Oral Assessment in Mathematics: Implementation and Outcomes

P. Iannone* and A. Simpson

*a School of Education and Lifelong Learning, University of East Anglia, Norwich, United Kingdom

*b School of Education, Durham University, Durham, United Kingdom

* Corresponding author: Paola Iannone. School of Education and Lifelong Learning, University of East Anglia, Norwich, NR4 7TJ, UK. Phone: 0044 (0)1603 591007, Email: p.iannone@uea.ac.uk
Abstract: In this paper we report the planning and implementation of an oral assessment component in a first year pure mathematics module of a degree course in mathematics. Our aim was to examine potential barriers to using oral assessments, explore the advantages and disadvantages compared to existing common assessment methods and document the outcomes in terms of student views and performance. Our findings suggest that concerns about staff workload, students’ anxiety and fairness were only partially realised and that oral assessments may have a role to play in enriching the current assessment diet in mathematics.

Keywords: Assessment at university, oral assessment, mathematics, UK
1. INTRODUCTION

The experience of assessment of the majority of mathematics students in the UK seems to be based predominately on high stakes, closed book examinations at the end of almost every module. Iannone and Simpson (2011) noted that, across most of their sample, the minimum proportion of closed book examinations that mathematics undergraduates can take for credit towards their final degree mark is over 50% and, for universities in the top half of a public league table, this minimum proportion is between 80 and 90%. While noting that there is some variety, in the form of coursework, essays and projects (normally focussed on particular modules such as statistics or history of mathematics), they describe a very restricted assessment diet.

There is evidence, however, that mathematicians are concerned at the need to preserve the special status of their subject with respect to how it is taught and assessed (LMS, 2010). The prevalence of closed book examinations after decades of scholarly work advocating projects, presentations and other ‘innovations’ (Hirst and Biggs, 1968; Berry and Houston, 1995; Haines and Crouch, 2005) suggests that mathematicians see the closed book examination as the ‘gold standard’ of assessment.

In any case, the use of oral forms of assessment in mathematics (apart from the ubiquitous doctoral viva voce) seems entirely absent in the UK.

This was not always the case, nor is this situation universal. Stray (2001) noted that the oral examination was standard practice in the UK until the beginning of the 20th century, falling from favour because of accusations of bias and the apparent efficiency of written exams. However, many mainland European nations such as Hungary, Italy and the Czech Republic maintain an oral assessment component as an important (or
even a majority) part of their assessment diet. While the US appears to also be
dominated by closed book examinations, there is evidence of some use of oral
assessment (Gold, 1999; Nelson, 2011).

There is only a small research base on which to ground consideration of the use of
oral assessments. Hounsell et al. (2007) in a comprehensive review of the literature,
found less than 2% of the papers they surveyed address the form of assessment we
consider here. They noted that much was “frustratingly vague, omitting basic details
about sample sizes and populations … did not attempt any formal evaluations of the
efficacy of the innovations described; only one contained a quantitative evaluation,
with two more giving qualitative data” (p34) and suggest that this area was much in
need of further research.

Among the more scholarly work, Joughin (2010) distinguishes between three types of
oral assessment: presentation, application (such as trainee doctors undertaking live
diagnoses with an actor-patient) and interrogation (covering everything from short-
form question-and-answer to the viva). In this paper, we will be concerned with the
latter form, for which we will use the less pejorative term “oral performance
assessment”. In our case “oral performance assessment” is assessment where the
student is working on a problem with a tutor, while able to both write and converse
about their attempts at a solution and the mathematics underpinning it.

Clearly there are a number of obvious concerns which can be raised about oral
assessment. The two most common ones relate to anxiety and fairness.

There is a perception that an oral assessment may make students more anxious than
other forms of assessment. Henderson, Lloyd and Scott (2002) looked at the use of
oral assessment in social work and found that, before taking them, students did indeed
harbour great anxiety about the assessment. However, they noted that, after
experiencing them, students held much more positive views and, by the time they had
entered professional practice, they recognised their importance in terms of
authenticity and value to later employment. Huxham, Campbell and Westwood
(2012) noted a similar balance between anxiety and value/authenticity in research
with biology undergraduates and Hounsell et al. (2007) question whether the anxiety
is inherent in the orality of the assessment or simply because the assessment is novel.

There is little empirical research which examines the comparative anxiety raised by
different assessment methods, though Marshall and Jones (2003) showed that while
oral clinical examinations led to more anxiety amongst medical students than written
examinations, assessed presentations in a seminar setting caused still higher levels of
anxiety.

Concerns about fairness can dominate discussions about oral assessment. There are
obvious concerns about bias (whether conscious or unconscious) because the nature
of most forms of performance assessment (whether music performance, driving test or
language oral) means they cannot be anonymised. It can be argued, however, that the
removal of bias from any assessment method is difficult: even handwriting style in an
exam can lead to bias (Briggs, 1980). Thus, the key concerns which need to be
addressed are monitoring and moderation. Since, by their nature, oral assessment need
not involve a permanent written record, traditional forms of checking and second
marking will not be feasible. However, the availability of cheap video recorders does
allow for alternative ways of monitoring the process and moderating the marks.

In contrast, oral assessments might be considered fairer than some forms of
assessment in other ways. Joughin (1998) notes that oral assessment makes
plagiarising very difficult and, unlike examinations, oral assessment with a tutor prevents one small gap in knowledge completely stalling a solution.

Levesley (2011) highlights many current concerns with assessment in mathematics, including mathematicians’ sense of ownership of the choice and mix of assessment methods and risk aversion when it comes to seeking alternatives. In particular, he called for an exploration of the role for oral assessment.

As part of a larger project on the assessment of undergraduate mathematics, we examined the issues relating to the implementation of oral assessment in one undergraduate module. In particular we were interested in two issues:

- What are the practicalities of implementing a pilot oral assessment?
- What are the outcomes in terms of performance and students’ attitudes to and views of the experience of oral assessment?

2. THE PRACTICALITIES

The first issues we faced in developing the project were real and imagined institutional obstacles. In wanting to trial a very small-scale experiment in using oral assessment, some of the people involved were unsure of where the institutional authority lay to allow a variation to the standard assessment procedure. This might have been the result of trying to implement the project in a short time frame, but there was a sense that institutional inertia might lead to delays and acted to suppress change (as suggested by Bryan and Clegg, 2006). Levesley (2011) states the “main obstacle [to changes in assessment] is conservatism inside the mathematics sector, and risk averseness in the quality management function of universities, who are anxious about QAA inspections” (pg. 21). However, the reality within our study was that we found a
lack of understanding of the institution’s risk appetite for trialling alternatives and the procedures that needed to be followed, rather than a real risk aversion. Indeed, the appropriate authorities within the mathematics department were very supportive and seemed genuinely interested in what we were trying to do and the appropriate university-level committee chairs were happy to allow the change outside of normal change routes, again expressing interest in the process and outcome.

It was agreed that the assessment could replace one piece of coursework on a first year module on graph theory. The module covers basic notions of graphs, Euler’s formula, paths and cycles, colouring, Kruskal’s algorithm etc. The module is not a core one, but is taken by the large majority of mathematics students, with 108 people registered to take it.

The course is delivered through 20 lectures and 10 tutorials. The tutorials accompany a piece of coursework consisting of 2 to 4 problems handed out in given week, with students handing in written solutions the following week and getting their marks back and engaging with discussion about difficulties during the tutorial in the week after that. The coursework accounts for 10% of the final mark of the module (the remainder coming from a written examination at the end of the academic year). The tutorials take place in eight small groups (with 12-15 students in each group), with marking taking around an hour per group and the tutorial taking an hour per group. Thus, the total staff time allocated to a standard piece of written coursework assessment and its associated tutorial for one week is 16 hours.

We worked with the course lecturer and the two course tutors (doctoral students paid to mark coursework and give the tutorials) to implement the oral assessment in a week towards the end of the module, when the lectures were covering planarity and
trees. The lecturer assigned four questions (shown in figure 1) a week before the oral assessment. In that week, a short introduction to the process of the oral assessment was given verbally to the students in a lecture and more detail about the timing and organisation was sent in an e-mail. In these and in our other communications with students, mindful of the issue of anxiety, we called the oral assessment a “one-to-one tutorial”. Students were encouraged to ask questions or raise issues. In fact only one student raised issues about nerves related not to the one-to-one tutorial or its assessment as such, but to the fact that the tutorials would be videoed.

[Fig 1 goes here]

Each student was required to attend a 10 minute session, described to them as follows:

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!)

The students were also told that the format of the one-to-one tutorial would be that they would discuss two questions, one of which they could choose and the second of which would be chosen by the tutor. Since questions A and B require a proof and questions C and D require the implementation of an algorithm with some supporting reasoning, the assessors decided that when the student chose a question from one pair,
the assessor would choose a second question (using some random method, such as tossing a coin or drawing lots) from the other pair.

The one-to-one tutorials were conducted by the lecturer, one of the usual postgraduate tutors and the authors of this paper. The assessments took place in the time slot reserved for the usual tutorials for this module. At any one time, four assessors were needed for the tutorials, but given that there was no marking of written work, the total staff time spent in conducting the assessment was 18 hours (instead of the 16 hours of marking and tutorials in other weeks).

One of the advantages of oral assessment noted by Joughin (1998) is that the interaction allows for dialogue that is genuinely individual, though this also “gives assessment an inherent unpredictability in which neither party knows in advance exactly what questions will be asked or what responses will be made” (pg. 371). Therefore, prior to the start of the assessment the tutors discussed how the sessions were to be organised and developed a set of contingent questions – areas of mathematics which they might discuss depending on the solution strategy taken by the student. For example, in question A, the contingent questions covered whether the students have a clear idea of the role played by the complete graph $K_n$ and can explain why it has $\frac{1}{2}n(n-1)$ edges; whether they have a sense of where the relationship $e \leq 3v - 6$ comes from; whether they understand where the proof breaks down for $v = 10$ and if that means that the statement is false for $v = 10$?

In the oral assessment sessions, students were given the opportunity to write notes and ideas on a blackboard or on paper, and to discuss their solution orally.

At the end of each session, the tutor was asked to award a mark. Unlike written coursework which is marked entirely on the correctness (or otherwise) of the solution,
oral assessments allow an evaluation of the extent to which a solution is understood by discussing it with the student. It also permits students who might have been unable to solve the problem because of a small gap in knowledge or a misunderstanding to be helped towards a solution and still gain credit for what they do know.

The mark was awarded according to an assessment matrix (figure 2), which was also shared with the students beforehand to give them an understanding of what criteria the tutors would use when awarding marks.

3. OUTCOMES

3.1 Marks and assessment

In terms of marks, the outcomes were broadly similar to those achieved in other weeks in the tutorial-and-coursework system. Figure 3 shows the mean marks (with standard error bars) for each of the four questions, the students’ total score and the marks across the previous weeks of the module (all adjusted to be out of 5). Clearly the comparability across weeks and questions is a crude measure, as performance will inevitably depend on the questions set each week.

In terms of the students’ choice, the proof-style questions A and B were more popular (with 46 and 31 students choosing them respectively) while the algorithmic questions C and D were chosen by a small minority (each being chosen by 11 students).

It is also worth noting that the attendance at the oral assessment was broadly in line with attendance at tutorials and submission rates for weekly homework. Only 9 of the
108 students failed to attend the one-to-one tutorial while, on average, 8.5 students failed to attend a weekly tutorial and 9.2 failed to submit their coursework in the other weeks of the module.

3.2 Students’ views

The week after the oral assessments, asked at the beginning of a lecture students were to complete a questionnaire about their experience of the one-to-one tutorials as an alternative to

a) the weekly coursework
b) a final examination.

It should be noted that the oral assessment described here had been a direct replacement for the weekly coursework, so their responses to that part of the questionnaire can be considered to be based in their immediate experience. However, at that stage, none of them had then taken an end-of-module examination at the university, so their responses were based on a hypothetical comparison; though all of the students would have considerable very successful experience of written examinations in mathematics prior to university.

The questionnaire was based on the Assessment Experience Questionnaire (AEQ) designed by Gibbs and Simpson (2003). It asked students to rate 11 statements (shown in figure 4) on a five point Likert scale as follows:

+2: Much more accurate of written weekly sheets than one-to-one tutorials

+1: More accurate of written weekly sheets than one-to-one tutorials

0: About the same for written weekly sheets and one-to-one tutorials
-1: More accurate of one-to-one tutorials than written weekly sheets

-2: Much more accurate of one-to-one tutorials than written weekly sheets.

With a second section of the questionnaire asking identical questions comparing the one-to-one tutorials with a final examination.

[Fig 4 goes here]

The questionnaire was completed by 85 students and proved to be highly reliable ($\alpha = 0.81$).

Figure 5 shows the results of the questionnaire for the comparison between the oral assessment and weekly homework. There is a clear sense that students see oral assessment as having more positive aspects than the weekly sheets. In terms of being significantly more positive (with $p < 0.05$), students see oral assessments as making them think about the material, encouraging them to understand things better, being challenging, being less likely to forget the material after the assessment and not allowing them to get high marks without understanding. While they do note that they the oral assessment relies more on memory than the weekly sheets, this is understandable given that they were not allowed to use notes during the oral assessment.

[Fig 5 goes here]

The second part of the questionnaire asked for a similar comparison between oral assessments and the final examination and here we see a quite different set of responses (figure 6). While slightly less clear cut, there is a notable tendency to see examinations in the more positive light. In terms of being significantly more positive (with $p < 0.05$) students see exams as having clearer instructions, making them think
more about the material, requiring them to learn more as they prepare, encouraging them to learn more than the course material and bringing things together more.

[Fig 6 goes here]

In addition to the questions asking them to compare assessment methods, the students were asked to fill in an open response box to comment on “your experience of the one-to-one tutorials and anything you would like to say about how you are, or would like to be, assessed in mathematics”. In total 73 students wrote some comments and these tended to be thoughtful and reasoned statements about the oral assessments.

The trends in the comments generally followed the numerical responses in being positive about oral assessment as an alternative to weekly coursework sheets, but negative about them as an alternative to examinations.

A number of students noted the extent to which oral assessments focussed on understanding:

“…having 1-1 guidance was much more helpful in getting me to understand the problem instead of just watching someone do the problem.” [questionnaire number 66]

“…I found it very useful having to explain my solution as it furthered my understanding and gave me the opportunity to discuss it rather than do a half-hearted homework that I only receive feedback from a fortnight later. It was a highly rewarding experience.” [53]

They also noted that the nature of the oral assessment ensured that they prepared more than for a normal tutorial:
“…I prepared a lot more than for the normal homeworks.” [49]

“1-1 tutorials felt more tailored to where I was personally and could move to new directions. However, I did prepare more for this than normal weekly tutorials. Discussing it one to one meant I understood the point of what the question was asking.” [57]

The students often noted the immediacy and value of feedback:

“I found it very useful because a lot of the learning comes from the homework and if you’re wrong a lot of the time (like me) then you just get a cross and no explanation as to why you were wrong.” [40]

“working individually on a written piece and getting it back marked often doesn’t reveal much.” [47]

However, many raised the issue of anxiety and this appears to have been exacerbated for some by the short time frame:

“I got quite anxious about the 1-1 tutorial, which was ok as a one-off, but if I had one on a weekly basis and if the assessment from it were to count towards my final mark, I think I would be a bit stressed about the whole thing.” [76]

“I found preparing for the 1-1 tutorial quite stressful and nerve-racking, but I probably learnt more than I would have in a normal tutorial. However, given that it was more daunting than usual, I don’t feel I was able to perform as well as usual.” [67]

“Some people (like me) do not like 1-1 things, get nervous and don’t respond as well. Most people are used to written exams.” [84]
“The 1-1 tutorials are helpful if there is time to go into detail. A 10 minute slot turns out to be very rushed and I didn’t gain all that much from it, as unless you reel out a model answer there is not time to take the opportunity to ask questions/clarify understanding.” [72]

The issue of whether the oral assessment might have been better had it not been tied to a grade came up from some students:

“I think 1-1 tutorials would be really good if they weren’t assessed as you would be able to talk to lecturers on a personal level. This way you can have a chance to ask questions and receive explanations in a way not possible in a bigger group. Also being assessed means you wouldn’t want to ask the questions in case you lose marks.” [27]

“…in a 1-1 tutorial which is being assessed you don’t want to ask questions which you don’t understand in case it effects [sic] the marks of the assessment.” [7]

Those who compared the oral assessment to the final examination in their comments were clear that they felt the final examination was a more appropriate form of major assessment, although this was not without some hedging:

“I think sticking to a written examination is best – a 1-1 tutorial is not a fair form of assessment for the end of the year.” [81]

“I think a conversational part to examination would be good – perhaps to help you prepare (before a written exam) and to develop interest and depth (go beyond topics covered in class) … In general a 3-hour exam seems to work quite well but some variety would be good.” [59]
This last theme - the role oral assessment might play in a richer assessment diet - came across in a large number of responses:

“I think it would be better to have a combination of both, similar to the … weighting of tutorials to final examinations. The risk with purely final exams is that one bad day can affect your mark quite significantly” [79]

“I really like the idea of 1-1 tutorials replacing some formative homeworks as you have the ability to ask questions as you go.” [15]

“I would feel dubious about solely being assessed via the 1-1 tutorial, although having a few which accounted for 10-30% of the course in addition to an examination would be welcomed” [35]

“I found the 1-1 tutorial fairly enjoyable and I think having them regularly would be beneficial. However, I also find group discussion in my regular tutorials very useful – perhaps alternating the two?” [13]

4. DISCUSSION

Huxham et al. (2010) note five main advantages to oral assessments: they develop oral communication skills, they are authentic (as they more accurately mimic job interviews or the way people defend ideas in verbal exchanges), they can be seen to be more inclusive, they can be a powerful way of evaluating understanding and they are more difficult to cheat in.

The students in our sample particularly focus the fourth of these issues, highlighting the ability of oral assessments to gauge understanding and the inability to get high marks in them simply from parroting answers. They do note concerns with them, however. The stress that this type of assessment can cause was noted by many in the
open comments, though as Hounsell et al. (2007) note “it is not clear whether oral assessments are scarier or just more novel” (p34).

When comparing oral assessment to two other forms of assessment – weekly coursework sheets and written final examinations – the students gave quite different opinions. They were much more positive about them in relation to coursework sheets which they feel lack immediate feedback. The students also felt that with weekly exercise sheets it was easy to gain marks with answers which were not really understood, and then forget about the material once the coursework was returned.

However, the students were much more conservative when it came to final examinations. They appear to see the closed book examination as the ‘gold standard’ of assessment (or “best form of assessment from all the forms of assessment possible” as one student put it). Of course, all the students in this sample are highly successful in systems dominated by written examination (they would have the equivalent of grade A or above in A-level mathematics). However, given the thoughtfulness of their written comments, that alone may not account for their response.

The key messages from the experience of implementing the oral assessment were that (apart from the resource involved in setting up any experimental form of assessment) the effort involved was broadly similar to a tutorial and coursework system and the performance was also broadly similar. The experience was generally positive leading many students to suggest it as a possible alternative or addition to the existing forms of assessment and they clearly see the potential value for gauging and promoting understanding of the material, while not replacing the final, high stakes, written exam.

Iannone and Simpson (2011) note the existing restricted diet of undergraduate assessment. It is also clear that both staff and students see the closed book
Oral assessment in mathematics examination as continuing to be the ‘gold standard’ for assessment in undergraduate mathematics. However, the experience documented here suggests that some use of oral assessment in less high stakes settings may be one way of enriching that assessment diet.

ACKNOWLEDGMENT

We would like to thank the MSOR Network in the UK for funding this project through the Mathematical Sciences HE Curriculum Innovation Project, part of the National HE STEM Programme.

REFERENCES


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


*PRIMUS*, 21(1), 47-61


**Notes on Contributors:**

**Paola Iannone** is lecturer in educational research in the School of Education and Lifelong Learning in the University of East Anglia. Her research interest is teaching and learning mathematics at university level, with particular focus on assessment and proof and proof production.

University of East Anglia

School of Education and Lifelong Learning

Norwich, NR4 7TJ

Tel: 01603 591007, p.iannone@uea.ac.uk

**Adrian Simpson** is reader in mathematics education in the School of Education at Durham University and is also the Principal of Josephine Butler College. His research has primarily focused on students’ thinking across the transition from school to university mathematics, the transition to independent graduate study and mathematical logic and rationality.

Durham University

School of Education
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 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 contradict 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 weight $w(ij) = i + j$;
(b) $K_5$ with vertices \{1, \cdots, 5\} and weight $w(ij) = 1 + j^2$ where $i < j$.

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

FIG. 1: Homework tasks
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. 2: Oral Assessment Matrix
FIG. 3: Average marks for each question, total and other weeks
Tackling the assessment really makes me think.
I learn more from doing the assessment than from studying the course material.
In completing the assessment you can get away with not understanding and still get high marks.
The assessment gives very clear instructions about what you are expected to do.
When I tackle an assessment it is not at all clear what would count as a successful answer.
The assessment is not very challenging.
Preparing for the assessment is mainly a matter of memorizing.
Doing the assessment brings things together for me.
I learn new things while preparing for the assessment.
I understand things better as a result of the assessment.
I'll probably forget most of what I learned for the assessment after the assessment is over.

FIG. 4: Assessment Experience Questionnaire Statements
FIG. 5: AEQ Results: oral assessment vs weekly sheets

- Relies on memory
- Makes me think
- Assessment leads to better understanding
- Brings things together
- Not clear what would be successful
- Preparing leads to new learning
- Learn more from course material
- Not challenging
- Clear instructions
- Afterwards I'd forget most of it
- Can get high marks without understanding
FIG. 6: AEQ Results: oral assessment vs exam