THE LONDON SCHOOL
OF ECONOMICS AND
POLITICAL SCIENCE

The impact of high stakes oral performance assessment on students' approaches to learning: a case study

lseresearchonline@lse.ac.uk
https://eprints.lse.ac.uk/

The impact of high stakes oral performance assessment on students' approaches to learning: a case study

Paola Iannone[1] · Christoph Czichowsky[2] · Johannes Ruf[2]

Abstract

This paper presents findings from a case study on the impact of high stakes oral performance assessment on third year mathematics students' approaches to learning (Entwistle and Ramsden, Understanding student learning, 1983). We choose oral performance assessment as this mode of assessment differs substantially from written exams for its dialogic nature and because variation of assessment methods is seen to be very important in an otherwise very uniform assessment diet. We found that students perceived the assessment to require conceptual understanding over memory and were more likely to employ revision strategies conducive to deep learning (akin to conceptual understanding) when preparing for the oral performance assessment than when preparing for a written exam. Moreover, they reported to have engaged and interacted in lectures more than they would have otherwise, another characteristic conducive to deep learning approaches. We conclude by suggesting some implications for the summative assessment of mathematics at university level.

Keywords Case study · Oral performance assessment · Students' approaches to learning · University mathematics

Mathematics is an academic discipline that employs very few summative assessment methods at university level and researchers (Steen, 2006) have often called for a variety of its summative assessment. Amongst the methods in use, written, timed exams dominate and indeed in a recent survey of assessment methods in England and Wales, Iannone and Simpson (2011) found the written, timed exam to be the modal assessment method across their sample. At the same time, the general education literature perceives uniformity of summative assessment as problematic (Brown, Bull, & Pendlebury, 2013) and indicates that traditional written timed exams alone are not suitable to assess all the competencies that may be desirable to assess (Birenbaum et al., 2006).

✉ Paola Iannone
  p.iannone@lboro.ac.uk

1   Mathematics Education Centre, Loughborough University, Loughborough, UK

2   Department of Mathematics, London School of Economics and Political Science, London, UK

In this paper, we report findings from a case study investigating the implementation of oral performance assessment, a dialogic form of assessment, in two modules offered to third year mathematics students at a research-intensive university in the UK. The motivation for the case study is twofold. On the one hand, we seek to introduce variety of assessment in university mathematics by including a dialogic assessment method. On the other hand, we wish to investigate whether this dialogic assessment method could bring changes in students' experiences and approaches to learning, for the module where it was implemented. To expect changes in students' approaches to learning following a change in summative assessment is plausible (Harlen & Deakin Crick, 2003; Scouller, 1998), although the factors that induce such changes are not clearly understood, as the study of Gijbels and Dochy (2006) shows. The case study of this paper aims to explore such factors, as well as other areas of impact of the assessment change, in the context of university mathematics.

In the following sections, we first introduce the framework of deep and surface approaches to learning. We will then review the relevant literature on summative assessment of university mathematics and the literature on oral assessment in university mathematics. Finally, we will introduce the research questions of the study and the findings from the data.

1 Approaches to learning and students' perceptions of assessment

We frame this study in the work of Entwistle and Ramsden (1983) and Marton and Säljö (1997) on the relation between approaches to learning and differences in outcomes of learning. Marton and Säljö (1997) theorise that the differences in outcomes of learning depend on the differences in the processes that students adopt to achieve such learning and find two distinct ways of engaging with the text (text intended in a wide sense as the object of learning). They call those approaches to learning. Following the work of Entwistle and Ramsden (1983), these approaches are called deep learning and surface learning. Entwistle and Ramsden (1983) characterise deep learning as learning with the intention of understanding, aimed at creating links between the material read and driven by intrinsic motivation. Surface learning, on the other hand, is learning for memorising, passively accepting ideas and information, driven by external motivation with an emphasis on assessment results.

Having defined these two ways of engaging with the process of learning, and assuming that higher education aims to lead students to engage with deep learning, Marton and Säljö (1997) set out to investigate the factors that impact on the approaches to learning. Through a series of empirical studies, they found that the keys to influence the approach to learning were the learners' interpretation of what is demanded for learning and the predictability of the tasks associated with learning. The interpretation of what is demanded for learning is akin to the clarity of the requirement of learning, for example, what questioning is valuable for learning. Such questioning could be of the type: what is the relation between this text and what I have read before? Is this argument logically sound? The authors also argue that predictability of questioning exerts high influence on approaches to learning: if students can predict the tasks that are associated with learning, they are likely to comply with the perceived requirements without engaging in the text in a deep away. In the case of mathematics, predictability could be interpreted as being able to predict the nature and format of the questions that will be included in an exam paper: both in terms of topics (e.g., there will be two questions on topic X and 1 question on topic Y) and for the sequencing and nature of questions (e.g., there will be one or two definition questions, one seen proof question and maybe one seen application question

followed by one unseen application question for one given topic). Because this framework was obtained through a series of empirical studies, some of which collected students' self-report on approaches to learning, Marton and Säljö (1997) also discuss the use of this method. They write that because there are few ways of investigating approaches to learning, asking students to describe how they approach a learning task is a good proxy for understanding how that learning task appears to them, and what they think that learning task requires. Indeed, self-reporting is often used in the higher education literature; studies in this area report how students' perceptions of the learning task and their engagement with the task are factors in the investigation of approaches to learning (see the review by Struyven, Dochy, & Janssens (2005) for examples of the use of self-report and approaches to learning).

While Marton and Säljö (1997) focus on the demands of learning and discuss ways to influence learning approaches, Entwistle and Entwistle (1991) make the direct link between these demands and the demands of summative assessment. The authors observe that the way students approach learning is influenced by what they perceive as a summative assessment method to demand to be successful. They conclude that the questioning included in an assessment task is bound to influence how the students approach learning. The framework of approaches to learning has been used to describe learning processes in higher education; the evidence in Struyven et al. (2005) indicates that the relationship between students' perceptions of assessment demands and approaches to learning is not straightforward: simply changing the summative assessment demands does not necessarily foster deep learning. Gijbels and Dochy (2006) found that introducing assessment which should have fostered deep learning facilitated instead surface learning. Their explanation concerns the lack of consideration, in previous studies, of contextual factors such as time required for preparing for the assessment and study load. They argue that the students perceived the workload associated with the assessment as too high and reverted to surface learning in a bid to make the most of the time they perceived they had. Results such as this indicate that while it may be easy to induce surface learning approaches by changing the assessment to something the students perceive to require memorisation, stimulating deep learning by changing assessment may prove to be more difficult.

If we apply this framework specifically to university mathematics, we can draw a parallel between relational understanding and deep learning, procedural understanding and surface learning, where procedural and relational understanding are taken in the sense of Skemp (1976). Many of the characteristics associated with relational understanding, such as being able to link mathematical concepts, knowing that something is true and why, are akin to the characteristics of deep learning while memorising and applying procedures without knowing why, a prominent characteristic of surface learning, is also prominent in Skemp's procedural understanding. In this paper, we adopt the approaches to learning framework to investigate the impact of a new method of assessment on students' engagement with learning.

2 Summative assessment of mathematics at university level

Summative assessment of university mathematics in the UK is dominated by closed book exams, specifically timed, written exams in which the student has no access to external materials, such as lecture notes or textbooks. Iannone and Simpson (2011) surveyed the assessment methods in England and Wales for university mathematics and found that the median contribution of assessment by closed book exams towards the final degree mark was 72% and that most departments have closed book exams accounting for more than 50% (when averaged across all their modules) of the final mark. They also found that summative

assessment methods other than the closed book exam in mathematics degrees are mostly linked to other disciplines usually taught within mathematics degrees. For example, open book exams—specifically timed written exams where the student has access to external materials such as lecture notes or textbooks—are often used for statistics or essays for history of mathematics and mathematics education. This is to say that although the authors found some variety of assessment methods in mathematics degrees, assessment methods other than the closed book exam were used predominantly for subjects other than mathematics.

Since this situation is likely to be similar in many other countries, as some of the literature reviewed below indicates (see Bergqvist, 2007, for Sweden and Mac an Bhaird, Nolan, O'Shea, & Pfeiffer, 2017, for Ireland), it is not surprising that much of the literature on summative assessment of mathematics at university discusses the types of reasoning assessed by closed book exam as they are currently designed. The analysis of types of reasoning elicited by current summative assessment is important for our study in the light of the link posed by Entwistle and Entwistle (1991) between perceived assessment demands and approaches to learning.

One example of the classification of reasoning associated with exam questions is Lithner's (2008) framework. Lithner (2008) divides the reasoning required by mathematical tasks into two categories: imitative and creative reasoning. In the imitative reasoning category, Lithner (2008) includes questions that can be solved by performing known algorithms and recalling solutions by memory as the solution is familiar to the student. The creative reasoning category includes questions whose solution involves a new sequence of actions (new to the student), or where the argument proposed is plausible and shows the correct use of the properties of the objects involved in such reasoning. Similar to the terms used in Skemp's (1976) framework, we can characterise imitative reasoning as surface learning and creative reasoning as deep learning. Lithner's (2008) classification of mathematical tasks has been used to analyse questions in university exams in mathematics. For example, using this classification, Bergqvist (2007) found, in the Swedish context, that it was possible to pass an introduction to calculus closed book exam by employing only imitative reasoning while the tasks involving creative reasoning consisted mostly in asking students to create specified examples. A similar study in Ireland (Mac an Bhaird et al., 2017) found that there were more creative reasoning tasks in the assessment of mathematics for mathematics students than in the assessment of mathematics for non-mathematics specialists (e.g., engineers) and that for the latter it was possible to achieve high marks in calculus exams by just employing imitative reasoning.

The implications of these findings and those of others (e.g., Boesen, Lithner, & Palm, 2010; Darlington, 2014) are problematic: if we assume, with Entwistle and Entwistle (1991), that students adapt their approach to learning to what they perceive as the assessment requirements, then the format of written exams as it is currently implemented is not encouraging students to adopt deep learning approaches. This is not to say that the written exam cannot be a good summative assessment of relational understanding, but that the literature reviewed above seems to indicate that currently this assessment is not designed in that way often. In the light of these findings, and of the uniformity of the assessment diet of mathematics, it is reasonable to investigate the potentiality of assessment methods other than the written, timed exam.

3 Oral performance assessment

There is very little literature on oral assessment in higher education, despite this method being widely used in many countries (see Kehm, 2001 for the German context, but this assessment is also widely used for example at Italian universities). Several types of distinct dialogic

assessment methods usually go under the label of oral assessment. Joughin (2010) distinguishes between three types:

- *presentation*, where the student has prepared a set of slides and presents something akin to a small project, often followed by a short questions and answers session;
- *application*, such as the assessment of law students performing a court case simulation;
- *interrogation*, covering everything from short-form question-and-answer to the viva voce used, in many countries, as the final assessment for doctoral students.

This paper is concerned with the latter type and given the negative connotation that the word interrogation may have in the English language, we re-name this type of assessment oral performance assessment in line with previous work (e.g., Iannone & Simpson, 2015). In mathematics, this assessment takes the form of a question-and-answer session with the student writing at the board (or using pen and paper) while answering the questions that the assessor poses. Such questions can be of a theoretical nature (state seen definitions, proofs or theorems, explain passages in proofs) or of a more applied one (working out examples or using algorithms appropriately). The structure of the oral performance assessment allows also for student-tailored contingent questions used by the assessor to probe students' understanding. Those questions are often about clarifying and explain what the student has said. This is to say that even if there is an existing list of questions for all students, the examiner will ask follow-up questions tailored to the student's response aimed at clarification of answers or encouraging the student to rethink their answer.

We focus on oral performance assessment both to explore the potential of an assessment method that is used yet under-researched (Iannone & Simpson, 2015) and because it is substantially different from the closed book exams due to its dialogic form. Moreover, some scholars have questioned the status of written exams as sole indicators of how much students know and can do. Schoultz, Säljö, and Wyndhamn (2001) observe that

> When analysing performance on written tests, the specific difficulties introduced by this particular communicative format are seldom recognised. It is as if writing in solitude in the context of a test is an unbiased indicator of what people know or understand. (p. 214)

The authors compared, in the school context, performance on written and oral test questions in science, where the oral test was an oral performance assessment. They found that in the oral context students reasoned and conveyed their understanding better than in the written context, which often is hampered by issues of the understanding of terms. They also argued that students were less able to engage in scientific discourse in a written exam, specifically without an interlocutor. Once students were allowed to interact in a dialogic manner, they were able to engage in scientific discourse. The study on students' proof production conducted by Stylianides (2019) echoes these findings. In the school context, Stylianides (2019) found that students were more likely to produce arguments that met the standard of proof when working in a dialogic manner with interlocutors rather than when working in a written form by themselves.

Relevant to the types of reasoning currently assessed in written form are studies, such as the one by Alcock and Weber (2010), that show how a student can reproduce a written piece of mathematics to the satisfaction of a reader (or an assessor) while at the same time declaring lack of understanding of the topic. Alcock and Weber (2010) report the case of Carla, a first year student, who was able to produce correct written answers to questions in real analysis but admitted lacking any sense of understanding of it. We argue that if the written outcome produced by a student like Carla was assessed in a written exam, it would have most likely

accrued high marks. In this case, would this high mark have given any indication of the student's relational understanding of the topic?

Few papers discuss the use of oral performance assessment in university mathematics. Some report on small implementations and outcomes of oral performance assessment (Iannone & Simpson, 2015; Odafe, 2006) while others investigate the views of mathematicians on summative assessment and oral performance assessment, in particular (Videnovic, 2017a, 2017b). Videnovic (2017a, 2017b) suggests that the mathematicians interviewed believe that written exams alone cannot give a complete indication regarding the relational understanding of mathematics and that oral performance assessment is a better means to assess relational understanding. Iannone and Simpson (2015) report the implementation of oral performance assessment as coursework in a year one Graph Theory module, assessed mostly (90%) by closed book exam. They conclude that oral performance assessment can be a suitable summative assessment for mathematics and can take the role of "assessment for learning" (Black, Harrison, Lee, Marshall, & Wiliam, 2004). Their implementation, however, was not high stakes: the oral performance assessment replaced a small part of the final module mark that did not contribute towards the students' final degree mark. Therefore, we have no indication of the impact on students' approaches to learning that the oral performance assessment may have if it is the only high stakes summative assessment for a module.

Hence, for our study, we pose the following research questions:

RQ1: Did the oral performance assessment impact on students' approaches to learning in the two modules in which it was implemented? What were the factors that contributed to or prevented such impact?

RQ2: Were there any other consequences of the introduction of such assessment?

4 Methods

This project follows the design of an instrumental case study (Stake, 2000) with a student centred-focus. This design was chosen for the uniqueness of the case (Yin, 2011), and the focus on the oral performance assessment is motivated by the potential of this assessment method, especially for mathematics, to motivate students to adopt a relational understanding approach to the subject. The case is the introduction of the oral performance assessment as high stakes summative assessment in two optional modules in the third year of study in probability and financial mathematics in one high profile university in the south of England.[1] The modules, MA and MB, were taught by the second and third authors of this paper. The requirements for entering this degree course are among the highest in the UK; the second and third year of studies contribute the most to the final degree classification. The oral performance assessment was the only summative assessment for each of the two modules and it was held in May at the end of the academic year while the modules were taught in the Autumn (MA) and

---

[1] University degrees in the UK are typically modularised. To progress from one year to the next, each student needs to accrue a certain number of credits by sitting appropriate modules. Each module has between 3 and 4 contact hours per week, depending on the stage at which it is offered, and can be compulsory or optional. MA and MB are two optional modules offered to students in their third and final year of studies. It is also common for UK universities to have only one exam session in the summer and one remedial exam session (for students who failed exams in the summer) in September. This is why both MA and MB were assessed in the summer session (May). In the UK, it is also common to talk about degree classifications as well as final marks. The degree classifications for UK students are First Class (100–70), Upper Second Class (69–60), Lower Second Class (59–50), Third class (49–40) and Fail (39–0).

Spring (MB) respectively. In previous years, both modules had been assessed only by a closed book exam. Each module had three contact hours per week over ten weeks, two hours of lectures and one hour of seminar, which is the norm for third year modules in mathematics in the UK. There were 19 students enrolled (five female students) in MA and 11 in MB (three female students). All students who took MB also took MA. At the point of choosing the modules (end of the second year), the students were told that the only summative assessment for these modules would be the oral performance assessment. It was necessary to make sure the students realised that this assessment method would be used as they would have not experienced it anywhere else in their degree course. Detailed information about the assessment was given to the students at their module choice meeting at the end of the second year.

To familiarise the students with this type of assessment, a formal mock oral assessment was held halfway through the module MA (see Appendix 1 for a sample of questions on this mock assessment). This formative assessment was held behind closed doors to replicate the exam setting as closely as possible. For MB, the small size of the class allowed for one mock exam to be held each week, one for each student, during the seminar sessions so that all students could see the way in which the exam would be carried out. The mock assessment was formative only, students were not required to take part, but all took advantage of the opportunity. After the mock exam for MA, the students were invited to an interview and 5 students volunteered. The students were assured that their lecturers would not see the interviews before they had graduated and left the university to reassure them that their views would not impact on their grades in any way. The interviews lasted

Table 1 Information regarding the data collected for the case study

Data	When	Purpose
Syllabus and other contextual material for the modules (description of the assessment to the students, relevant emails, etc.)	September	Background data
Questionnaire	First class (MA - October)	Investigation of students' perceptions and preferences for summative assessment prior to the start of the module
Observations of four mock oral exams for MA – field notes	November	Background data
Interviews with five students after the mock exam	November	Investigation of students' experience of the mock exam and their expectations of the oral performance assessment
Interviews with seven students after the oral performance assessments for MA and MB	May	Investigation of students' experience of oral performance assessments
Interview with the External Examiner[a] of the degree course	June	Investigation of issues of quality assurance and feasibility of the oral performance assessment
Four videos of four oral performance assessments for MA	June	Background data
Some data about the previous cohorts for MA and MB (number of students, distribution of marks)	June	Part of the background data for comparing uptake for the module and marks

[a] Universities in the UK use an external peer review model for quality assurance of their degree courses. One or more academics from another university are assigned as External Examiners to each degree course. Their role is to review the syllabus and each item of assessment for each module, to review a sample of the coursework and marking and to advice in matters of curriculum or assessment change. They participate at the final exam board and produce a document (the External Examiner Report) that comments on all aspects of the degree course

between 15 and 35 min, were audio-recorded and fully transcribed. The interviews were semi-structured (Flick, 2018). A list of questions was devised before the interviews and used for all participants, but the interviewer also followed the flow that the conversation took in their questioning. Immediately after the oral performance assessment for both modules, students were invited again to be interviewed. This time seven students replied; the five students interviewed after the mock exam and two more. These interviews lasted between 10 and 40 min and were audio recorded and fully transcribed. All but one of these interviews were collected via video conference.

During the first class of the module MA, the first author of this paper introduced the study and asked students to fill in a questionnaire. This questionnaire comprised the Assessment Preferences Inventory (API, Iannone & Simpson, 2015) and the Assessment Experience Questionnaire (AEQ, Gibbs & Simpson, 2003), which compared students' perceptions of closed book exams and oral performance assessment. The External Examiner was also interviewed, after the main exam board for the degree course. This interview followed a semi-structured format, was collected via video conference and lasted 40 min. Ethical approval for the study was sought and obtained in the institution where the first author works. Information regarding all data collected can be found in Table 1.

4.1 Data analysis

4.1.1 The questionnaire

A questionnaire was administered to the students during the first class of the module MA. It comprised the API (Iannone & Simpson, 2015) and the AEQ (Gibbs & Simpson, 2003) comparing the closed book exam to the oral performance assessment on 11 statements on a five-point Likert scale. The text of the questionnaire is included in Appendix 2. The questionnaire was preceded by a short biographical part and was completed anonymously. The questionnaire aimed to ascertain the preferences and perceptions of summative assessment of this cohort to then be compared to those reported elsewhere in the literature (Iannone & Simpson, 2015).

4.1.2 The interview data

The data analysis started during the data collection phase and was carried out by the first author of the paper. An early analysis was deemed necessary to inform subsequent data collection phases and the structure of the interviews held after the oral performance assessment. The interview schedule for the second round of interviews (included in Appendix 3 together with the one for the first round) was designed following a preliminary analysis of the interviews in the first round. Analytical memos were produced for each interview. After the analytical memos were written, the interview transcripts were coded using descriptive codes (Saldaña, 2015). The term descriptive indicates codes that summarise the content of a part of the transcript. These codes (Saldaña, 2015) aim to construct meanings relevant to the central issue of

the case study: the impact of the introduction of the oral performance assessment on students' approaches to learning and experiences of the modules. A code list was kept throughout the first coding and the final list of codes with descriptions of meanings was obtained. The first author of this paper, who is not an academic at the university where the case study took place, hence an outsider for the participants, coded the interviews through two iterations of coding phases. The codes obtained in the first phase were grouped in umbrella themes and organised in a thematic network (Attride-Stirling, 2001), reported in Fig. 3. Once the thematic network was devised, the interview extracts grouped in each theme were coded again for nuances of meaning and explanation of rules and mechanisms via the use of pattern coding (Saldaña, 2015) to "understand the how" of what had been found (Saldaña, 2015, p. 154) in the thematic network. Other data sources were analysed separately and used as contextual data.

5 Findings

5.1 Context: the students' questionnaire

The whole of the MA cohort ($N = 19$) returned the questionnaire. The average age of the students was 20.5 years and the marks (out of 100) at the end of the first year of their degree were between 55 and 90, with 12 students in the 70–100 bracket confirming that this was a very high achieving cohort of students. Figure 1 shows
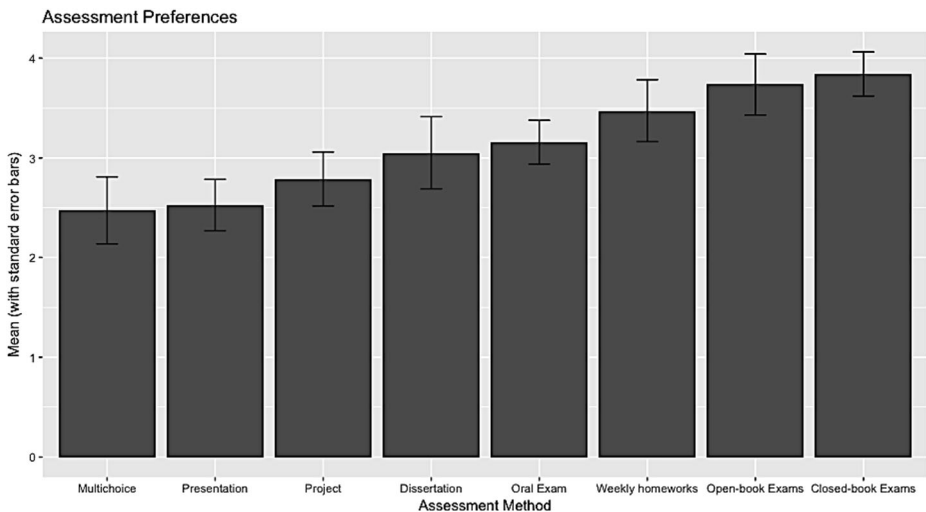

Fig. 1 Mean response (on a five-point Likert scale) for preference for each of the eight assessment methods in the API (with standard error bars) for the 19 students in the sample
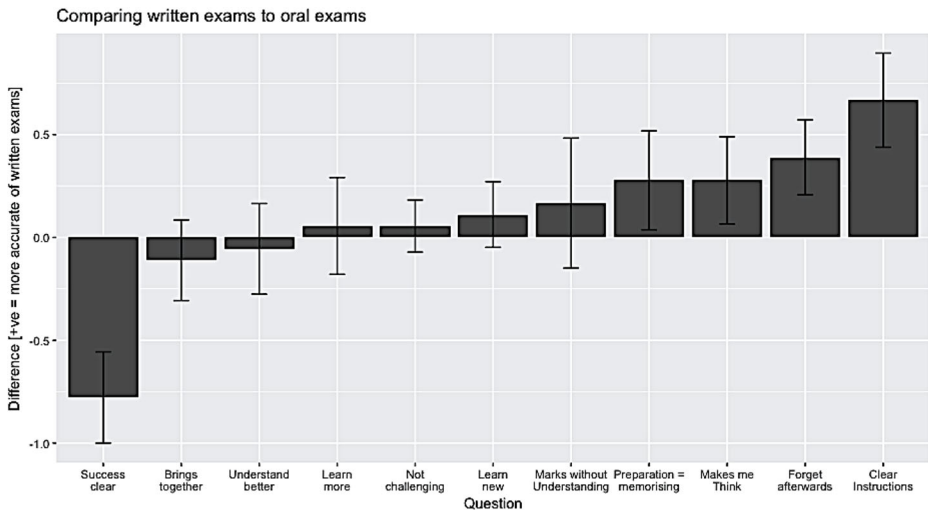
Fig. 2 Mean responses (on a five-point Likert scale, from + 2 (more accurate for written exams) to − 2 [more accurate for oral exams]) for 11 statements in the AEQ (with standard error bars) for the 19 students in the sample

the students' preferences for summative assessment. Similar to what is reported in Iannone and Simpson (2015), the students preferred closed and open book exams, which in the context of the very uniform assessment diet of university mathematics (Iannone & Simpson, 2011) would be the most familiar to them. However, as also found in Iannone and Simpson (2015), the oral performance assessment is listed as the fourth preferred assessment method, more than projects or dissertations, and multiple choice is the least preferred method.

Figure 2 shows the comparison between closed book exam and oral performance assessment. Whereas our students saw closed book exams as coming with clearer instructions and clearer indications on how to succeed, they also saw them as being associated with memory, and they saw themselves as more inclined to forget the material covered by the closed book exam than for the oral performance assessment. Note that the views of these students agreed with the views of the students in Iannone and Simpson (2015): the cohort involved in that study consisted of first year students, while the participants in the current study were at the end of their degrees. Yet, their perceptions of summative assessment seemed similar. We hypothesise that those views seem unchanged perhaps because of the uniformity of assessment experiences.

5.2 The interviews with the students

After the interviews were coded, two overarching themes emerged: process and impact. Figure 3 describes the resulting thematic network with umbrella themes and subthemes within those. Here, we discuss only the impact of the oral performance assessment, because in the process theme we found codes related to this specific implementation. These might not be relevant to other implementations under different circumstances.
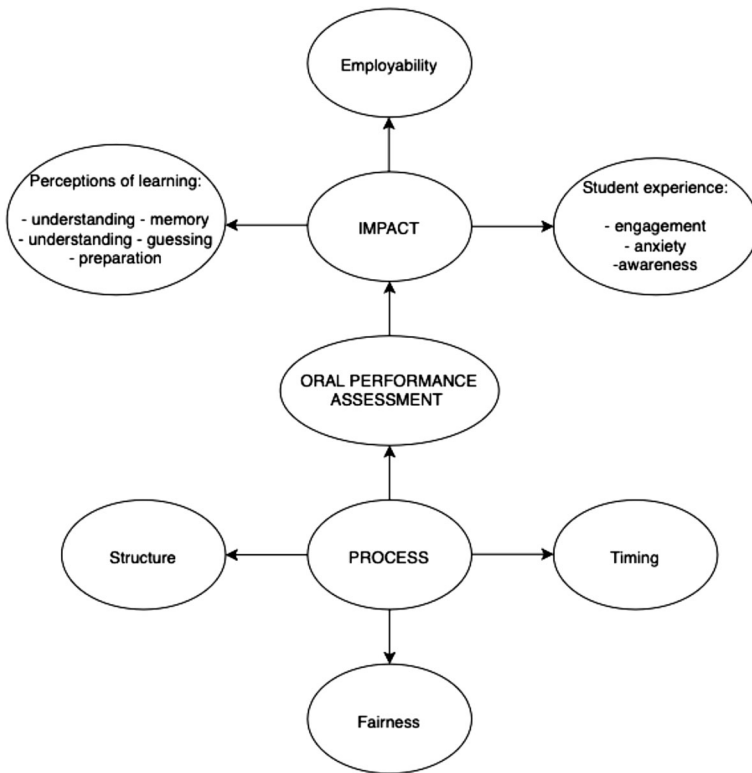
Fig. 3 Thematic network of codes emerging from all interviews, with the intervention (oral performance assessment) and the two overarching themes indicated by capital letters

5.3 Impact

The students talked at great length of the impact of the oral performance assessment on aspects of their learning experience. Most students report that the presence of an unusual mode of summative assessment did not deter them from selecting the modules, which were not compulsory. Comparison of the enrolment data for the two modules across three years shows that the cohort sizes are similar across this period, indicating that any impact on the uptake of the oral performance assessment may have been small (see also Figs. 4 and 5). This comparison does not explain whether the profile of the students taking these modules is different from previous years. It could be that students who took the module were higher achievers than those
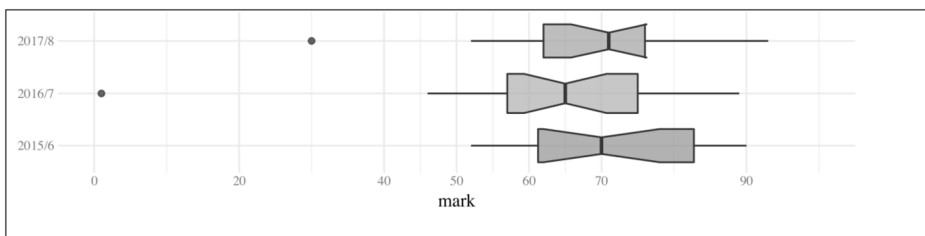

Fig. 4 Mark distribution for MA in AY 15/16; 16/17; and 17/18 (the year the study took place). The cohort sizes are 18, 25 and 19 respectively
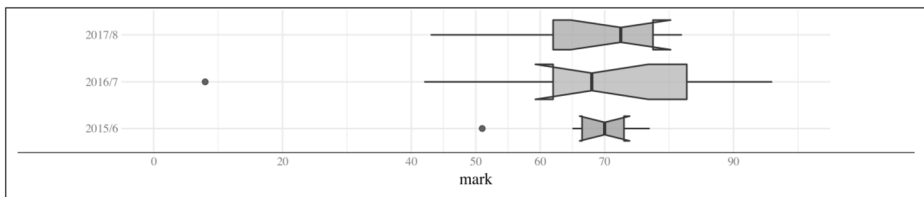
Fig. 5 Mark distribution for MB in AY 15/16; 16/17; and 17/18 (the year the study took place). The cohort sizes are 7, 14 and 11 respectively

in previous years, although the questionnaire shows a spread of marks similar to previous cohorts. From the thematic analysis of the interviews, we can find three areas of impact: on students' perceptions of learning, on student experience and on employability skills. In the remainder of this section, we will discuss these three areas in turn.

5.4 Impact on students' perceptions of learning

5.4.1 Understanding versus memorising, understanding versus guessing

The students juxtapose relational understanding to memorising when speaking about learning mathematics. The presence of the oral performance assessment was perceived to foster the need for relational understanding:

> But in the oral exam, it's less about tricks and more about probing your understanding deeper and deeper and deeper. (Rob, post[2])

While often closed book exams are seen to be only about understanding the format of the questions and being able to replicate some of the answers, oral performance assessment is perceived to be mostly about probing understanding further. Moreover, the students thought that it is not possible to achieve high marks in the oral performance assessment without this deeper understanding; on the other hand, for a written exam:

> Sometimes in written exams, you guess a method and you get it right and it works, and you get the marks. In an oral exam, it's very difficult to get away with that kind of solution where you just guessed something, because they [the examiners] will ask questions, and they will be able to tell that you were just lucky, and you didn't really understand the material. (Alvyn, pre)

This indicates that students may try to adopt a deep learning approach for this module, supported by the appropriate questioning if they expect that the oral performance assessment will require deep learning to succeed. One student even suggests that the oral performance assessment may be akin to the process of doing mathematics collaboratively either with a peer or with the lecturer. These results are interesting and confirm previous research findings that also indicate that the oral performance assessment was perceived by students to require understanding to gain high marks (Iannone & Simpson, 2015). Of more interest to our research questions, however, are the findings regarding the impact of the oral performance assessment on the approaches to learning the students employed.

---

[2] Transcript convention: the participants' name have been changed to maintain confidentiality; pre after the participant's name indicates that the extract is from the first round of interviews while post that it is from the second round of interviews. An ellipse such as […] in the interview transcript indicates some omitted text while three dots … indicate a pause in the speech.

5.5 Preparation for the assessment

All students reflected on how they prepared for the oral performance assessment. Some reported to have prepared in the same way as for closed book exams indicating that the lack of understanding of the requirements of this task prevented them from being able to prepare appropriately for the new assessment. These students assumed that the questioning associated with oral performance assessment would be predictable and it would be akin to the questioning in written exams in previous years. Indeed, all reported that the usual way to prepare for written exams consists of accessing past exam papers, solving the questions and then marking their solutions against those provided by the lecturer and those of their peers. This way to prepare for written exams assumes that the format of the exam remains similar across the years, making this assessment, as it is currently implemented, predictable in the sense of Marton and Säljö (1997). This predictability may motivate students' perceptions that to perform well in a closed book exam memorising is enough:

> I think it is a good idea to do oral exams in mathematics, because otherwise it's really easy, especially for these courses that are more theoretical, to just memorise certain things and not understand anything and still […] do well during the exam. (Joey, post)

Students who prepared by solving questions from past exam papers realised after the oral performance assessment that this method did not help them:

> Right, so that's also one of my problems with oral exams. So I didn't know how to prepare. So, the way I prepared for it is just … I thought just doing more exercises would make me more prepared so I just did the normal stuff. I looked at the past paper exams and I solved all of them […]. So, I learned all I could know, made sure I memorise all the definitions, all the major theorems, and how the implication will go one way, I had good examples. […] So basically, the way I practiced, I prepared for paper exams, but then I realised it's not that helpful for oral exams. (Ralph, post)

For Ralph, the lack of understanding of the demands posed by this new assessment disrupted his preparation. When he was asked whether he worked with peers towards preparing for the oral performance assessment he replied:

> So because some of the past exams didn't have solutions, so yeah, so I worked with my friend just to … we didn't do like our own mock oral exam, like we tried a bit, but we felt like that's just silly. Like, we're not going to know what will actually be on the exam, so we just did the normal paper exams. (Ralph, post)

Ralph's experience shows how some students found it difficult to think about summative assessment differently than they have been used to. For an oral performance assessment, it is less relevant than for a closed book exam to think in terms of "what will actually be in the exam," as the questions and especially the contingent questions are tailored to the student performance, and there are few, if any, questions that will be asked of all students. Other students reported that they had prepared differently for the oral performance assessment, like Alvyn:

> I think for this one [the oral performance assessment], well what I did actually was practice like asking my friends like, questions like, asking for the proofs, and what we can really come up with difficult questions ourselves … What helped was like practicing orally and reading all. I think understanding the proofs is much more important because you will be asked to justify a lot. (Alvyn, post)

This is echoed by Joey's experience:

> For the oral exam, the way I tried to prepare for it was to try to know the core fundamentals really well so that I would be able to explain what I was doing. (Joey, post)

The interviews also hint at the reasons for these different strategies: some students thought that the format of the oral performance assessment would be similar to that of the closed book exam and revised accordingly (displaying a mistaken sense of predictability), while others revised by trying to engage in deep learning and questioning each other. Recall that predictability, used in the sense of Marton and Säljö (1997), refers to the possibility of predicting the kind of questioning that is associated with the learning task. In contrast, from our data, we may also define the concept of familiarity. Familiarity refers to understanding how the oral assessment is structured, the role of the contingent questions, the perceptions that there will be questions that will ask to explain statements made during the oral performance assessment, and that these questions will not be the same for all students. The students seemed to indicate that it was this familiarity that helped them with the appropriate preparation while (a mistaken sense of) predictability led them to prepare in ways that were not helpful for the assessment. Indeed, familiarity was often obtained from the experience of the mock exam:

> The other one [the mock for MB] is more casual but then you get to see it more often, you get to see other people doing it, asking what they are getting wrong like maybe they're not stating assumptions properly, so you're going to observe and ready yourself which I think is quite good, and the mock, its good because you get to see how it is and being with the lecturer asking questions or yeah, you done it too I guess, observe, learn from the people. (Alvyn, post)

Alvyn, as well as other students, indicated that the format of the mock assessment may be of importance in creating familiarity with the assessment method. Summarising, the introduction of the oral performance assessment changed how some students prepared for the assessment. Most students tried to prioritise relational understanding over memory, but whereas some students were able to change the way they prepared and thought about assessment requirement in a different way than they were used to, others were unable to make this change, only to find out after the oral performance assessment that their preparation had not been as useful as they had hoped. Issues of familiarity and mistaken sense of predictability are of relevance to students' ability to change preparation patterns and the format of the mock oral assessment seems to have had a big impact on making the students feel familiar with the assessment method.

5.6 Impact on students' experience

5.6.1 Engagement

Engagement was referred to in the interviews both as engagement with the teaching in class and with the subject. The students reported that the oral performance assessment motivated them to engage in class more:

> Honestly, I prefer the oral exams generally. I could've sat MA and MB as written exams and I would've been fine, but I think particularly for MB, it made the course more interesting and engaging because during the classes, you're forced to kind of participate

and engage and during the lectures, you felt that you were forced to participate and engage. (Tony, post)

This way of engaging during lectures was also perceived as more interesting by some students and may have become part of their motivation to engage with the subject:

I do feel like you get more engaged, like first of all, it kind of motivates you to go to class a bit more just because you want- and when you're in class, you want to participate because you know like in the end, you're actually going to have to talk about it … (Alvyn, post)

Note that this engagement that seemed to be motivated first by external factors—namely the wish to understand the requirements and types of contingent questioning of the new assessment method—may become engagement motivated by enjoyment, which is what the quote by Tony above seems to indicate by the use of the word "interesting". Others noticed that engagement in class was necessary to be successful in the assessment and reflected on this after experiencing the oral performance assessment:

I mean I don't know how I would have prepared better for the oral exam I had. I would advise probably talking a lot to the lecturer, making sure that you're very … that you know very clearly what the lecturer likes to assess either by going to office hours or by interacting a lot in lectures and class ... just have a deep understanding of the material … (Joey, post)

The last part of Joey's quote may hint at the second type of engagement: engagement with the mathematics at a deep level, which is also most likely what the lecturer would like to convey to their students and would like to assess. Joey's reflections suggest that the two types of engagement should go hand in hand; engagement with the teaching of the module leading to and being motivated by engagement with the mathematics. Joey's suggestion of the need to engage comes from his perceptions about the requirement of the assessment. The students' perceptions of this type of assessment may indicate that they are likely to engage:

So, for making sure that the students [taking any module] engaged deeply into the content that you present them, I think oral exams are a lot better [than closed book exams]. (Rob, post)

5.6.2 Assessment anxiety and awareness of assessment requirements and process

Issues of anxiety and awareness of assessment requirements and processes are intertwined throughout the students' interviews. Not surprisingly, all students commented on high stakes assessment anxiety. Anxiety in oral performance assessment is linked both to the presence of an assessor and the lack of awareness of the method's process and requirements. Many students, not accustomed to facing their assessor during an assessment task, perceived this situation as stressful:

So firstly, it's because in an oral exam setting, when you are stuck at one problem, you're under the constant pressure of your professor looking at you. I mean kind of trying to nudge you, of course in a good way, but then it just makes me a bit nervous and more difficult to actually think. (Ralph, post)

In an oral assessment, some students reported, there is no place to hide:

… so yeah, just looking and seeing that there is an examiner watching you while you're doing it is a bit more stressful than a written exam where you're in a room filled with 80 other students. (Alvyn, post)

This sense of anxiety was strong, and attempts by the assessor to ease the tension may be acknowledged by the students but may be perceived to be ineffective:

Yeah, he [the examiner] was making it very comfortable, he said sure and I could sort out the right stuff on paper, and the thing about me, myself, is that I am always unsure of stuff, so like … But I was … I was freezed on that a little bit. (Marshall, post)

However, some recognised that being used to the examiner from engagement in class would help in this situation. Some students discussed the role of the mock assessment in easing their anxiety. They reported that the most effective mock assessment was the one which was informal and public, where the students could become accustomed to the style of contingent questioning and the process of the oral performance assessment through the year:

I think it made the class (MB) a bit more informal and it meant that by the time we get to the final exam, you've seen it done so many times that it isn't an alien concept anymore. It was less of a concern. Going into MA, I was very unsure as to the style of questions, as to a lot of things. (Tony, post)

And as Annie observed, these mock oral performant assessments were not anxiety-provoking:

Yes… maybe because it [the mock] has no marks I am not nervous about that. (Annie, pre)

After the actual exam, some students felt that the experience of the oral performance assessment was better than they had envisaged:

I thought it would be more pressurised, more formal than it was. It was a lot more relaxed, which is quite nice. (Tony, post)

Awareness of the method's requirements emerges also as a strong theme from the interviews connected to assessment anxiety. The students conflate the idea of awareness related to the structure and process of the oral performance assessment with that of predictability. Lack of awareness of process and methods was often mentioned in the interviews as a factor inducing anxiety. The students were not aware, for example, that in an oral performance assessment it is not only the number of questions a student answers that determines the mark obtained, but above all the way in which the questions are answered, and that the students will not all be asked the same questions, as it is the case from a written exam:

Yeah, an oral exam is hard. It's hard to do so in 30 minutes, it's hard to make sure that you ask enough of the student so that they can show you what they know, and that was my concern, but I didn't feel … I mean obviously I was nervous, but I didn't feel like the extra nervousness made the exam worse. (Joey, post)

Lack of insight into what questions could be asked also contributed to anxiety. Students felt they could not tell what questions would be asked—and that the questions they were asked were very different both from what they expected and what they had experienced before:

It's about like different definitions of continuity ... the delta-epsilon definition which is actually not part of the problem that matters here because that's something for real

analysis. Like I can do, but then, it's from last year, so that's why I was thrown off balance ... (Ralph, post)

To gauge whether anxiety regarding the new assessment had influenced marks, we compared performance on the modules in the past three years. We found that the mark distributions were similar (Figs. 4 and 5) indicating that the new assessment modes had not much impact on student achievement overall.

5.6.3 Employability

All interviewed students mentioned the importance of effective oral communication in future jobs, and most thought that practicing such skills was important:

So, I'm planning to go into sales after I graduate in one of the best banks in London and I'm going to have to explain pretty complex things, hopefully complex, to people who don't necessarily have the same complex understanding. So, forcing myself to learn how to break a theorem which may be very complex down to its small constituent parts and then explain each one piece by piece, I think that ability to simplify and kind of piecemeal is very useful in that regard. So it's probably quite a good life skill, just generally. (Tony, post)

Rob discussed how this type of assessment mirrors situations which may occur in the workplace:

I'm just thinking back to my experience in my internship where I had to explain complex like sort of mathematical concepts to clients, and you have to do it on the spot. And if they have any questions, you have to answer on the spot as well, there's no such thing as like, 'Give me 10 minutes to think about it.' You know, an oral exam really simulates like a real-life experience ... (Rob, pre)

The students in the case study saw immediately the relevance of the oral performance assessment experience to what they may be asked to do in the workplace, which probably depended on their previous experience of internships mostly in financial firms. They also acknowledged that perhaps the experience of the closed book exam does not foster the most useful employability skills and oral communication skills may be overlooked in their very uniform assessment diet.

6 Discussion

We recall that the two research questions of this study were:

RQ1: Did the oral performance assessment impact on students' approaches to learning in the modules in which it was implemented? What were the factors that contributed to or prevented such impact?

RQ2: Were there any other consequences of the introduction of such assessment?

The research questions were motivated by the lack of insight into the reasons why a change in summative assessments may result in a change in students' approaches to learning (Scouller, 1998) and by the very uniform assessment diet that mathematics has in the UK at least. We chose oral performance assessment as there is evidence that this dialogic form of assessment may help students express their relational understanding, especially in STEM subjects (Schoultz et al., 2001; Stylianides, 2019). Moreover, Brown et al. (2013) suggest that uniform assessment diets lead to reproductive learning, therefore, contributing variety to students' uniform assessment diet may help

fostering deep learning approaches in mathematics. Recall also here that we draw a parallel between deep and surface learning as characterised by Marton and Säljö (1997) and relational and instrumental understanding as described by Skemp (1976), and we assume with these authors that university education aims to foster deep learning.

For RQ1 we evidence the impact by focusing on factors that the learning approaches theory (Marton & Säljö, 1997) found to be characteristic of deep learning. The first factor is the students' perception that oral performance assessment requires learning for relational understanding. As we noticed, this is not a new finding (Iannone & Simpson, 2015), but it is important in this context as this perception may, in the right circumstances, help students to adopt deep learning approaches. If students had found the assessment to require memorising, then it would not have been plausible to expect evidence of deep learning. The second factor we found regards questioning in connection to preparation for the assessment, as opposed to replicating strategies and predicting questions. The role of questioning in preparation for the assessment is highlighted by students who were able to change their preparation in a way in which Marton and Säljö (1997) suggest will foster deep learning, namely by introducing peer questioning for relational understanding. The change in exam preparation may have been beneficial to students' relational understanding, as it not only encourages them to work collaboratively, but also to formulate their own questions, rather than using the ones which have been prepared by their lecturers. Indeed, as Lithner (2004) reports, this may be a viable strategy for relational understanding, as going through pre-prepared exercises (from textbooks or past exam papers) is likely to foster procedural understanding.

Engagement is the third factor associated with deep learning approaches. The students distinguish between engagement with the subject and engagement with the lectures and peers. While we could consider engagement with the subject akin to learning for relational understanding, we may hypothesise that engagement with the lectures and peers may facilitating a shift to internal motivation for learning. Engagement that was first initiated by external factors (e.g., students wanted to be prepared for an assessment method which they did not know much about and wanted to gauge what their lecturer might be asking in the oral performance assessment) became enjoyment and intrinsically motivated engagement. Intrinsic motivation is a characteristic of deep learning approaches, too. Moreover, attendance and class engagement are correlated to attainment, as Newman-Ford, Fitzgibbon, Lloyd, and Thomas (2008) have shown in their study. In a higher education context, the UK, where attendance to lectures is sometimes problematic and motivation to engage low, and where occasional non-attendance is the norm (Moore, Armstrong, & Pearson, 2008), the introduction of the oral performance assessment may potentially be one way of increasing such attendance.

We note that not all students changed the way in which they prepared for the oral performance assessment. Those who did not, found that revision techniques that served them well in the past were not helpful for this new assessment. This observation brings to the fore the role that familiarity with the assessment plays. By familiarity, we mean the knowledge and understanding of the structure and procedure of the assessment. In the case of oral performance assessment familiarity denotes understanding the role of contingent questioning and that this contingent questioning will be unique to each student. In this context, familiarity also means realising that attainment is not linked to the number of questions answered, but to a large extent to the quality of the answers in terms of handling contingent questions and the ability of students to explain their answers. This is distinct from predictability, as predictability is meant to be the ability to predict what type of questions will be asked in the assessment, as we have described in the theoretical framework. Some of the students were able to become familiar with the structure of the oral performance assessment through attendance to the mock exams while others, perhaps because they only took the module MA which had only one formal mock, were not. The students who became familiar with the oral performance

assessment prepared by mimicking this format: asking each other questions and asking each other to explain their reasoning. Those who did not become familiar with the oral performance assessment reverted to revision strategies that were appropriate to other, very different assessment methods. We argue that failing to foster changes in students' learning approaches and engagement with the subject that should originate from changes in assessment may depend, at least in part, on familiarity. Our data seem to imply that to become familiar with the oral performance assessment students need to experience it and witness the way in which it is implemented also by observing others, as the students enrolled in the module MB had done. The lack of familiarity with the new assessment method and the lack of understanding of its requirements may drive students to change their preparation in undesirable ways, or in no way at all. We argue that this could be the case especially in a subject like mathematics where students are familiar with a few assessment methods and often very successful in at least the most common one: the written exam.

Regarding RQ2 we found two more factors that impacted the student experience: assessment anxiety and oral performance assessment for employability skills. While our students reported being anxious about the oral performance assessment, more than they would be for a written exam, comparing their exam results with results from previous cohorts that took the same modules but were assessed by a written exam revealed that their marks did not differ significantly. Being used to the presence of the assessor as well as to the requirements of the assessment plays a role in reducing anxiety. Students felt that being used to the assessor through engagement in class, for example, would reduce their anxiety levels. Of course, all high stakes assessment induces anxiety in students (Lotz & Sparfeldt, 2017), therefore it is not feasible to think anxiety can be completely eliminated. Some literature suggests ways in which oral performance assessment anxiety can be reduced: Nash, Crimmins, and Oprescu (2016) suggest that public performance anxiety can be reduced with formative, non-assessed activities designed to allow students to become familiar with the format and with public speaking. This is reflected also in our data where the students acknowledge that the experience of the mock assessment had helped them overcome anxiety, especially when the mock exam was public, and students could not only experience their mock exam but also observe that of their peers. Finally, our study indicates that the oral performance assessment can foster employability skills different from those fostered by the written exams and that students ascribe value to this opportunity. The data show just how relevant to real work situations the oral performance assessment was for the students: it was clear to them that being able to explain complex ideas to others was a skill which is rarely practiced in the assessment of mathematics degrees and yet very important in the working life they were about to enter.

Our research has implications for the summative assessment of mathematics at university. While the written exam is certainly an efficient way of assessing mathematics, the reasoning that its questions require to be successful is mostly procedural in nature, as it is currently implemented. As we have also seen in our data, the current structure of written exams supports a kind of preparation that leads students to favour procedural thinking rather than relational thinking: students feel encouraged to develop exam preparation strategies that aim to predict the nature of the questions that will appear in the exam paper (as also observed by Gueudet, 2008). Moreover, as Stylianides (2019) also argues, assessing students exclusively by their written output may give the assessors a negatively biased version of what the students can actually do. The high stakes oral performance assessment, especially for small cohorts of students, can be used for mathematics as an assessment that gives information regarding students' abilities that may be of a different nature than those given by written exams and may contribute to varying this uniform assessment diet. This assessment also

contributes to develop presentation and oral communication skills that the written exam does not, and which are required in the workplace.

The innovation, however, needs to be carried out with caution. Implementations of the oral performance assessment need to be planned carefully and take into account the lack of familiarity with this format for some students, the resulting high anxiety levels and the fact that students may not know how to prepare for it. Communication with students regarding this assessment method, the introduction of a mock practice exam and advice on relevant and effective revision strategies for this assessment method may make these implementations successful for the students. Small trials of the oral performance assessment carried out during the first and second year of study (e.g., as in Iannone & Simpson, 2015) can also help prepare for the high stakes oral performance assessment. By experiencing this assessment method in a summative, but not high stake, context, students may become familiar with its requirements and with successful ways to prepare for it.

7 Concluding remarks

This study reported on a specific way of investigating the impact on approaches to learning caused by the change in summative assessment. We believe that mathematics is an interesting case because of its very uniform assessment diet. We acknowledge that our study did not investigate issues such as validity or reliability in relation to oral performance assessment in mathematics, and we recognise that those are important directions for future research. Finally, we note that due to the nature of the teaching implementation in this case study the cohort of students involved was necessarily small, therefore possibly impacting the generalisation of the results to other universities and other modules. However, we believe that the case study method and the variety of data we have collected offer insights not only in what the effects of this teaching innovation were and its benefits/drawbacks for the students involved but also on the factors that brought about the effects observed. After all, discussing the role of case studies in educational research, Flyvbjerg (2006) acknowledges that:

> Predictive theories and universals cannot be found in the study of human affairs. Concrete, context-dependent knowledge is, therefore, more valuable than the vain search for predictive theories and universals. (p. 224)

Appendix 1

We provide here a sample of the questions asked on one day as part of the mock oral performance assessment for MA. Each question was followed by contingent questions tailored to the student's answers. Each day of mock assessment had a different set of starting questions.

Exam 1

- What is a probability space?
- State the definition of a σ-field and probability measure.
- Can you give an example of a probability space?
- How do you define the Borel-σ-field $B([0,1))$ on $[0,1)$?
- Let $\Omega\{1,2,3,\}$ and $G = \{\{1\},\{3,4\}\}$. What is $\sigma(G)$?
- How do you construct the Lebesgues measure P_L on $([0,1), B([0,1)))$?
- Explain the statement of Caratheodory's Extension Theorem.
- Set $A = Q \cap [0,1)$. What is $P_L(A)$? How do you compute it?
- Prove the statement on the "continuity of measure" that you just used.

Appendix 2

In this appendix, we include the text of the questionnaire we gave the students at the start of the project and we referred to the two sets of questions related to their perception of assessment and oral performance assessment when compared with closed book exams adopted for the questionnaire of this study. The first is the API (Iannone & Simpson, 2015), and the second is the Gibbs and Simpson questionnaire comparing students' perceptions of assessment with oral performance assessment and closed book exam. For further details on the use of the scale and its uses/reliability and measures see Gibbs and Simpson (2003).

The questionnaire

About you

1a. Please indicate your age

I am _____ years old.

1b. Please indicate your gender by ticking the appropriate box

☐ M   ☐ F   ☐ Prefer not to say

1c. Please tell us what your final average grade at the end was of

Year 1 _____ Year 2 _____

About assessment

1. To what extent would you want your achievements in the course to be assessed by each of the following: (from 1 hardly at all to 5 almost exclusively)

		1 2 3 4 5
Multiple choice examination	(e.g., a test taken in an exam room, where for each question you have to select one response from five possible choices)	
Written examination with no support materials	(e.g., a test taken in an exam room, with a separate booklet in which you write solutions, but where you are not allowed to use a calculator, books or any other support materials)	
Written examination with support materials	(e.g., a test taken in an exam room, with a separate booklet in which you write solutions, but where you are allowed a copy of the standard textbook for the course)	
Weekly example sheets	(e.g., a test which you complete in your own time over the course of a week, based on the material covered in the course over that week)	
Project coursework	(e.g., a piece of written work submitted in response to a question or problem, undertaken throughout a number of weeks)	
Project presentation	(e.g., an oral presentation of the results of a project, undertaken in response to a set question or problem, after working on the project for a number of weeks)	
Oral examination	(e.g., working on a mathematical problem on a chalkboard or piece of paper with a tutor present who can provide suggestions or check errors as you work on it)	
Dissertation	(e.g., a substantial piece of written work, on a set topic or problem, undertaken throughout a long period, such as a term or two)	

About assessment by written exam and assessment by oral exam

Now compare written exams to oral exams. Please rate the statements below where:

-2: Much more accurate of oral assessment than written exams.

-1: More accurate of oral assessment than written exams.

0: About the same of written exams and oral assessment

$+1$: More accurate of written exams than oral assessment

$+2$: Much more accurate of written exams than oral assessment

	-2 -1 0 1 2
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 memorising.	
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.	

Appendix 3

Pre-interview protocol

1. Which assessment methods have you experienced in mathematics as a student (at school? at university?)

2. How would you like your mathematics efforts to be assessed? Why?

3. What do you think are the characteristics of a student who is successful in each one of these assessment methods (in mathematics)?

(list of the common assessment methods showed to the students - see Appendix 2)

4. Have you ever experienced an oral assessment before?

5. Do you prefer to be assessed by exams (closed book) or by oral assessment? Why?

6. What did you think about the mock oral exam? Was it the same/different than what you were expecting?

7. What part did the assessment method (if any) played in your choice of this module? Explain.

8. If you were to design an assessment schedule for mathematics students in your year at university (comprising more than 1 method but less than 4) which assessment methods would you include? Why?

Post-interview protocol

1. What was your general impression of the oral exam?

2. What type of questions did you have? Where these questions you had expected? More/less hard? Same/different than a written paper?

3. Did the exam differ from the mock exam much? Did the mock help? Explain.

4. How do you think it went?

5. How did you prepare for it? Like for a written exam? Differently? Why?

6. How do you usually prepare for written exams?

7. Do you think that the oral exam helped in understanding the material better/worst?

8. Would you do it again? Why? (or why not?)

9. What would you say about the assessment to students taking this module next year?

References

Alcock, L., & Weber, K. (2010). Referential and syntactic approaches to proving: Case studies from a transition-to-proof course. *Research in Collegiate Mathematics Education VII*, 93–114.

Attride-Stirling, J. (2001). Thematic networks: an analytic tool for qualitative research. *Qualitative Research*, *1*(3), 385–405.

Bergqvist, E. (2007). Types of reasoning required in university exams in mathematics. *The Journal of Mathematical Behavior*, *26*(4), 348–370.

Birenbaum, M., Brueer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., … Nickmans, G. (2006). A learning integrated assessment system. *Educational Research Review*, *1*(1), 61–67.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, *86*(1), 8–21.

Boesen, J., Lithner, J., & Palm, T. (2010). The relation between types of assessment tasks and the mathematical reasoning students use. *Educational Studies in Mathematics*, *75*(1), 89–105.

Brown, G., Bull, J., & Pendlebury, M. (2013). *Assessing student learning in higher education*. Abingdon: Routledge.

Darlington, E. (2014). Contrasts in mathematical challenges in A-level Mathematics and Further Mathematics, and undergraduate mathematics examinations. *Teaching Mathematics and its Applications: An International Journal of the IMA*, *33*(4), 213–229.

Entwistle, N. J., & Entwistle, A. (1991). Contrasting forms of understanding for degree examinations: The student experience and its implications. *Higher Education*, *22*(3), 205–227.

Entwistle, N. J., & Ramsden, P. (1983). *Understanding student learning*. London, UK: Croom Helm.

Flick, U. (2018). *An introduction to qualitative research*. Los Angeles, CA: Sage.

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, *12*(2), 219–245.

Gibbs, G., & Simpson, C. (2003). Measuring the response of students to assessment: The assessment experience questionnaire. In *11th Improving Student Learning Symposium*, Hinckley (pp. 1–12).

Gijbels, D., & Dochy, F. (2006). Students' assessment preferences and approaches to learning: Can formative assessment make a difference? *Educational Studies*, *32*(4), 399–409.

Gueudet, G. (2008). Investigating the secondary-tertiary transition. *Educational Studies in Mathematics*, *67*(3), 237–254.

Harlen, W., & Deakin Crick, R. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice*, *10*(2), 169–207.

Iannone, P., & Simpson, A. (2011). The summative assessment diet: How we assess in mathematics degrees. *Teaching Mathematics and its Applications: An International Journal of the IMA*, *30*(4), 186–196.

Iannone, P., & Simpson, A. (2015). Students' views of oral performance assessment in mathematics: Straddling the 'assessment of' and 'assessment for' learning divide. *Assessment & Evaluation in Higher Education*, *40*(7), 971–987.

Joughin, G. (2010). *A short guide to oral assessment*. Leeds: Leeds Met Press in association with University of Wollongong.

Kehm, B. M. (2001). Oral examinations at German universities. *Assessment in Education: Principles, Policy & Practice*, *8*(1), 25–31.

Lithner, J. (2004). Mathematical reasoning in calculus textbook exercises. *The Journal of Mathematical Behavior*, *23*(4), 405–427.

Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics*, *67*(3), 255–276.

Lotz, C., & Sparfeldt, J. R. (2017). Does test anxiety increase as the exam draws near? – Students' state test anxiety recorded over the course of one semester. *Personality and Individual Differences*, *104*, 397–400.

Mac an Bhaird, C., Nolan, B. C., O'Shea, A., & Pfeiffer, K. (2017). A study of creative reasoning opportunities in assessments in undergraduate calculus courses. *Research in Mathematics Education*, *19*(2), 147–162.

Marton F., & R. Säljö. (1997). Approaches to learning. In F. Marton, D. Hounsell and N. Entwistle (Eds.) *The experience of learning. Implications for teaching and studying in higher education* (pp. 36–55). Edinburgh, UK: Scottish Academic Press.

Moore, S., Armstrong, C., & Pearson, J. (2008). Lecture absenteeism among students in higher education: A valuable route to understanding student motivation. *Journal of Higher Education Policy and Management*, *30*(1), 15–24.

Nash, G., Crimmins, G., & Oprescu, F. (2016). If first-year students are afraid of public speaking assessments what can teachers do to alleviate such anxiety? *Assessment & Evaluation in Higher Education*, *41*(4), 586–600.

Newman-Ford, L., Fitzgibbon, K., Lloyd, S., & Thomas, S. (2008). A large-scale investigation into the relationship between attendance and attainment: A study using an innovative, electronic attendance monitoring system. *Studies in Higher Education*, *33*(6), 699–717.

Odafe, V. U. (2006). Oral examination in college mathematics: An alternative assessment technique. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *16*(3), 243–256.

Saldaña, J. (2015). *The coding manual for qualitative researchers*. London, UK: Sage.

Schoultz, J., Säljö, R., & Wyndhamn, J. (2001). Conceptual knowledge in talk and text: What does it take to understand a science question? *Instructional Science*, *29*(3), 213–236.

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, *35*(4), 453–472.

Skemp, R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching*, *77*(1), 20–26.

Stake, R. E. (2000). Qualitative case studies. In N. K. Denzin, & S. Y. Lincoln (Eds.), *Handbook of qualitative research* (pp. 443–466). London, UK: Sage.

Steen, L. (2006). Supporting assessment in undergraduate mathematics. The Mathematical Association of America. Washington, DC.

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment and Evaluation in Higher Education*, *30*, 331–347.

Stylianides, A. J. (2019). Secondary students' proof constructions in mathematics: The role of written versus oral mode of argument representation. *Review of Education*, *7*(1), 156–182.

Videnovic, M. (2017a). Oral vs. written exams: What are we assessing in mathematics? *IMVI - Open Mathematical Education Notes*, *7*(1).

Videnovic, M. (2017b). Evidential vs non-evidential beliefs in case of oral assessment: A sequel. *Open Mathematical Education Notes*, *7*(2), 43–47.

Yin, R. K. (2011). *Applications of case study research*. London, UK: Sage.