Teaching Mathematics and its Applications: An International Journal of the IMA (2022) **41**, 22–31
https://doi.org/10.1093/teamat/hrab007 Advance Access publication 6 May 2021

How we assess mathematics degrees: the summative assessment diet a decade on

Paola Iannone[*,†] and Adrian Simpson[‡]

[†]*Mathematics Education Centre, Loughborough University, Epinal Way, LE11 3TU, UK*
[‡]*School of Education, Durham University, Leazes Lane, Durham DH1 1TA, UK*

[*]*Corresponding author. Email:* p.iannone@lboro.ac.uk

[Received 31 July 2020; accepted 15 April 2021]

Two previous studies mapping university mathematics students' summative assessment diet in the UK revealed a clear picture. In general, there was a dominance of closed book examinations with a strong relationship to departmental league table position. The decade since then has seen many changes in higher education in the UK, particularly in the strength of the student voice. The study we report here replicated the earlier work to see if there has been an impact on the assessment diet. While the analysis shows a very small decrease in the use of closed book examinations, this may be accounted for by the addition of adjunct modules, rather than a broadening of the assessment diet across mainstream mathematics topics.

1. Introduction

A previous pair of related studies (Iannone & Simpson, 2011, 2012) mapped university mathematics summative assessment practices across the UK. The outcomes demonstrated that by far the most common summative assessment method for mathematics in the UK was the closed book examination, with other assessment methods mainly associated with modules included in mathematics degrees but which were not focused on learning new mathematics.

While the proportion of closed book examinations (across modules) varied between institutions, it was uniformly high: ranging from just less than 50% to over 90%, and there was a strong correlation between league table position of the department and proportion of closed book exam suggesting that this assessment method was seen as the 'gold standard' in mathematics. The message that emerged was that not only was the mathematics assessment diet across the UK fairly uniform but also the closed book exam was still the overwhelmingly dominant form of assessment for mathematics.

The data in those studies, collected during the academic years 2009–2010 and 2010–2011, reflected the pattern of assessment as it stood at the time. Since then, much has changed in the higher education landscape in the UK. At the policy level, for example, consumer laws now cover universities, and a new Office for Students (OfS) has been formed which has shifted the focus of the sector regulator from institution to student (Dandridge, 2019). In 2018, the newly formed OfS introduced the Teaching Excellence and Student Outcome Framework (TEF) intended to assess the effectiveness and outcomes of teaching in higher education. Other changes have not been so centrally driven but are common across

the sector: for example, it is a near-universal requirement that university lecturers undertake training in teaching, with many institutions requiring Associate Fellowship of the Higher Education Academy. Finally, there has been increasing attention on assessment at university level both in the research literature and in the mathematics community, with both the Institute of Mathematics and Its Applications and the London Mathematical Society funding workshops on assessment of mathematics at the university level in 2019. It would be therefore plausible to expect these changes would be reflected in changed patterns of assessment. We therefore decided to revisit assessment data as it stands currently to find out whether there have been changes in the mathematics assessment diet in the past 10 years. As with the original work, we addressed three questions:

- What mix of methods are observed in mathematics departments in the UK?
- How are different assessment methods used in different topic areas?
- Are there links between the assessment diet and the types of university mathematics department?

2. Background

While the call for a varied assessment diet in mathematics has been made for some time in the research literature (e.g. Gold *et al.*, 1999; Houston, 2001; Steen, 2006), there is now also increased emphasis on such variety in policy documents. In 2019, the Quality Assurance Agency's new Subject Benchmark Statement for Mathematics, Statistics and Operational Research (QAA, 2019) highlighted the characteristics of assessment appropriate to these disciplines. The document recognizes the need for a variety of assessment methods to reflect, first, the variety of skills that the QAA believes undergraduate mathematics students need to master and, second, students' varying aptitudes for different assessment methods.

The research literature has also moved on since the time of the previous work (Iannone & Simpson, 2011, 2012). Its focus has increased in two main areas: the potential of e-assessment for both formative and summative functions and the use of dialogic forms of assessment.

E-assessment has gained much popularity in the past 10 years with systems like STACK (https://www.ed.ac.uk/maths/stack/) and NUMBAS (https://www.numbas.org.uk) now used by many universities, often for coursework components of first-year modules. The popularity of these systems depends not only on the time-saving aspects of electronic marking but also on the potential of such systems to have both a formative and summative function and to provide feedback rapidly. The research in this area mostly investigates the reasoning skills that questions built in these systems can successfully assess. This research sits in the context of findings both in the UK (Darlington, 2014) and elsewhere (Mac an Bhaird *et al.*, 2017) about the reasoning skills elicited by some closed book exam questions: both studies noting that it was possible to perform well in the exams in their sample by just implementing taught, mechanistic procedures. Sangwin (2019) investigated whether current closed book exam questions in linear algebra could be replicated in an e-assessment system and found that this is possible for most types of questions in the exam sample selected. In finding that most of the chosen sample could, the author questioned the value of posing such tasks in exam papers at the expense of more conceptual questions that cannot be solved by computer algebra systems. Rasila *et al.* (2015), however, held out the hope of assessing conceptual understanding through automated assessment systems. There does appear to be much in common between closed book and e-assessment methods in the sense that both involve the written expression of knowledge (albeit there may be substantial differences with fluency of handwriting and typing mathematics), they tend not to allow access to external materials or other assistance and they tend to be time limited to a few hours at most.

A second direction that research on summative assessment of mathematics has taken is that of investigating assessments which differ substantially from the written assessment, such as those taking a dialogic form. Videnovic (2017a,b) reported on her interviews with mathematicians teaching university modules, concluding they believe written exams—especially as they are currently structured—are not always a good indicator of students' abilities and that oral exams would be a better indicator of what the students can do. Oral exams, although popular in many other countries and occasionally in the UK for other subjects, are not commonly in use in UK mathematics departments. From the viewpoint of mathematics students, Iannone & Simpson (2015) and Iannone *et al.* (2020) highlighted the potential of oral assessment to foster revision strategies conducive to conceptual understanding, to enhance participation in classroom activities and to be generally appreciated by students for the immediacy of feedback.

Despite the growing interest in these two areas of summative assessment of mathematics at university level, research in this field is still limited and much of the discussion on summative assessment of mathematics takes place in professional journals such as the MSOR Connections (https://journals.gre.ac.uk/index.php/msor/index) and in dedicated professional conferences. In these forums, issues of implementation of e-assessment (Greenhow, 2019) and other innovative assessment activities such as using voting systems during lectures (Evans, 2018) are discussed and these forums act as a platform to share ideas amongst those who teach mathematics at university.

That growth in both research and professional literature also motivates a re-examination of the state of summative assessment practice a decade later. In what follows, we describe the methods of the study, the current findings and then a comparison between the assessment diet of mathematics students between the first studies and now.

3. Methods

The methods of the study reported in this paper closely resemble those used in Iannone & Simpson (2011). The data were drawn from degree programmes most closely fitting the common 3 year pattern (noting the increase in four year degrees in many institutions). In each case, this was the one coded as 'G100' in the Higher Education Statistics Agency's coding system. Sampling could not be fully random as not every institution has sufficient data publicly available; so the representatives were chosen according to the completeness of the set of available assessment information and to cover a wide range of institutions. The availability of more detailed public information regarding the assessment diet of mathematics degrees allowed the sample of the current study to be larger than the 2011 study, maintaining the same methodology. In order to obtain a good spread of institutions, a stratified sample of 23 universities was obtained by selecting one in every group of four in rank order from a publicly available league table of mathematics departments. Note that, while not the intention, no universities in Scotland (which has a somewhat different higher education system) ended up in the sample. This was also the case in Iannone & Simpson (2011); however, Iannone & Simpson (2012) did include Scottish universities and closely replicated the findings of the 2011 study. So, while the research question originally referred to the UK, the results should be seen as applicable only in the context of England and Wales.

As with the original work, given the freedoms and restrictions to choose modules at various institutions, to calculate an estimate of the percentage of marks accrued by closed book exams during the degree in one institution, we calculated the mean across all the modules on offer as part of the programme, weighted according to the credits each module accrued. Given the complexity of module choice, as in the previous work, we ignored any contribution to the assessment diet of students taking modules outside

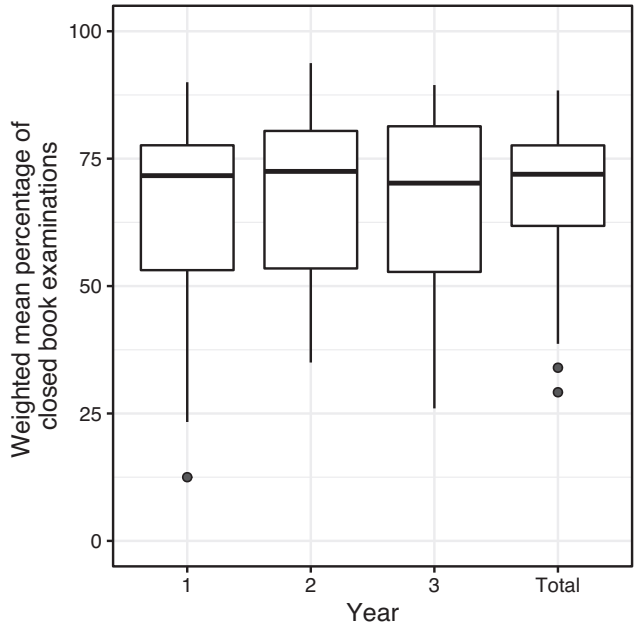

FIG. 1. Weighted mean percentage of closed book examinations, by year and in total.

those offered by the mathematics department. Thus, our analysis might not reflect the experience of a particular student able to select a more or less varied assessment diet than our simplified description.

Data for 1017 modules were collected and for each module we recorded the assessment pattern and number of credits. Data were coded according to the information provided in available course documentation, which might not always reflect actual practice (for example, what is described as a closed book examination at one institution might allow sufficient material to be used to be described as open book elsewhere). Although the data we collected are publicly available, we chose not to identify institutions in the analysis.

4. Analysis at degree course level

We collected complete datasets for 23 departments, represented here using the letters A to W, ordered by the largest weighted mean percentage of closed book examinations across their degree programme.

Considering the sample as a whole, we examined the contribution of closed book examinations to each year group and to the full degree, illustrated in Figure 1. Taking the mean across all their modules, weighted by credits, the median university used closed book examinations for 72% of their assessment. We observed no particular pattern in the use of closed book examinations between years.

In Table 1, we report the full dataset summary, the number of modules offered as part of the programme, the percentage of marks accrued by closed book exam in the first, second and third years

TABLE 1. *Percentage of marks accrued by closed book examinations, mean within years and across the whole set of modules weighted by credit value for each institution (with numbers of modules offered in parenthesis).* * *indicates a Russell Group institution,* $^{+}$ *indicates a Post* 92 *institution*

Department	Year 1	Year 2	Year 3	Total
A*	85 (6)	92 (12)	88 (27)	88 (45)
B*	88 (6)	94 (7)	72 (13)	85 (26)
C*	88 (8)	82 (18)	83 (62)	84 (88)
D	71 (10)	82 (9)	89 (18)	81 (37)
E*	72 (6)	87 (10)	80 (33)	79 (49)
F*	75 (6)	75 (13)	83 (39)	78 (58)
G*	79 (8)	68 (28)	85 (84)	78 (120)
H	76 (11)	79 (14)	68 (16)	74 (41)
I*	56 (8)	81 (17)	83 (26)	73 (51)
J	76 (7)	80 (9)	62 (12)	73 (28)
K^{+}	90 (6)	68 (6)	58 (8)	72 (20)
L	66 (8)	80 (13)	70 (24)	72 (45)
M	80 (8)	80 (11)	48 (15)	69 (34)
N*	73 (10)	64 (12)	70 (32)	69 (54)
O	50 (10)	72 (8)	78 (24)	67 (42)
P*	74 (8)	54 (30)	71 (34)	66 (72)
Q*	58 (9)	71 (13)	57 (28)	62 (50)
R^{+}	70 (6)	53 (11)	62 (14)	62 (31)
S	30 (5)	52 (11)	47 (16)	43 (32)
T^{+}	50 (6)	35 (6)	40 (13)	42 (25)
U^{+}	45 (4)	45 (6)	26 (22)	39 (32)
V^{+}	23 (6)	44 (6)	34 (8)	34 (20)
W^{+}	12 (4)	45 (6)	30 (7)	29 (17)

and the overall percentage. Of the departments listed in Table 1, nine belong to universities in the Russell Group and five to post-92 universities.[1]

In the previous work, we had been interested in the relationship between league table position and the use of closed book examinations. It can be argued that any relationship would depend on the choice and weighting of factors selected for inclusion in summary league tables which can vary considerably (Simpson, 2018). In this analysis, we examined the relationship between institutional grouping and use of closed book examinations. There was a strong association in our sample (Kruskal–Wallis rank sum test, $\chi^2(2, N = 23) = 10.2, p = 0.006, \epsilon^2 = 0.46$) with the Russell Group (median = 77.6) non-significantly dominating the unaligned group (71.9), but both significantly dominating the Post-92 institutions (40.2). Figure 2 illustrates the disparities between the groupings.

While league table scores are aggregates of scores for many different elements purported to assess institutional and departmental performance, most include elements of student satisfaction and research performance. The former is normally measured in the UK by the annual National Student Survey and the latter by the Research Excellence Framework. Given the genesis of the alignment groupings, the strong positive relationship with research assessment we found ($r^2 = 0.38, p = 0.003$) is to be expected. It is

[1] The UK changed the mechanism for granting university status in 1992, allowing many former polytechnics to gain university status (identified as 'post 92' here). The Russell Group is a self-selecting collection of 'research intensive world-class' UK universities. The categories are disjoint.

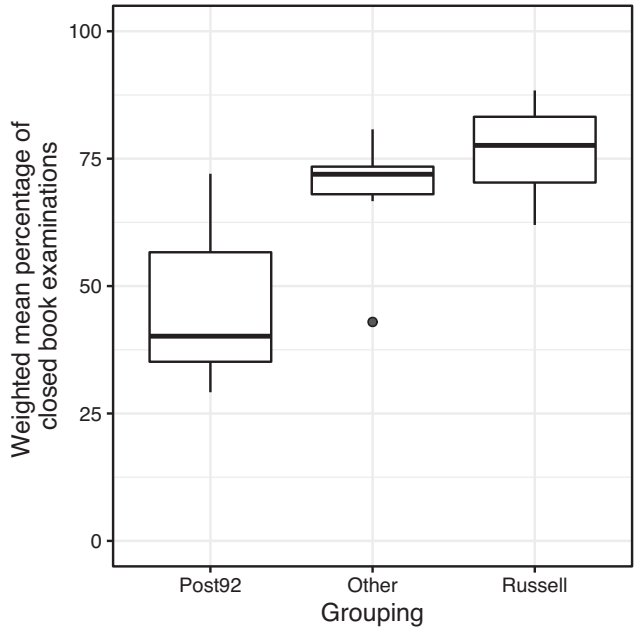

FIG. 2. The distribution of use of closed book examinations by institutional alignment.

interesting to note, however, that the use of closed book examinations has a modest *negative* relationship with student satisfaction ($r^2 = 0.19$, $p = 0.04$).

5. Analysis at module level

Following the analysis of the previous studies, we also investigated assessment patterns at module level, examining topics within degree programmes which appeared to have forms of assessment different from the overall pattern. The topics reported below, identified from a simple analysis of module title, are traditionally found in mathematics degrees (e.g. probability and statistics) or are topics which have seen wide introduction in mathematics degrees more recently (e.g. financial mathematics and skills modules).

5.1 Probability and statistics

Not surprisingly, all departments surveyed offered modules in probability and statistics. At the median department, 14% of modules were in this area, though one (T) has only a single module. There was no particular evidence of a difference in the proportion of closed book examinations in this topic compared to the modules as a whole (median percentages both 80%, Wilcoxon rank sum $W = 62861, p = 0.9$). While we noted few open book examinations in the dataset compared to the previous study, this may be an artefact of the way the data presentation has changed, with fewer institutions classifying, say, statistics examinations where students can access tables as open book, which was often the case in the dataset collected for the 2011 study.

5.2 History of mathematics and mathematics education

We found many departments that offered history of mathematics modules (4 from the sample of 23) and mathematics education modules (12 out of 23). Again as found in our earlier survey, these modules are assessed predominantly by coursework that often takes the form of an essay or a presentation followed by a Q&A session. Indeed, only two modules, one in history of mathematics and one in mathematics education, used examinations for any summative assessment. Some of the modules in mathematics education appeared to include a short school placement and seem to be aimed at students who are thinking about a teaching career. There is one department in the sample that also offers modules in mathematics and music and philosophy of mathematics (again assessed by essays).

All modules in this area appear to be optional and they are offered either in the second or third year of study. The assessment of these modules reflects the assessment of similar modules in social sciences and history departments.

5.3 Projects

Most departments (14 of the 23) offer a final year project, with many giving double the credit weighting for them. These projects can be in pure or applied mathematics or in statistics, with at least one institution offering projects in mathematics education to mathematics students. A small number of modules, although not always having 'project' in their title, are also assessed by a written report coming from either in-depth individual or group study. Such modules tend to be in the second year of study and may be considered as preparation for the final year project.

5.4 Computational and programming modules

Around 5% of the module in the sample are computational and programming modules. All departments surveyed offer at least one module with some computational aspect, and 18 out of the 23 offer at least one dedicated computational or programming module. These often still rely heavily on closed book examinations (median percentage = 60%) though this is clearly below the level for other modules ($W = 33876, p < 0.0001$). These modules often have a large coursework component, commonly involving writing a computer program in the students' own time.

5.5 Skills modules

Seven departments offered generic skills modules and 18 offered mathematics skills modules, mostly assessed by coursework in the form of reflective essays, group work, individual or group presentations. The generic skills modules are often focused on employability skills such as time management, CV writing and presentation skills. The mathematics skills modules offer a variety of content, from introduction to proof, through communicating mathematics, to basic logic and LaTeX typesetting. In this sample, four departments offered a problem solving module, which is a smaller proportion than in the previous samples.

6. Comparison with the previous studies

6.1 Overall comparison

The main lesson from comparing the data which reflects the assessment diet for undergraduate mathematics in 2010 with the assessment diet now is that little has changed. There are 19 departments for

which we have complete data for both dates. For these departments, there has been a small decrease in the mean total percentage of the use of closed book examinations from 72% to 68%. It is difficult to be very confident that this reflects a clear tendency (Wilcoxon signed rank $V = 52, p = 0.09$), and the changes have not been evenly spread across groupings. Indeed, in the post-92 universities, there has been a small increase from 45% to 47%. However, in the Russell Group institutions, there has been a decrease in the use of closed book exams from 83% to 76% and in the remaining group there has been a smaller decrease, from 68% to 66%. The relationship with grouping remains the same: the Russell Group and unaligned institutions dominate the post-92 institutions in their use of closed book examinations and while the Russell Group has a higher median, which is not statistically significant, the measure of association has decreased (from $\epsilon^2 = 0.70$ to $\epsilon^2 = 0.46$).

6.2 Comparison at module level

Through the analysis at module level, we found some similarities and differences with the analysis of the earlier data. The role of the project in the third year and the presence of mathematics education and history of mathematics modules have continued, with many departments offering these modules and assessing them predominantly by coursework. In mathematics education modules, there seem to be a growing number of departments that include a small placement component in secondary schools. These modules seem to be aimed at students who are thinking of a teaching career, rather than aimed at those with an academic interest in mathematics education. The number of computational modules and programming modules has increased with respect to 2010—with all departments in our sample offering such modules—and these use somewhat lower levels of closed book examinations. Statistics modules seemed to be assessed by open book exams more often in the previous studies than in the most recent dataset, but this may be an artefact of the way the data presentation has changed, with fewer institutions classifying, say, statistic examinations where students can access tables as open book. The increased presence of skills modules (whether generic or mathematically focussed) is also a notable change and these are commonly assessed with more coursework than the typical mathematics module. This may also account for a decrease in modules directly aimed at problem solving in mathematics as these mathematics skills are now included along with others in mathematics skills modules.

7. Discussion and concluding remarks

The evidence of this survey is that there remains a relatively low level of variety in the assessment diet. Mathematics departments still use closed book examinations to a great extent, albeit that the use still seems to covary strongly with institutional grouping, with those whose roots lie in previous polytechnic status using coursework to a much larger extent than older universities with more traditional roots.

Rather than seeing a diversification of assessment in existing modules, the small change in the use of closed book examinations may be the result of the growth of alternative modules. We observed more skills modules, more computational and programming modules and an increase in the number and weight of final year projects. Each of these changes acts to bring down the proportion of marks accrued by closed book examinations in total. That is, we may not be seeing a broadening of the assessment diet within modules but a broadening of the diet of modules across degree programmes. We may also argue that these changes, at least to some extent, are the result of an increased emphasis on graduate employability brought by the increasing emphasis on league tables and TEF (where employability is one of the metrics). Introducing adjunct topics in general skills modules, short school placements and programming modules

may support the aim of making graduates more employable given the necessity of classroom experience to apply for teaching qualifications and the general need for programming experience in many jobs.

Reflecting back to the QAA (2019) benchmark statement's suggested shape of assessment in university mathematics, we notice that overall the variety of assessment diet that is advocated by this document is not realized within mathematics modules but may be present in other modules across a mathematics degree. This may suggest that while it is now acknowledged by mathematics departments that graduates need to acquire a variety of skills, these can be assessed in subjects seen as peripheral to mainstream mathematics. For mathematics, the gold standard of assessment still remains as the closed book exam. This may reflect a continuing implicit belief that 'when analysing performance on written tests,... it is as if writing in solitude in the context of a test is an unbiased indicator of what people know or understand.' (Schoultz *et al.*, 2001, p. 214).

Acknowledgements

We would like to thank Cameron Metcalfe who collected the data for this study as part of his third-year project at Loughborough University.

References

Dandridge, N. (2019) The Office for Students: reflections on our first year. *Perspect. Policy Pract. High. Educ.*, 23, 158–161.

Darlington, E. (2014) Contrasts in mathematical challenges in A-level Mathematics and Further Mathematics, and undergraduate mathematics examinations. *Teach. Math. Its Appl.*, 33, 213–229.

Evans, K. (2018) An overview of the Pingo audience response system in undergraduate mathematics and statistics teaching. *MSOR Connect.*, 17, 25–31.

Gold, B., Keith, S. & Marion, W. (eds) (1999) *Assessment Practices in Undergraduate Mathematics*. Washington, DC: Mathematical Association of America.

Greenhow, M. (2019) Embedding e-assessment effectively. *MSOR Connect.*, 17, 63–70.

Houston, K. (2001) Assessing undergraduate mathematics students. *The Teaching and Learning of Mathematics at University Level: An ICMI Study* (D. Holton ed.). Dordrecht: Kluwer Academic Publishers, pp. 407–422.

Iannone, P., Czichowsky, C. & Ruf, J. (2020) The impact of high stakes oral performance assessment on students' approaches to learning: a case study. *Educ. Stud. Math.*, 103, 313–337.

Iannone, P. & Simpson, A. (2011) The summative assessment diet: how we assess in mathematics degrees. *Teach. Math. Its Appl.*, 30, 186–196.

Iannone, P. & Simpson, A. (eds) (2012) *Mapping University Mathematics Assessment Practices*. Norwich: University of East Anglia.

Iannone, P. & Simpson, A. (2015) Students' views of oral performance assessment in mathematics: straddling 'assessment of' and 'assessment for' learning divide. *Assess. Eval. High. Educ.*, 40, 971–987.

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. *Res. Math. Educ.*, 19, 147–162.

QAA (2019) Subject Benchmark Statement: Mathematics, Statistics and Operational Research. Accessed on 9 July 2020 at https://www.qaa.ac.uk/docs/qaa/subject-benchmark-statements/subject-benchmark-statement-mathematics-statistics-and-operational-research.pdf?sfvrsn=;e8f3c881_4.

Rasila, A., Malinen, J. & Tiitu, H. (2015) On automatic assessment and conceptual understanding. *Teach. Math. Its Appl.*, 34, 149–159.

Sangwin, C. (2019) Developing and evaluating an online linear algebra examination for university mathematics. *Eleventh Congress of the European Society for Research in Mathematics Education*. Utrecht, NL: Freudenthal Group; Freudenthal Institute; ERME, pp. 4121–4128.

SCHOULTZ, J., SÄLJÖ, R. & WYNDHAMN, J. (2001) Conceptual knowledge in talk and text: what does it take to understand a science question? *Instr. Sci.*, 29, 213–236.
SIMPSON, A. (2018) The structure of surveys and the peril of panels. *Stud. High. Educ.*, 43, 1334–1347.
STEEN, L. (ed.) (2006) *Supporting Assessment in Undergraduate Mathematics*. Washington, DC: The Mathematical Association of America.
VIDENOVIC, M. (2017a) Oral vs. written exams: what are we assessing in mathematics? *IMVI Open Math. Educ. Notes*, 7, 1–7.
VIDENOVIC, M. (2017b) Evidential vs non-evidential beliefs in case of oral assessment: a sequel. *Open Math. Educ. Notes*, 7, 43–47.

Paola Iannone is a senior lecturer in the Mathematics Education Centre at Loughborough University. Her research interest is teaching and learning mathematics at university level, with particular focus on summative assessment and oral performance assessment, the use of automated theorem provers in transition to proof modules and the transition from school to university mathematics.

Adrian Simpson is The Principal of Josephine Butler College and a professor in mathematics education in the School of Education at Durham University. His recent work has focussed on the use and misuse of evidence in education research though he is also interested in students' thinking across the transition from school to university mathematics and mathematical logic and rationality.