pupils' responses can be analysed to inform future teaching. Teachers can write questions themselves, and can create quizzes using their own questions or questions drawn from the question bank. A significant outcome is the crowd-sourced quality-checked question bank itself, and the subsequent anonymised analysis of the pupils' responses to identify common mis-conceptions.

Opportunity 1: the new computing curriculum

In September 2014 the Department for Education launched an entirely new national curriculum for Computing, replacing Information and Communication Technology. Uniquely in the world, computer science is now taught to every child, at every level from primary onwards.

This change gives schools a huge challenge. It amounts to introducing an entirely new subject at school, and one in which few teachers have enough subject knowledge, let alone well-tested pedagogy. There is a very real danger that the reality on the ground will not match the vision set out in the new curriculum.

In this context, focusing on assessment could be an extremely powerful force for good:

- Every teacher must assess their pupils' progress, for both formative and summative purposes.
- To do so they will, among other strategies, develop or acquire questions, tests, practical exercises, and the like.
- The assessments that a teacher chooses to use will powerfully affect the students' perception of what they are supposed to learn; both they and the teacher will seek for them to do well in them.

- Well-designed assessments therefore directly support teaching and learning; they make tangible the abstract statements of the curriculum; they expose mis-conceptions, they validly measure students' progress. On the other hand, poorly designed assessments directly lead to mis-directed teaching and learning. In short, we need tests worth teaching to.
- Designing good assessment is *hard*. It is all too easy to focus on aspects that can readily be tested, rather than on the ideas and skills that we want students to learn. Moreover research shows us that formative assessment is under-developed in England.

So the opportunity is this: a cloud-based assessment platform for computing, with rich high-quality content, would be a powerful influence for good on teaching and learning in computing. Rather than requiring each teacher to invent their own tests (with accidental imperfections and undue variations), we will save them work by offering a bank of assessment items, of high quality, that they can use immediately, for free. Teachers are hungry for such a thing, especially since the abolition of levels, so uptake would be immediate.

There would be other benefits too. The new curriculum is only two pages long, so there is a huge distance between its high-level goals and the detailed reality of a classroom. One very effective and concrete way to explain what is meant by the new curriculum is to develop high quality tests that embody it. So the process of developing good questions will give rise to dialogue about what exactly it is that we want our students to learn – and that is a very good conversation to have.

Opportunity 2: crowd-sourced assessment, with quality control

It is striking how traditional assessment remains, even in these high-tech days. Two basic approaches are widespread:

- **Centralised.** Publishers or other commercial organisations pay experts to develop tests which, in the best case, are field-tested and refined. When done well, this can lead to high quality, but such an approach needs a high level of initial financial investment, and tests necessarily come on-stream quite slowly.

Moreover, quality is variable. CEM has done item analysis and evaluation work with a number of awarding bodies, and has found that even assessments that are the product of detailed and sophisticated processes often contain poor items.

- **Do It Yourself (DIY).** Individual teachers create their own assessments, using a range of tools from paper through to online, niche tools, and administer them to their own classes. Although many of these will be very good, there is usually a significant range in quality and many are not very good. Creating assessments in this way can also be quite time-consuming for teachers.

Technology advances present another alternative:

- **Crowd-sourced assessments.** Many teachers already adopt the DIY approach. A crowd-sourced platform would allow them to share that work with others, and benefit from the work of others. In particular:
 - Teachers could create their own assessment items
 - They can construct tests drawing both on their own assessment items, and on a rich bank of assessment items supplied by others.
 - They can administer these tests to their students, as low-stakes formative assessment, with immediate feedback to both the student and teacher.

- They can benefit from data analytics showing their class results, over time.
- They can get feedback about the effectiveness of their own questions.

Within this framework, opportunities abound. For example, adaptive tests, where the difficulty of the question is adjusted in response to the student's success rate on earlier questions; more advanced forms of feedback to teacher and student; mechanisms to allow teachers to offer feedback on questions, and recognition for contributing; and so on.

The big challenge in crowd-sourced assessment is **quality control**: some questions will be better than others at assessing the learning that the national curriculum envisions; others might be ineffective, or even misleading. "Crowd-sourced assessment" could easily mean "low-quality assessment".

This is where our project is unique. CEM's work in assessment development, and in working with outside assessment developers to evaluate and quality-assure their assessments, gives us some important insights. CEM has developed algorithms that will generate measures of the quality of assessment items and produce templates for reporting and feeding back those measures to the teacher who created the assessment, in ways that will make sense to someone who has no specific assessment expertise. Crucially, this feedback will allow the author to withdraw or improve the immediate assessment items, but also over time to learn about what works in assessment.

A full quality assurance process will also harness the power of networks of users to make valid judgements about the alignment of assessments with desired learning. In an open network not all users will always make sound judgements, but we will develop analysis tools that allow us to identify those whose judgements are reliable, and create recognition and incentives within the system to encourage them. There may also be a role for commissioned experts, at least in the early stages.

To our knowledge, such a platform has never been built or deployed before. However, the 2015 [Macintosh Commission on Assessment Without Levels](#) recommended that

"The Commission recommends the establishment of a national item bank of assessment questions to be used both for formative assessment in the classroom, to help teachers evaluate understanding of a topic or concept, and for summative assessment, by enabling teachers to create bespoke tests for assessment at the end of a topic or teaching period."

This project addresses that recommendation head on, at least for the formative assessment part.

What we propose to do

We propose to put these two opportunities together, by **developing a hybrid cloud-based platform for assessment in computing, with crowd-sourced assessment items, and integral quality assurance so that the best assessments are promoted and shared.**

We focus on computing, because there is such a crying need for high-quality assessment in computing that quick wins are well within reach. Moreover, computing teachers are excellent partners in such an endeavour, partly because they are relatively tech-savvy, and partly because they are so hungry for support.

Within our focus on computing, our scope includes:

- The whole computing curriculum, as described by the [computing programmes of study](#)
- Both primary and secondary stage.

However our medium-term aspirations are much broader. This project is a pilot for doing the same thing in other subject disciplines.

We aim to produce *both* a directly useful platform and content, *and* evidence to drive research.

- The online testing system will be of immediate and direct help to teachers. It will be free at the point of use, provided development and ongoing support costs can be met some other way. It will report individual results to students, and class results to teachers.
- The platform will generate massive, rich, constantly-evolving data set that we and other researchers can use to develop and refine the concept and realisation of crowd-sourced assessment.

If we are successful, we will simultaneously achieve three goals:

1. **We will directly support hard-pressed computing teachers**, by giving them a platform that they can use to administer tests; and by working with them as partners to author a bank of high-quality assessment items, with integral quality control mechanisms.
2. **We will improve teaching and learning in the new computing curriculum.** If our assessments express the vision of the new computing curriculum (and that will be a challenge), teachers can look to them for concrete guidance about what that vision means.
3. **We will demonstrate, at scale and for the first time, the possibilities offered by a cloud-based, crowd-sourced assessment platform.** These lessons can then be applied in other subjects, and in other countries.

Focus

We propose to make progress by maintaining a laser-like focus:

- **Low-stakes, formative assessment only.** Teachers will decide what tests they want to administer, and when; and they will choose which questions to include in those tests.

The data will not be used for high-stakes purposes such as holding teachers or schools to account; we will not make it available to managers (e.g. for teacher appraisal), or Ofsted inspectors (e.g. for school inspection).

However, the results *will* be used, in anonymised form, to feed data analytics to improve the quality of the item bank, and other research purposes.

- **Machine-assessable question forms**, such as multiple choice questions. This has the huge merits both of removing marking load from teachers, and of making data from student answers available in machine-readable form to the analytics that we will use to drive improvement in the quality of the assessment items.

We will begin with multiple choice (actually a rather subtler medium than is often supposed) but we certainly hope to explore other machine-markable forms such as numerical answers, choosing sets of responses, drag-and drop, and the like.

We know that this will not cover the full range of capabilities that teachers need to assess (e.g. teamwork). We hope to offer a useful tool, not a universal one.

- **Quality control.** *The principal distinguishing feature of our proposal is our commitment to quality control.* It's no good simply assembling a large corpus of bad questions. Specifically,

we intend to use the data at scale from students taking tests in a feedback loop, to identify questions that “work” (have discriminating power, are well correlated with hand-curated, expert-written questions, etc), analyse responses to identify common misconceptions, and so on.

There is a serious research question here: *no one has attempted to start from crowd-sourced assessment items and use data (along with expert intervention) to drive quality.* Can it be done? No one knows for sure but (a) CEM have encouraging preliminary work that makes us believe that it will work and (b) if CEM and Cambridge Assessment can't do it, no one can.

The public good

In its initial stages Quantum is funded by philanthropic donors (and perhaps public money), and will be fuelled by crowd-sourced questions written by volunteers. It would not be acceptable for the fruits of this freely-offered support to be locked up in a commercial product, however good it might be.

To that end, the most valuable (and most expensive) aspects of the project will be available to all.

- **The basic DQ platform itself will remain indefinitely available, for free, to UK computing teachers.** Some advanced aspects, not funded by Quantum, may eventually be charged for but the ability to administer tests to a class, drawn from all the questions we gather, and get simple analytics of the results, will remain free indefinitely.
- **We will make all the questions, and their answers, publicly available for free as a corpus, in machine-readable form.** What this means is that a competing platform could draw on the exact same item bank. This will not be a one-off; we will make it continuously available.
- **We will make data from thousands of students taking thousands of questions available, for free, in suitable anonymised form, to *bona fide* researchers.** This is the dataset that Durham CEM will be drawing on, but others will be able to do so too.
- **Durham CEM will publish their quality-control techniques in the open literature.** In particular, we do not propose to patent or otherwise encumber them.

This approach leverages the “gift economy”, in which CAS specialises. For example, the CAS community has over 3,000 classroom resources, which were developed by teachers and others, and are shared, for free, under a Creative Commons license.

Partners

This project is ambitious. It will make a substantial, near-term impact on the delivery of the new computing curriculum; and will develop an approach to assessment that can be used in other subjects and jurisdictions.

To underpin that ambition, the project is a collaboration between three complementary partners, each an internationally recognised leader in their fields, and each with direct involvement of very senior colleagues:

- **Cambridge Assessment** is internationally respected for assessment. **Tim Oates** CBE chaired the Expert Group providing advice to the DfE Review of the National Curriculum, and since 2006 has led the large research group in Cambridge Assessment, a non-teaching department of Cambridge University. Tim provides research support and policy advice to the UK

Government and a large number of jurisdictions around the world. He is a member of the Ofqual Standards Advisory Group and Fellow of Churchill College Cambridge.

- **The Centre for Evaluation and Monitoring** (CEM) is Durham University's educational research centre devoted to making a positive impact on the outcomes of children and young people across the world. By employing scientific, evidence-based methods, CEM provides educators (teachers, school leaders, education authorities and jurisdictions) with high quality data and guidance to inform their practice and policy with the aim of improving pupils' educational outcomes. CEM has a long track record of developing high-quality, innovative assessments. **Professor Rob Coe**, a UK leader in educational assessment, with a seat on the DfE's Commission on Assessment without Levels and Ofqual's Standards Advisory Group, heads the centre.
- **Diagnostic Questions** is an online assessment platform for crowd-sourced multiple-choice questions, developed by a small British start-up. It is already used by over 22,000 teachers, mainly in secondary mathematics, across 1,600 schools in the UK. Beyond simply recording students' answers, DQ's platform asks each student to explain why they thought their answer was correct, uses this to explain correct solutions to the student, and also helps teachers understand their students' misconceptions. **Dr Simon Woodhead** is the co-founder and CEO, and has a background in mathematical statistics.
- **Computing At School** (CAS) is a grass roots organisation that was at the epicentre of the reform of the Computing curriculum. It has 21,000+ members, about three quarters of whom are school teachers, and is probably the single most trusted organisation for computing teachers. **Professor Simon Peyton Jones** is chair of CAS, and chaired the working party that drafted the new programme of study for Computing.
- **Naace** is the Association for all those committed to improving outcomes for learners, through the use of Education Technology. Naace has around 3,000 members and partners drawn from all sectors of the Education Technology Community. Its two main concerns are helping the community understand how Education Technology creates added impact on learning, and how to plan and deliver all aspects of the Computing Programme of Study. To achieve these, Naace has a wide range of products and services for members and non-members.
- **Teachers and computing professionals.** We hope to work in partnership with teachers themselves, and with computing professionals outside education, in a collaborative endeavour to write, review, curate, tag, and improve questions on computing. Success will depend on building trust, in an area (assessment) where teachers often feel under pressure.

This combination offers research expertise, experience of assessment in depth and at scale, and unrivalled access to classroom computing. Specifically

- CAS has immediate access to a 15,000+ constituency of classroom teachers; they are hungry for help, relatively tech-savvy, and eager to help develop best practice in computing education.
- Teachers will only use assessments that they believe in. CEM, CA, and CAS are all highly trusted brands with enduring reputations; anything offered by this group will be trusted.
- CEM, CA, and CAS all have huge "reach", through their existing rich network of relationships with schools. Anything they offer is likely to see very rapid uptake.
- DQ has an existing platform that *already* fulfils the basic project requirements: crowd-sourcing, cloud-based, the logistics of accounts and data collection, and an active user base.

Funding

The Quantum project is entirely funded by generous donations from

- Microsoft
- Google
- ARM

Their donations are unrestricted, and do not encumber the project with commercial obligations, other than to acknowledge, with huge thanks, the generosity of this support.

What is the plan?

We plan an initial two-year pilot project, kicking off in mid-2016. We envisage four main work areas:

- **Content.** Sourcing, writing, curating, reviewing, questions about computing. “Crowd-sourcing” is not a magic bullet. We need to actively seek content, guide authors, give them feedback, review their work, look for gaps in coverage etc. This requires a serious investment of time from subject experts.

We will also run workshop to identify more clearly what it is that we are seeking to measure, especially in computer science. (The “constructs” in assessment jargon.)

- **Quality control.** Both the research and the actual execution of the new analysis and quality-assurance mechanisms will require significant resources. But the results will be of broad applicability, certainly beyond computing.
- **The platform itself.** DQ’s platform does a fair proportion of what we need, but we will need to modify and develop it further. For example, it only handles multiple choice questions right now, and we want to move beyond that.

The project *absolutely needs* a platform and DQ’s platform is a very good one; but (unlike some projects) we do not see the platform as the solution; it is just an essential vehicle.

- **Evaluation.** We must spend some significant effort evaluating the effectiveness and impact of our work.

Long term sustainability

In its initial stages, Quantum is a donor-funded project. But if it is successful, there will be the opportunity to develop the same ideas across other subjects and other countries. Donor funding will never scale to this level of ambition, and some form of business proposition the only way to make it sustainable.

We are not sure what form this will take; it will be a nice problem to have. The open-source movement itself has demonstrated that there are ways to make open-source projects into sustainable business propositions without losing their central ethos; it is just that some care is needed. But whatever form it takes, we will not compromise the principles set out above.