Proceedings Paper:

Rossiter, J.A. orcid.org/0000-0002-1336-0633 (2018) Using quizzes instead of paper based exams to assess control topics. In: Proceedings of the 2018 UKACC 12th International Conference on Control (CONTROL). Control 2018: 12th International UKACC Conference on Control, 05-07 Sep 2018, Sheffield, UK. IEEE , pp. 422-427. ISBN 978-1-5386-2864-5

https://doi.org/10.1109/CONTROL.2018.8516759

Using quizzes instead of paper based exams to assess control topics

J A Rossiter
Dept. ACSE
University of Sheffield. S1 3JD, UK
j.a.rossiter@sheffield.ca.uk

***Abstract*—This paper focuses on the potential and practice of alternative forms of assessment to the conventional written exam. Readers will be interested in issues such as why would a staff member want to replace the written exam and how can I be sure that the replacement is equally effective at capturing student attainment. The paper summarises a trial at the University of Sheffield where the written exam was replaced by a computer quiz which was marked automatically.**

***Index Terms*—Assessment, exams, quizzes, student attainment**

I. Introduction

There has been a relatively long history of using computer aided assessment (CAA) as part of a holistic approach to student assessment and learning. In the UK this was pioneeered by the mathematics community [3], [5], [6] but subsequently has become widespread both in Universities but indeed also in schools [4]. A number of learning and teaching projects and evaluations [7], [8], [11], [17] established that giving students access to an environment which allowed them to self-test their problem solving and numerical computations had significant positive benefits to overall learning and attainment.

Subsequently, the author and indeed many others have transferred these insights into more general engineering curricula and found similar benefits (e.g. [9], [10], [12]–[14]). In practice, the CAA component often takes the form of several low weighted quizzes spaced throughout semester. In some cases students are allowed to take the quiz as often as they like to improve their mark and thus motivate learning. In other cases, practice quizzes may be available for preparation, but the summative quiz may only be attempted once. Nevertheless, the common pattern is that students have access to an environment containing a large data base of questions where they get immediate feedback on their attempts. Consequently they can identify where they are struggling and either use the supplied answers to deduce where they are going wrong, or seek more focussed guidance in tutorials. A key underlying point is that they are in charge of providing their own feedback and thus more empowered to take control of their own learning.

A key advantage of CAA is that once the quiz is set up, the staff member need only update settings such as availability dates and time limits and thereafter, irrespective of the size of the class, the rest of the assessment is managed solely by the computer. Thus, a staff member who previously was faced with the dilemma of providing regular homework and feedback at the cost of significant time handmarking, can now provide the regular homework and feedback, but with minimal cost to their time, beyond the initial (admittedly substantial) effort in creating the CAA.

Alongside the use of CAA, on the whole staff have continued to use end of year examinations worth 60% or more of the module marks. A simple argument that is used (anecdotally), is that most quiz environments available to staff[1] only allow relative simplistic questions such as those which have simple algebraic answers. Thus quiz environments are easy to use for assessment of the low level learning outcomes of a module, but much less easy to use for higher level skills. A complete assessment must include a substantive component which looks at higher level skills such as problem solving, multi-step questions, insight, creativity and so forth. Moreover, with a typical quiz environment, the answer supplied by the student is right or wrong; there is no credit available for working.

Despite these observations, the author felt it was worth challenging the assumption that a simple quiz environment was not easily able to capture the assessment which a written exam does. That is, in a written examination a student may get marks for working even when all their calculations are incorrect if they have demonstrated core understanding of the mportant processes, algorithms and steps. This paper first presents the arguments for why the anecdotal assumption about the value of written exams is perhaps overstated and then considers how a limited quiz environment can still capture enough of the important observations to give an equally valid measure of student attainment. Some test cases are used to evaluate the proposed approach.

Section II looks at the limitations of conventional written examinations and section III then proposes and illustrates the sort of questions which will capture the same marking accuracy, but allow automation by computer. Section IV discusses the efficacy of this with real students and the paper then finishes with some conclusions.

II. Marking conventional written examinations

The paper is set in the context of control engineering topics and in the light of a much broader range of work looking at student education and engagement as a whole [15], [16].

[1]We discount cases such as maplesoft and stack [2], [5] which are more powerful but not available to most academics.

However, the focus here is solely on what would traditionally have been a written examination type assessment and not laboratories or other activities.

Written examinations, at least in principle, are subject to relatively strict quality assurance procedures (e.g. [1]). The examination paper along with indicative solutions and mark schemes are checked both internally by a staff member not involved in the module and also by an external examiner (usually a senior academic from a different institution). Without consent from both of these, the examination will not be accepted as suitable for module assessment.

Although the precise guidance placed on the auditors and examiner will vary across institutions and departments, common themes are:

- Progression from straightforward parts the majority of students can do successfully through to harder parts which only a minority will complete. Questions should therefore be able to distinguish different achievement levels marked against the module learning outcomes.
- Exams should not be excessively long with students able to complete them in well under the time available (usually about half the time is a common rule of thumb). Exams assess understanding and problem solving but not speed of writing!
- The draft solutions and mark schemes should be complete enough for the auditors to: (i) assess the effort involved in answering the question; (ii) the level of understanding required to complete each part and thus judge progression against learning outcomes and also; (iii) to act as an objective statement of how marks are awarded to ensure consistency.

A. Limitations of traditional exam marking

This latter point is particularly problematic when it comes to judging the worth of conventional examinations. Consistency of marking in engineering topics often reduces to awarding marks for successful completion of key steps, correct computations, or making key observations. A student who fails to correctly calculate a core variable or note a key observation may end up with zero, that is the mark scheme is often implemented as a binary process.

In principle one could argue that a student with *correct working* but an incorrect initial calculation should be allowed to score significant marks. However, the practicalities (time requirements) of following through student calculations from incorrect starting points for every student who does this means this is not feasible in general, especially with large cohorts. Moreover, later calculations in an algorithm are often invalidated by the use of an incorrect start point. Taken together, this means that in practice students who make incorrect calculations early in a question can often achieve only minimal marks for working.

B. Possible consequences of weaknesses in examination processes

Although written examinations still occupy the largest proportion of marks for most traditional engineering degrees, it would be incorrect to suggest that they are a highly accurate measure of student ability. They are a snapshot of a student's ability to solve, on a given day, a somewhat arbitrary selection of problems from the syllabus with arbitrary numbers and somewhat arbitrary markscheme. A student making a silly typo or mistake early on in a question may not score the marks their ability and knowledge deserves and a range of students making a range of different mistakes may end up with similar marks, irrespective of their abilities, due to the need for a mark scheme that can be implemented consistently. The examiner's hope is that, on balance over a large number of assessments, student marks will average out to represent what each individual deserves. Similarly, any given examination with a large number of students, is expected to give a range of marks to represent the range of abilities in the class, but this is due to averaging and need not imply accuracy for any given individual.

A final and very critical point is that class sizes have been steadily increasing in recent years so that exam marking has moved from 50-100 scripts to often 300+ or more. This means that marking by hand is become both too onerous and of course less reliable due to a combination of marker fatigue and/or the use of multiple markers.

III. USING QUIZZES IN LIEU OF WRITTEN EXAMINATIONS

In view of the observations in the previous section, a trial was undertaken with a multi-disciplinary cohort taking a core control module. The written exam was replaced by a computer quiz which is marked automatically, in its entirety. A major benefit is the saving of staff time marking with large cohorts and improved consistency of treatment of each student, offset partially by more time in developing the quiz than required to produce typical model answers.

The basic argument is that a computer is able to implement a rigid marking scheme more reliably than a human (who is prone to lapses in concentration). The differences between a rigid marking scheme and one that looks carefully at student work for evidence of understanding that is not explicit in the required answers is often minimal and thus inconsequential in the light of all the other approximations and arbitrariness in the assessment process.

The evidence for such a proposal is that the spread of marks delivered by a computer quiz should be similar in profile to those delivered by a traditional examination. Although the author's institution does not routinely scale marks to ensure consistency across modules, for Universities that do this, it is important that the marks have a sensible spread so that any scaling can be defended. More critically, a spread of marks is evidence that the assessment adequately distinguished, on average, between students of different abilities.

A. Designing computer based questions

Rather than taking the common model of multi-choice questions, the author decided to unpick a traditional examination question and identify which aspects would typically be on a mark scheme and then construct questions to capture whether the student deserved the mark for these.

- Many multi-step problems have mark schemes based around interim calculations. Consequently it is straightforward to set up quiz questions which assess the accuracy of the interim calculations.
- A typical question will have a few computations and or problem solving steps followed by evaluation. Evaluation is often presented as a short paragraph though in practice *ideal answers/mark schemes* are often just a set of bullet points or keywords. Assessment of this can be captured with multi-choice questions (select as many as apply) where a number (possibly a large number) of alternative interpretations are presented. To prevent guessing where several answers may be true, negative marking is applied to incorrect selections.
- In order to avoid issues whereby a student gets zero because of a silly error in step 1 which then impacts the whole question, a reasonable number of marks should be available for what I will call parallel computations. That is, the students do a number of straightforward independent computations (based on foundational knowledge for pass/fail) which can be assessed separately and then bring these together for the later and harder parts of a question. One example is Q1 in the next subsection which in affect has separate credits for each root-loci observation.
- Including some questions which depend on student working being perfect is still reasonable as these questions help distinguish between good and excellent student performance.
- It is straightforward in the quiz environment to assess relatively high level learning outcomes, for example linked to control design, by providing perfect bode plots for some cases and ask for calculations based on those plots. Such credits are then not dependent upon students having provided a correct sketch, thus breaking the dependence on student working. For examples see Q8-Q10 in the following subsection.

B. Examples of questions

This section gives a selection of questions to demonstrate how the quiz can be designed and implemented in an efficient fashion. The reader may note that by making the number of options in a single question very large, one can capture a large number of learning outcomes/skills and avoid breaking down problems into questions with ‘predictable’ answers. Nevertheless, once the student has performed the necessary computations (such as producing a sketch), selecting the correct options should be rapid.

Q1. Sketch the root-loci for $G(s)$ and select whichever of the following statements apply? Do not guess as incorrect answers will carry negative marks, so only select those you are sure are correct.

- The root-loci has 3 asymptotes.
- The root-loci has 4 asymptotes.
- The root-loci has 2 asymptotes.
- The asymptote directions are -180, 60 and -60 degrees.
- The asymptote directions are -180,+180, +90 and -90 degrees.
- The asymptote directions are 0, +120 and -120 degrees.
- The real axis between -2 and -1 is on the loci.
- The real axis between 0 and 2 is on the loci.
- The real axis between infinity and -4 is on the loci.
- The system is closed-loop stable with low values of gain.
- The system is closed-loop stable with high values of gain.
- The system is closed-loop unstable with low values of gain but closed-loop stable with high values of gain.
- For high values of gain, the closed-loop system has 1 unstable closed-loop pole.
- For high values of gain, the closed-loop system has 2 unstable closed-loop poles.
- For high values of gain, the closed-loop system has 0 unstable closed-loop poles.
- The compensator $K = (s + 4)/(s + 1)$ would improve the root-loci.
- Using positive feedback would improve the root-loci.
- The compensator $K = (s + 1)/(s + 4)$ would move the centroid of the root-loci asymptotes well into the LHP.
- The closed-loop system is expected to have smooth behaviour with low values of gain but will have non-oscillatory and divergent behaviour for high values of gain.
- The closed-loop system is expected to have smooth but slow behaviour with low values of gain but will also have convergent oscillatory behaviour for high values of gain.

Q2. Sketch the Bode diagram for $G(s)$ and hence select whichever of following statements are true. Do not guess as incorrect answers will carry negative marks, so only select those you are sure are correct.

- I expect the system to be unstable with unity negative feedback.
- I expect the system to have unstable behaviour with unity negative feedback and proportional compensator of 4.
- A lead compensator with appropriate pole and zero will stabilise the system for any chosen gain cross over frequency smaller than w=10rad/s.
- For this system, a lag compensator for increasing the low frequency gain should have corner frequencies in the region of 0.1rad/s.
- For this system, a lag compensator for increasing the low frequency gain should have corner frequencies in the region of 1rad/s.
- A good value for the frequency to centre a lead compensator would be around w=0.5rad/s.
- The margins for this system with a unity proportional gain are satisfactory but would be improved even further

with a slight increase in gain.

- The margins for this system with a unity proportional gain are poor but would be improved with a small decrease in gain (say 20 to 40%).
- The margins for this system with a unity proportional gain are poor but would be improved with a decrease in gain of at least a factor of 2.
- The margins for this system with a unity proportional gain are high and would be improved with a small increase in gain.

For the same system, calculated questions can be used to assess core computations, for example.

Q3. For the system $G(s)$ with unity negative feedback, where is the centroid of the asymptotes in the root-loci plot? Give your answer to three sig. fig.

Q4. In the Bode diagram of $G(s)$, what is the slope (in dB/decade) of the gain asymptote for $w < 1$? Give your answer to three sig.fig.

Q5. What is the asymptotic phase of $G(s)$ for low frequency (in degrees)? Give your answer to three sig.fig

Q6. What is the phase margin (in degrees)? Give your answer to two sig.fig.

Q7. The system is expected to unstable in closed-loop with unity negative feedback? True or False?

To remove the dependence on student working for assessing higher level learning outcomes, the examiner can provide a number of Bode/Nyquist or other diagrams and pose some questiosn arounds these, for example:

Q8. Your job is to design a lead compensator which improves the phase margin to around 60 degrees at w=0.5rad/s. How much phase uplift is required at this frequency?

Q9. What value of proportional compensator (in dB) will give a phase margin of about 50 degrees?

Q10. For the following Bode diagram of an open-loop stable $G(s)$, what would be the gain margin for $G(s)M(s)$, in dB where $M(s) = 6$?

IV. Evaluation

Exams based on the above philosophy were used early in 2018 in two different engineering departments (systems and chemical) within the author's university and this section displays a profile of the student marks. It may be conjectured that the computer based examination will likely mark slightly more harshly than a human in that if a numerical answer does not meet the required accuracy it will score zero, whereas sometimes examiners can be more lenient and award partial marks. Hence, the exam average may be slightly lower, however, we expect this to a small order affect, that is to affect just a few marks and for only a few students.

A. Illustration of mark profiles

For two different assessments, histograms of the marks are presented in figures 1,2. Two observations jump out.

1) The quizzes have been effective in giving a spread of marks with a well defined peak roughly in the middle of the class. Hence the quizzes are effective at distinguishing between students of differing competence levels.
2) Far more of the year 2 students scored low marks due to errors in elementary numerical computations of foundational material. This is an important weakness in the student body to be aware of which is made much clearer with the quiz data which penalises such errors more objectively than handmarking may do

Remark 1: *Many marks were lost by year 2 students were due to incorrect signs (a critical error as this is LHP or RHP that must be penalised) and other similarly careless typos which a competent student should spot immediately due to the inconsistency with other data they are producing. It is possible that a handmarked exam would offer a few partial marks in this scenario which would modify the histogram of figure 1 slightly to the right (say 2-3%) at the lower end.*

Remark 2: *Perhaps a reflection of the year 2 data in particular is a growing staff perception that students have become too dependent on tools such as MATLAB and consequently have lost the ability to perform straightforward computations such as gain and phase reliably and an inability to ask questions such as:* does my answer make sense?

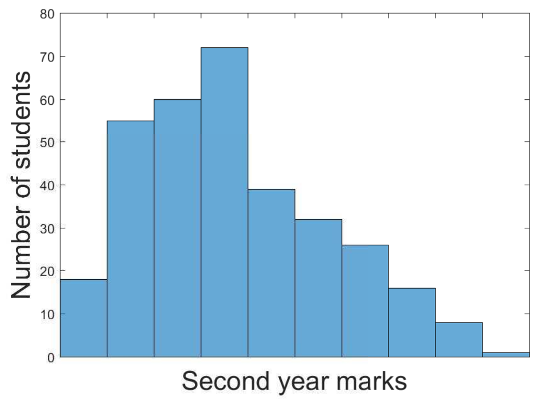

Fig. 1. Histogram of marks for examination 1 (year 2 students).

B. Reflections on the marking and auditing process

Staff who have been setting and marking exams for many years will recognise that a core process in marking is first to assess whether your anticipated marking scheme is fit for purpose. Over the process of marking the first 10 scripts it usually becomes transparent if some questions have not been interpreted as expected, perhaps for good reason, and consequently the mark scheme needs modification in order to be fair. Of course, one may even discover some typos that students did not notice but worked with. Unsurprisingly, with computer marked assessments this process is still needed!

It is critical to be able to revisit questions and in particular the chosen mark scheme after students have done the exam.

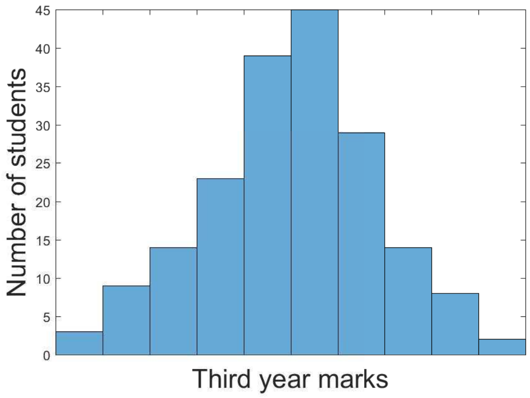

Fig. 2. Histogram of marks for examination 2 (year 3 students).

This should be performed alongside the scripts collected from the students using a sampling approach across a range of performance levels.

- Perhaps inevitably with a large number of questions, it is possible that some questions have minor numerical errors or assumptions that could critically affect student answers and thus the slack in the accuracy demanded may need to be modified. Sometimes this is only evident after looking though a few scripts after which the mark scheme can be updated. The computer re-marks all the submissions instantly!
- Looking through scripts also helps identify where questions were not as clearly phrased as originally thought perhaps leading to student confusion. Again, marking schemes may need modifying to take account of this. One example could be the marks for different multi-choice options.

The examiner also used the time looking at the hard copy scripts students submitted which contained all their working and sketches to compare marking against what the quiz had awarded. Specifically, the trial is interested in the validity of the quiz for giving a good representation of student competence, or alternatively, did the quiz fail to capture sound student understanding which may be evidenced on pen and paper? A large number of scripts were sampled and the conclusions were:

- Where students had numerous incorrect computations in the same thematic area, this was due to conceptual or other fundamental mistakes, so no credit could be awarded for working. If they has just a few incorrect computations and some correct, the overall mark they achieved was fair.
- For students with very low marks, that is a clear fail, hand marking may have given an extra 2-3% on the basis there was a semblance of some understanding and due to a desire to be generous, albeit the computations were mostly incorrect. In truth, this is irrelevant if students are far below the pass threshold.
- For students in the middle band of say 40-70%, there was rarely cases where the mark given would differ due to the evidence provided on the hard copy script. This is partially because the quiz tended to give a few extra marks for some parts and was far more directive about the key steps/computations required so in effect fewer marks were available for *working* and in general any differences thus cancelled out.
- For students scoring over 80% a bigger difference could be perceived as for the very hardest components, perfection is requested by the computer to get all the marks whereas a hand marker may be slightly more lenient.

In summary, there are minor differences between the marks a quiz delivers and what hand marking would deliver, but these differences did not effect critical decisions such as pass/fail and classification. Moroever, as one can argue that marking schemes and the marking process itself are always to some extent arbitrary, hence these differences are inconsequential as long as the marking scheme adopted is fair.

C. *How accurate is hand marking in practice?*

Many scripts are very scruffy and it is quite hard to detect or follow specific student solutions and thus marking is somewhat difficult as you are trying to make judgements with an inevitable inconsistency across different scripts. The author's experience is that a significant minority of scripts are simply unreadable and trying to give a mark which fairly represents the student understanding is very difficult and certainly unlikely to be precise. Forcing students to put final answer on the computer removes this vagueness and makes marking much more objective and thus it could be argued, more consistent and thus fairer.

V. Summary conclusions from staff perspective

The author's summary is that the trial was a success and delivered an effective assessment which could save significant staff time in the future as well as deliver marks reliably.

1) The assessment is just as fair as a hand written equivalent and delivers marks which properly distinguish different levels of performance.
2) For classes with 100s of students, there is a significant saving in marking time and indeed likely improvements in consistency of marking as human factors such as fatigue are avoided.
3) It is essential to do a sample marking after the exam with some of the collected scripts as this may bring to light issues with the phraseology of the questions and/or the accuracy demanded which necessitate a modification of the marking scheme.
4) Often handwritten scripts are very messy and hard to follow and thus mark objectively. The computer quiz removes this challenge and possible inconsistency.
5) Creating the exam takes slightly longer than a traditional exam paper due to the need to break questions down, but in truth this is only exposing the thinking an examiner should be going through anyway. This is more than

compensated for by the ability to do sample marking only after the exam.

REFERENCES

[1] Higher education funding council for England, http://www.hefce.ac.uk/
[2] Maplesoft, https://www.maplesoft.com/
[3] Mathcentre, http://www.mathcentre.ac.uk/ (Checked Dec 2017)
[4] MyMaths, https://www.mymaths.co.uk/ (Checked Dec 2017)
[5] Mathematics stack exchange, https://math.stackexchange.com/ (Checked Dec 2017)
[6] SIGMA, http://www.sigma-network.ac.uk/ (Checked Dec 2017)
[7] A.C. Croft, M. Danson, B.R. Dawson and J.P. Ward, Experiences of using computer assisted assessment in engineering mathematics, Computers & Education, Volume 37, Issue 1, August 2001, Pages 53-66
[8] D.A. Lawson, The effectiveness of a computerassisted learning programme in engineering mathematics, International Journal of Mathematical Education in Science and Technology Volume 26, Issue 4, 1995.
[9] J.A.Rossiter, L. Gray and D. Rossiter,Case studies of the resources students use, IFAC world congress, 2005
[10] J.A. Rossiter, D. Rossiter and G. Diercks O'Brien, Experiences in the use of web-based delivery for first year engineers, WBE 2004, Innsbruck
[11] J.A. Rossiter and A.C. Croft, Engaging engineers in learning, HEA annual conference, Edinburgh, 2005
[12] D Rossiter and J.A. Rossiter, Applications of Online Pedagogy to a First Year Blended Learning Module Using a VLE, International Conference on Innovation, Good Practice and Research in Engineering Education, 2006.
[13] J.A. Rossiter and D. Rossiter, A blended learning module design approach to engage new students, Second International Blended Learning Conference, Bedford, 2007. (ISBN 978-1-902806-81-5)
[14] J.A. Rossiter, L. Gray and G Diercks O'Brien, A case study in using Theory of Change to improve teaching practice in a control department, International conference on engineering education, 2007.
[15] J. A. Rossiter, D. Giaouris, R. Mitchell, P. McKenna, Typical control curricula and using software for teaching/assessment: a UK perspective, IFAC world congress, 2008
[16] J.A. Rossiter, B. Pasik-Duncan, Br.LI. Jones, R.M. Murray, L. Vlacic, S. Dormido, A Survey of Good Practice in Control Education European Enginering Education Journal, to appear, 2018
[17] G. Sim, P. Holifield and M. Brown, Implementation of computer assisted assessment: lessons from the literature, Journal ALT-J Research in Learning Technology, Volume 12, Issue 3, 2004.