
Mapping University Mathematics Assessment Practices

Edited by

Paola Iannone
University of East Anglia

Adrian Simpson
Durham University



Chapter 23

Performance Assessment in Mathematics: Preliminary Empirical Research

Paola Iannone and Adrian Simpson

Abstract This paper outlines the experience of introducing a form of oral performance assessment into an undergraduate degree module. While oral assessment is commonplace in many countries, it has all but disappeared from undergraduate mathematics in the UK and we explore some of the issues regarding implementing this form of assessment, some of the potential advantages and how this particular form of oral assessment was used with a group of first year undergraduates. We discuss the outcomes in terms of students' performance and student and tutor views of the assessment process.

23.1 Introduction

The issue of how to assess undergraduate mathematics has been a significant one for decades. Most recently, Levesley (2011) noted a number of current challenges regarding assessment including

- the potential for conflict between mathematicians' ownership of assessment methods and the requirements of external quality assurance and management systems
- conservatism and risk aversion with universities
- the need to assess efficiently
- the need to avoid repetitive testing

Iannone and Simpson (2011) noted the assessment diet for students tends to be very restricted, with closed book examinations overwhelming all other forms of assessment. While most universities include elements of coursework, projects or

Paola Iannone, e-mail: p.iannone@uea.ac.uk
School of Education and Lifelong Learning
University of East Anglia

Adrian Simpson, e-mail: adrian.simpson@durham.ac.uk
School of Education
Durham University

dissertation assessment, the written examination contributes the vast majority of the marks towards the final degree classification.

In contrast, Stray (2001) noted that the oral examination was the norm until the start of the 20th century when it fell out of favour as a result of political issues and because written examinations were seen as more efficient. However, oral assessment in mathematics remains commonplace in many mainland European countries such as Italy, Hungary and the Czech Republic and there is evidence that they play a role in some US institutions, including as synoptic exams or oral assessments in group work (Gold, 1999).

We developed a project to explore how an oral assessment component might be introduced to a mainstream undergraduate mathematics module, by replacing one piece of coursework with a one-to-one tutorial in which students discussed the mathematics surrounding standard homework questions with a tutor.

23.2 Oral assessments - the research evidence

The concept of an oral component to assessment can cover a range of methods. Joughin (2010) contrasts three forms of oral assessment: *presentation* on a prepared topic (which may be individual or in groups); *interrogation* (from short form Q&A, through to the still ubiquitous doctoral viva) and *application* (where candidates apply their knowledge live in a simulated situation, e.g. when medical students diagnose an actor playing the part of a patient).

While presentations already play a role in many undergraduate mathematics degrees (Iannone and Simpson, 2011) our project was specifically focussed at an oral assessment of the second form. Rather than the pejorative term *interrogation* we call these *oral performance assessment*, since it requires students to work live on a problem with a tutor.

Joughin (1998) notes that oral performance assessments may have advantages in allowing probing of knowledge and more accurately resembling a real world problem solving situation in which one solves a problem in a dialogue with a colleague. Despite this, there is little literature examining oral assessment methods (Hounsell et al., 2007, note that only 2% of the works in their survey of assessment across all disciplines discussed oral performance assessment of this type).

Students' perceptions of oral assessment have been explored in a number of contexts. In social work, Henderson, Lloyd and Scott (2002) noted a marked difference between the generally negative views of them (particularly in relation to anxiety and usefulness) before they were undertaken, more positive ones afterwards and, once students had graduated and were practising professionals, a positive sense of their value and authenticity. Huxham, Campbell and Westwood (2012) similarly noted the issue of nervousness being tempered by an understanding of their usefulness and authenticity of oral examinations in a study involving biology students.

While there are some articles which discuss the implementation of oral assessment in mathematics (Gold, 1999; Nelson 2011), many of these are presentations

and do not address the performance element of oral assessment. There is little empirical research in the use of orals in mathematics, their outcomes and student attitudes to them. This project was designed to examine the implementation of one particular variant of oral performance assessment in an undergraduate mathematics module and describe the student and staff experiences of that implementation.

23.3 Pre - implementation issues

While many mathematicians express concerns about the ways in which undergraduate mathematics is assessed (LMS, 2010) and oral performance assessments have been proposed as potential solutions (Levesley, 2011), one area we were interested in was the practicalities of implementation and we did find some obstacles to implementing oral performance assessment. While all the staff involved were willing and, in many cases, generous in giving time to the project, there was a nervousness about permitting this form of assessment. These seemed to focus around four main areas: institutional constraints, fairness, anxiety and preparedness.

Some people involved were unsure where the authority lay to vary the assessment procedure or whether it would be acceptable to the institution. This may have been related to the short timescale in which the project was implemented, requiring making a minor variation to the assessment process during the academic year. The inertia in systems governing teaching and assessment practices in many universities can appear to lead to delays of years to get even minor changes approved and can stifle change (Bryan and Clegg, 2006). In our case, because the change was minor and the timescale tight, approval was given from the appropriate committee chairs in the University outside the usual change routes. There appeared to be no real barriers to implementation and some genuine interest was expressed in the outcomes of the project.

As we prepared the assessment, a number of people raised the issue of fairness. By their nature, oral performance assessment (as with other performance assessments such as music and drama performance, oral language tests, driving tests, etc.) cannot be truly anonymous and this can give rise to concerns about bias. However, one could argue that the potential for bias is impossible to eliminate from any assessment system: it is possible that one can even be biased by handwriting style with anonymised written examination scripts (Briggs, 1980). Thus the issue of fairness is really one of monitoring and moderation: if there are concerns about the fairness of marking with written work, it can always be re-examined and assessment procedures often undergo moderation to reduce any potential for bias. While having a second assessor in each tutorial was considered, given the existence of cheap, high quality video recorders, we instead opted to record the tutorials to enable marks to be challenged and moderated (with the understanding that they would be deleted at the end of the assessment process unless students explicitly agreed to their use for research). Another issue of fairness raised concerned students with English as an additional language. However, we felt that difficulties with English would be

equally likely in written submissions and the opportunity to help students express themselves orally might be easier in a suitably sensitive tutorial setting.

A number of people voiced concerns about the level of anxiety a one-to-one tutorial might cause. However, the regular tutors for the module noted that their previous encounters with students on an individual basis had shown they were more comfortable with direct conversation than speaking in a tutorial setting with other students present. This fits with the evidence from Marshall and Jones (2003) that while anxiety was higher for oral clinical examinations than written examinations, it was higher still for seminar presentations.

The anxiety level of oral performance assessment may be higher than that for written assessment partly because students have considerably less experience of them. While we had no opportunity to give students the chance to practise this form of assessment, as much as possible was done to deal with this. Assessors were asked to make the setting and initial part of the conversation as comfortable as possible, students were given a short talk about the planned assessment which emphasised the conversational nature of the tutorial and this was re-emphasised in the information about the organisation of the tutorials.

23.4 Implementation

The assessment took place in a first year module on graph theory. While not a compulsory module, it is taken by the majority of first year students, with 108 students registered. The marks for the module are made up from 10% for solutions to homework problems and 90% for a written examination taken at the end of the year. The oral performance assessment replaced one set of homework problems and took place during normal tutorial time (which in other weeks would have been used to discuss and return the homework problems).

Each student was asked to attend a 10 minute session described thus:

The idea of the one-to-one tutorial is to help you express what you understand, not to catch you out. It is perfectly acceptable to ask the tutor for help or for the tutor to give you guidance or to help correct any errors you've made which might make it difficult for you to get to the answer. You'll be able to use the blackboard or paper to write things down (though you shouldn't bring complete answers or notes with you - we want to talk to you about the problem and its solution, not just hear you read an answer out!).

As normal, they were set the homework problems during the previous week, so they had at least six days to work on them. The problems were the following:

- A: Prove that if a graph has at least 11 vertices, then either it or its complement must be non-planar.
- B: Show that every connected planar graph with less than 12 vertices has a vertex of degree 4 or less. [Hint: argue by contradiction to get a lower bound for the number of edges which contradicts the upper bound which follows from Euler's formula]

C: For each graph find a minimum spanning tree and prove it is unique:

- (a) Q_3 with the usual binary vertex label and weigh $w(ij) = i + j$.
- (b) K_5 with vertices $\{1, \dots, 5\}$ and weigh $w(ij) = i + j^2$ where $i < j$.

D: Draw all forests on 5 vertices and justify your answer.

The students were told that in the tutorial, of the four problems set, they could choose one to discuss first and then the tutor would choose another. Because two of the questions required a proof (A and B) and two required the use of an algorithm and some reasoning about the outcome (C and D), it was agreed that when the student chose a question from one pair, they would be assigned a question from the second pair through some random process (tossing a coin or drawing lots). Students were also informed that they would be videoed to allow for marks to be moderated.

Tutors discussed the idea of *contingent questions* – areas they could explore around the solution, depending upon the quality and form of the response the student gave. For example, in question D, depending on the kind of method employed, the student could be asked to explain how they would check to ensure that no two forests were isomorphic, how they would prove that their method gave an exhaustive list or how they might use their solution to estimate the number of forests on 6 vertices.

The tutorials normally took place in groups of 12-16. Five people acted as assessors (the course lecturer, two postgraduate tutors who normally ran the tutorials and the two authors). For each tutorial slot, four assessors were needed. However, since no extra marking was required, the resource overhead was not as large as it would first appear. The 16 hours of staff time normally used for marking work and delivering tutorials were replaced with just over 18 hours of staff time doing the one-to-one tutorials.

At the end of each 10-minute slot, the tutor awarded a mark based on an assessment matrix (Figure 23.1) which had also been shared with the students.

The week after the assessments were completed, students were asked to fill in a short questionnaire adapted from the Assessment Experience Questionnaire (AEQ from Gibbs and Simpson, 2003) comparing the one-to-one tutorial with written coursework. They were also all contacted to see if they were prepared to attend a short interview to discuss the assessment and their experience of it. The course lecturer and the two postgraduate students involved in conducting the assessments were also interviewed.

23.5 Outcomes

Each question was marked out of 5 and the performance was generally good. The mean marks for each question A, B, C and D were 4.00 ($\sigma = 0.93$), 4.16 ($\sigma = 1.09$), 4.06 ($\sigma = 1.01$) and 3.89 ($\sigma = 1.18$) respectively (with no statistically significant differences between any pair of questions) and the students averaged 7.96 across the

Grade Solution		Key ideas and application	Clarity and explanation
5	Complete solution outline given with no extra help needed	Clearly identified key ideas behind the problem and shown how they apply elsewhere	Explains clearly and concisely, even in unfamiliar areas
4	Complete solution given with some extra help	Identified key ideas or shown how solution approach might apply elsewhere	Explains clearly and concisely in prepared areas and generally clear elsewhere
3	Complete solution given with substantial extra help	Has identified some key ideas, but may not fully distinguish key ideas from calculations or details OR shown some sense of wider application of solution	Explanations need a little probing to clarify
2	Complete solution not obtained, but some key steps made without help	Does not have key ideas or any sense of wider application	Explanations need to be drawn out at length
1	Complete solution not obtained, but some key steps made with help	Does not have key ideas or any sense of wider application	Has difficulty giving any explanations

Fig. 23.1 Assessment Matrix

two questions they were asked. Over the previous 6 weeks of coursework (covering the graph theory component of the course) the average mark was 4.29 ($\sigma = 0.95$).

While the marks obtained for a given tutorial are highly dependent on the choice of questions, it is worth noting that the one-to-one tutorial marks and their spread seem broadly similar to those for previous written homework. The attendance was also similar to previous weeks: of the 108 students registered on the course, only 9 failed to attend their assigned one-to-one tutorial. This compares with an average of 8% each week failing to attend a tutorial and 9% failing to submit the homework.

The issue of anxiety is one which pervades the literature on oral assessment and had been a concern in planning, but only one student expressed any concerns before the assessment. One further telling indicator that the level of anxiety was much less than that supposed before implementation came from the administration of the videos. To comply with ethical research standards, students were asked to give their consent for the use of the videos for research purposes: if students were unduly nervous about being videoed one would have expected few to opt in. In fact 97 out of the 99 students agreed to their use.

Figure 23.2 shows the mean (with standard error bars) of the students' scores for each statement in the AEQ. They were asked to rate each statement as being more accurate of weekly sheets or more accurate of the one-to-one tutorials (on a five

point Likert scale from +2 to -2). The AEQ proved highly reliable ($\alpha = 0.81$) and while space precludes a full analysis of the results, there were some interesting and surprising findings worth reporting.

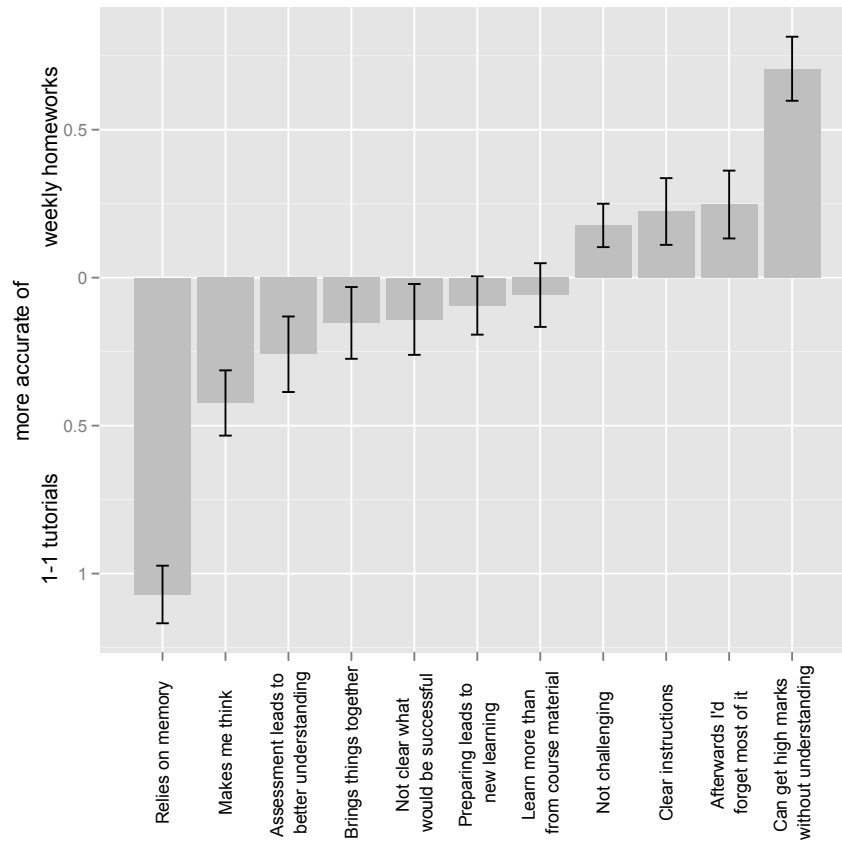


Fig. 23.2 AEQ Results

The students saw the one-to-one tutorial as making them think significantly more about the material ($t(84) = 3.84, p < 0.001$) and helping them understand things significantly better ($t(84) = 2.02, p < 0.05$) than weekly example sheets. They felt that with the weekly example sheets it was easier to get away with not understanding and still get high marks ($t(84) = 6.52, p < 0.001$), that they were not as challenging ($t(84) = 2.41, p < 0.05$) and they were significantly more likely to forget the material learned ($t(84) = 2.16, p < 0.05$) than one-to-one tutorials. However, they also thought that the one-to-one tutorial relied more on memory than example sheets ($t(84) = 11.00, p < 0.001$).

In addition to the comparative questions, students were given some free text to discuss their experience of the one-to-one tutorials and many took the opportunity to give substantial responses. While, again, space prevents a detailed analysis of these,

the general tone came across clearly through repeated comments. Many mentioned that they had prepared more than they would for an ordinary weekly homework and tutorial and felt that they needed to understand the material more deeply to be successful. However, many also mentioned that they found the situation more stressful and the pressure of the short timescale only added to this. Some commented on the need to rely on their memory more than for written homework, but many noted that there was value in individual working and feedback tailored to the gaps in understanding.

Many mentioned the issue of fairness and consistency, particularly in relation to the contingent questioning. Some felt that the tutors being able to tailor questions to their solutions was a benefit of the system, but others were concerned that not everyone got identical questions. Quite a lot of students noted both benefits and disadvantages and the general sense was that the oral performance assessment could have a role to play in a more mixed diet of methods, alongside (rather than replacing) written example sheets and exams.

In the interview with the course tutors, some of the advantages and drawbacks of this type of assessment were discussed. The tutors strongly agreed that performance assessment of this kind allows the assessor to find out 'what they [*the students*] do know rather than what they don't know' and allows the assessor to get to know the students individually and quickly understand whether they are struggling, or indeed whether they are coping well and on top of the material. A setting like the one-to-one tutorial does not allow a student to 'hide' amongst his/her peers and lets the assessor offer targeted help. Concerns were raised about accountability (although the assessors acknowledged that the marks could be moderated by using the videos) and resources. On the whole the tutors considered this as a positive experience, but felt that much more work went into this type of assessment than in the usual marking of weekly exercise sheets. However, this may have been because they were comparing the effort related to preparing and piloting a new form of assessment with a regular, well systematised form: the actual time spent in assessing was broadly similar to the time given to a week's tutorials and marking.

In conclusion, there were a number of concerns raised during implementation: many students reported levels of anxiety, though generally the students reported a positive experience. The workload and the results in terms of attendance and marks were broadly similar to the ordinary coursework/tutorial system, but the students and staff reported that the assessment led to improved understanding, was less likely to enable students to gain high marks without understanding and was more likely to engage the students in thinking about the material.

References

- Briggs, D. (1980) A Study of the influence of handwriting upon grades using examination scripts, *Educational Review*. 32(2), 186-193.

- Bryan, C. & Clegg, K. (2006) *Innovative Assessment in Higher Education*, Routledge.
- Gibbs, G. & Simpson, C. (2003) Measuring the response of students to assessment: the Assessment Experience Questionnaire. 11th *International Improving Student Learning Symposium*, Hinckley, England.
- Gold, B. (1999) *Assessment Practices in Undergraduate Mathematics*, Washington DC: Mathematical Association of America, 143-145.
- Henderson, J., Lloyd, P. & Scott, H. (2002) 'In the real world we're all put on the spot at some time or other, so you need to be prepared for it': An exploratory study of an oral method of assessing knowledge of mental health law. *Social Work Education*. 21(1), 91-103.
- Hounsell, D., N. Falchikov, J. Hounsell, M. Klampfleitner, M. Huxham, K. Thompson, & S. Blair (2007) *Innovative assessment across the disciplines: An analytical review of the literature*. York: Higher Education Academy.
- Huxham, M., Campbell, F. & Westwood, J. (2012) Oral versus written assessments: a test of student performance and attitudes, *Assessment & Evaluation in Higher Education*. 37(1), 125-136.
- Iannone, P. & Simpson, A. (2011) The summative assessment diet: how we assess in mathematics degrees. *Teaching Mathematics and its Applications*. 30(4), 186-196.
- Joughin, G. (1998) Dimensions of oral assessment. *Assessment & Evaluation in Higher Education*. 23(4), 367-78.
- Joughin, G. (2010) *A short guide to oral assessment*, Leeds Met Press.
- Levesley, J. (2011). Taking control of the assessment agenda. In Rowlett, P. (ed) *Report of the HE Mathematics Curriculum Summit*, York: Higher Education Academy, 21-23.
- London Mathematical Society (LMS). (2010). *Mathematics Degrees, their Teaching and Assessment*. Accessed via <http://futurelms.wordpress.com/2010/04/08/teaching-position-statement/>
- Marshall, G. & Jones, N. (2003) A pilot study into the anxiety induced by various assessment methods, *Radiography*. 9, 185-191.
- Nelson, M. (2011) Oral assessments: improving retention, grades and understanding, *PRIMUS*. 21(1), 47-61.
- Stray, C. (2001) The shift from Oral to Written Examination: Cambridge and Oxford 1700-1900. *Assessment in Education*. 8(1), 33-50.

Mapping University Mathematics Assessment Practices
Published 2012.
University of East Anglia
ISBN 978-1-870284-01-1

The Intellectual Property Rights (IPR) for the material contained within this document remains with its respective author(s).

This work is released under a Creative Commons Attribution-NoDerivs 2.0 UK: England & Wales Licence as a part of the National HE STEM Programme.



Photographs on the cover are reproduced courtesy of Durham University, and under Creative Commons license from pcgn7 and ILRI.