Students' preferences in undergraduate mathematics assessment

Existing research into students' preferences for assessment methods has been developed from a restricted sample: in particular, the voice of students in the 'hard-pure sciences' has rarely been heard. We conducted a mixed method study to explore mathematics students' preferences of assessment methods. In contrast to the message from the general assessment literature, we found that mathematics students differentially prefer traditional assessment methods such as closed book examination; they perceive them to be fairer than innovative methods and they perceive traditional methods to be the best discriminators of mathematical ability. We also found that although students prefer to be assessed by traditional methods they are also concerned by the mix of methods they encounter during their degree, suggesting that more account needs to be taken about the students' views of this mix. We discuss the impact of the results on the way general findings about assessment preference should be interpreted.

Keywords: Students' preferences, assessment of ability, undergraduate mathematics, mixed methods.

Introduction and background

There appears to be widespread agreement that students' preferences for assessment methods influence their learning and how they interact with teaching (Scouller 1998; Harlen and Crick 2003; Gielen, Dochy and Dierick 2003; Boud and Falchikov 2007). Birenbaum (2007) argues that finding out about students' assessment preferences and what factors underpin those preferences is increasingly important in institutions with a service orientation which seek to accommodate students' views. However, Birenbaum goes on to note that research into students' preferences is relatively scarce and focused mainly on preferential choice between assessment formats.

Much of the work in this area tends to compare multiple choice tests to essays and traditional examinations. It points to a preference for multiple choice questions or short

response questions over traditional essay style or more in-depth questions. Moreover, they suggest student preferences are biased towards methods with low discrimination (giving high marks irrespective of ability) and which can be prepared for easily. For example, Traub and MacRury (1990) suggested that students in their sample preferred multiple choice questions over essays because they felt they would achieve higher scores on them and were easier to revise for. Similarly, both Scouller (1998) and Zeidner (1987) implied that a preference for multiple choice tests was based on students viewing them as covering a wide range of the syllabus superficially, avoiding the deeper focus required of a traditional essay or examination question.

When the exploration of assessment methods widens beyond a comparison of multiple choice with essays and examinations, the research suggests a general dissatisfaction with the traditional methods. For example, Sambell, McDowell and Brown's (1997) study, which was based on a wide selection of students on 13 different degree programmes, suggested that exams were "invariably seen as a regrettable or second-rate situation" (p357) and that alternative forms of assessment such as projects and presentations were both preferable and more exacting. Moreover, Sambell et al. suggest that these alternative forms were also seen as fairer than traditional examinations. Similarly, Kniveton's (1996) study of students across 7 degree programmes suggested that continuous assessment was seen as fairer and a better discriminator of ability than examinations.

More recently however, Joughin (2010) has questioned the way in which some key studies on students' perceptions of assessment have been interpreted. He suggested that the oversimplified way in which these studies have been used does not take into account

their context and the cognitive demands both of the assessment type and of the discipline of study.

This latter issue is of particular interest when the sources of the data for much of the research in this area are examined. Struyven, Dochy and Janssens (2005) conducted a comprehensive review of research on students' perceptions about assessment which analysed the findings of over fifty different empirical studies. Table 1 shows an analysis of the academic subjects studied by the students who took part in these studies.

[Table 1 goes here]

It is clear that the research has not drawn evenly from all areas, with large samples from applied subjects such as psychology and medicine. Indeed, over three quarters of the students in these studies come from just four broad disciplines: psychology, biological sciences, engineering and economics, which may not be representative of the spectrum of subjects.

Biglan's (1973) seminal paper classified academic subjects by identifying key underlying dimensions (such as hard vs. soft and pure vs. applied) and this classification has been used repeatedly to distinguish discipline areas (e.g. Becher, 1989; Wareing, 2009). An analysis of the samples in the subjects in Table 1 against this classification shows that three of Biglan's four quadrants – hard-applied, soft-applied, and soft-pure subjects – are relatively well explored. However, there is only one study from the hard-pure area: a small scale study of physics students' views of portfolios as an assessment tool which did not compare or analyse preferences for different assessments.

More recent work which draws on Biglan's classification scheme (Becher 1989, 1994) suggests that there are significant differences between cognitive demands and social structures of academic disciplines. Becher (1994) notes the different structuring of knowledge across disciplines, where the pure-hard sciences are characterised by knowledge which is "…cumulative; atomistic (crystalline/ tree-like); concerned with universals, quantities, simplification; resulting in discovery/explanation" (p154) while the soft-applied subject structure is "functional; utilitarian (know-how via soft knowledge); concerned with enhancement of [semi-] professional practice; resulting in protocols/procedures" (p154). Such differences are also reflected in the views of professional bodies: for example, the London Mathematical Society (LMS 2011) note "there are significant cognitive differences between learning in mathematics and in most other subjects".

Given the importance of disciplinary differences, Joughin's (2010) concern for the lack of appreciation for academic context in existing literature and the lack of existing research reporting the voice of those studying hard-pure subjects, this paper examines whether students within that area might have similar patterns of assessment preference and rationales for that preference to those in the literature in general. Kolb (1981) argues that the heart of the hard-pure category is mathematics so our expectation was that potential issues would be most clearly highlighted by focusing on mathematics students.

Assessment of mathematics at university in the UK

Despite a rich literature on advanced mathematical thinking, there appears to be no existing work which takes an empirical approach to investigating students' preferences for assessment methods in undergraduate mathematics. There is, however, a body of

scholarly writings in mathematics education advocating the introduction of new and – for mathematics – 'innovative' assessment forms.

These include projects (Berry and Houston 1995), poster presentations (Houston 2001), multiple choice questions (Haines and Crouch 2005; Ramesh 2009), oral assessment (Levesley 2011) or a combination of projects, posters and presentations (Povey and Angier 2006). The rationale behind the calls to innovate is that traditional closed book examinations are not seen as fostering understanding of this subject and may be seen as responsible for the association that students appear to make between being good at mathematics and simply having a good memory (Schoenfeld 1989).

The mix of assessment methods encountered, the connection between methods and subject matter and the differential contribution of methods to final marks is described by Iannone and Simpson (2013) as "the assessment diet". In mathematics, the assessment diet in the UK is surprisingly uniform (Iannone and Simpson 2011). By far the most common assessment method is the closed book examination. Iannone and Simpson found a correlation between university ranking and increased use of closed book examinations and found, in almost all institutions, this form of assessment accounts for the large majority of the final mark.

There may be many reasons for this. Gibbs (2006) noted that, although there have been strong pressures on higher education to adopt innovations in assessment, other pressures act to retain traditional systems: "Instead of being imaginative and innovative, assessment reverts to simple and crude basics" (p21). Elsewhere, others have suggested that the mathematics community is resistant to change and should be held responsible

for the lack of innovative assessment at university (Burton and Haines 1997). However Becher's (1994) work on disciplinary differences suggests that mathematics demands different cognitive processes from other disciplines and this, rather than an inherent conservatism, might largely account for the uniform assessment diet in the field.

So, the general literature suggests students have a strong dislike of traditional forms of assessment and prefer those giving high marks with low discrimination; but mathematics is dominated by very traditional forms of assessment – particularly in those institutions with high entry requirements. Thus we would expect to see a level of dissatisfaction with assessment methods amongst mathematics students and an inverse correlation between discrimination and preference.

To explore this, we developed four research questions:

RQ1: Do students on mathematics degrees show the same general tendency to prefer less traditional forms of assessment?

RQ2: What forms of assessment do mathematics students consider best at discriminating between students of different abilities?

RQ3: To what extent are student preferences related to their experience of assessment or to their views of how well methods discriminate?

RQ4: What factors underlie students' preferences about assessment?

Methods

Our research adopts a mixed methods design (Johnson and Turner 2003). The first three research questions are designed to reflect those asked in the existing general assessment preference literature, and the approach we took mirrored the existing

methods in many studies. That is, following Birenbaum (1994), Gijbels and Dochy (2006) Amin, Kaliyadan and Al-Muhaidib (2011) etc. we adapted the Assessment Preferences Inventory (API) for our student population to obtain quantitative data on student preferences.

To obtain a suitable sample size, we administered the API as an online questionnaire at two universities (which we designate Uni1 and Uni2). Both mathematics departments in these institutions are research intensive and demand high entrance grades from their students. The assessment diet of these institutions is representative of that of other similar research-intensive institutions in the UK (Iannone and Simpson 2011). For example, in both institutions coursework accrues marks only in the first year and all assessments in the first year are 'progression only' (i.e. marks accrued during this year do not count towards the final degree classification).

Assessment in all years is largely dominated by closed book examinations, though both universities use a somewhat wider range of methods in later years. For example, Uni1 has a module assessed by a project in year 3 and Uni2 has a similar module in year 2. After we administered the questionnaire, we conducted follow up semi-structured interviews with some respondents to focus on the final research question: the reasons for their assessment preferences.

Quantitative Methods

In adapting the API to mathematics we included assessment methods which feature in the current assessment diet of mathematics in the UK, or which have been suggested in the literature as suitable for assessing mathematics (see Appendix 1). The API was

converted to an online questionnaire for ease of access across both institutions. Dillman and Bowker (2001) suggest that while type of delivery does have an effect on response, this tends to be confined to different modes of delivery (e.g. oral – such as a telephone questionnaire – versus written) and that there are no significant differences between the responses given to written questionnaires and the equivalent questions delivered online and so there were few concerns about delivering the API online.

An email was sent to every undergraduate student on mathematics degrees at each university. They were invited to click on a link taking them to the online survey. At Uni1, 96 students responded (19% of the population) and at Uni2, 52 students responded (16% of the population). These response rates are in keeping with this type of delivery (Sax, Gilmartin and Bryant 2003). In total 16 of the responses were removed from the analysis for missing data.

Table 2 shows the distribution of the remaining data: while there is a relatively even distribution between genders, there is some skew towards students in year 1 in Uni2.

[Table 2 goes here]

Both institutions offer an MMath, four year degree programme in addition to the three year BSc, though only the minority take the extra year; thus we have conflated the small number of year 4 responses with the year 3 responses.

The two universities have very similar assessment diets and we shall see that, unsurprisingly, there were few differences in response – there were no interaction effects for either the preference or discrimination responses with the university (except

one mentioned in the analysis below). Thus, for most of the analysis, the two groups are conflated. In contrast to Kniveton (1996), there were also no main or interaction effects involving gender and so this factor also plays no further role in the analysis.

The questionnaire (reproduced in the Appendix) was split into two sections – students' views of assessment methods, levels of discrimination and students' preference for assessment methods. The sections were presented in random order. In each section, the respondents were asked to rate each of eight assessment methods on a five point Likert scale.

Results

Our first research question concerned whether mathematics students conform to the tendency seen in the general literature to prefer less traditional forms of assessment. Table 3 gives the mean (and standard deviation) of the student responses to the questions regarding their preferences for the different assessment methods. It also gives the results of pairwise *t*-tests (with *p*-values scaled for Bonferroni correction) which show that there are no significant differences between closed book, example sheets and open book; between open book and projects; projects and dissertations; dissertations, oral exams and multiple choice and between oral exams, multiple choice and presentations. The other differences are significant (at the 0.05 level).

[Table 3 goes here]

Figure 1 shows these results graphically with the means (with standard error bars).

[Figure 1 goes here]

This shows a strong preference for traditional assessment methods, in marked contrast to the findings of the general assessment preference literature.

A two-way analysis of variance showed that there was a significant interaction between assessment method and university ($F(7,1040) = 36.6$, $p < 0.001$) but post-hoc t-tests with Bonferroni adjustment showed that this was due to only one difference: while students at Uni2 had rated closed book examinations highly, they had not rated them quite as highly as students at Uni1 ($t(85) = 3.80$, $p = 0.002$).

Recall that the students in our sample have a very restricted assessment diet. All will have had extensive experience of closed book examinations and weekly example sheets. As the years progress however, students do get some occasional experience of other forms of assessment; though some methods, such as oral examinations, play no role at all in their assessment diet. Only some of the final year students will have had (on-going) experience of dissertations.

One might argue that the preference for closed book examinations and example sheets thus only reflects a conservatism on the part of the students: they tend to prefer what they are used to (and demonstrably successful at, given the role closed book examinations played in their admission to these highly successful institutions). However, this would not explain the relatively strong preference for open book examinations (which few will have experienced) and the low preference for multiple choice (which, while playing little role in undergraduate mathematics assessment for these students, are methods they would be familiar with and presumably successful at

during their school studies). Moreover, if there was some underlying conservatism, we would expect to find a quite different pattern of preference as the years progress, particularly given the influence in later years of projects on final marks. So we explored if there was an effect for year group on preference. This also allowed us to address part of our third research question: the extent to which experience affects students' preferences.

Figure 2 shows an interaction plot with the mean preference responses for each assessment method for each year group. A two-way anova shows a significant interaction for assessment method with year group ($F(14,888) = 32.3$, $p = 0.023$). An examination of the interaction plot suggests that the largest differences lie in both dissertations and presentations being differentially preferred by finalists compared to year 2 students. However, Bonferroni corrected post-hoc *t*-tests indicate these differences should not be seen as significant ($t(68) = 2.46$, $p = 0.272$ for dissertations and $t(68) = 2.80$, $p = 0.112$ for presentations). These non-significant differences aside, the interaction plot shows remarkable level of unanimity across years.

[Figure 2 goes here]

Thus, while there is some suggestion in our data of a small influence of experience, the student preferences do not appear to be determined by an inherent conservatism. However, even if this was the case, the results still stand in contrast to the general sense from the literature that students prefer less traditional forms of assessment: in our sample, mathematics students rate less traditional assessments such as presentations and projects, along with multiple choice, amongst their least preferred options. Mathematics

students appear to prefer the more traditional forms, such as closed book exams and weekly example sheets.

Students' perceptions of assessment discrimination

Our second research question concerned students' perceptions of the extent to which different forms of assessment discriminate on the grounds of ability.

Table 4 gives the mean (and standard deviation) of the student responses to the questions regarding the extent to which assessment methods distinguish between those who are good and poor at mathematics. It also gives the results of pairwise *t*-tests (with Bonferroni correction) which show that closed book exams are seen as significantly more discriminating than all other methods; dissertations, open book exams, example sheets, oral exams and projects have no significant difference between them, but are significantly more discriminating than presentations and multiple choice (with the former significantly more discriminating than the latter).

[Table 4 goes here]

Figure 3 shows these differences graphically by displaying the means with standard error bars. There was no significant interaction effects for discrimination with year group ($F(14,888) = 21.19$, $p = 0.09$). The effect size for the extent to which closed books are seen as more appropriate to discriminate between good and poor mathematics students is relatively large ($d = 0.71$), as is the effect size between presentations and multiple choice ($d = 0.71$), while the effect size between projects and presentations is very large ($d = 1.141$) That is, students tend to perceive the assessment methods in four

distinct groups: closed book as by far the best at discriminating on the grounds of ability, a group of dissertations, open book, example sheets, oral exams and projects as somewhat in the middle; presentations as less discriminating still and multiple choice as by far the poorest discriminator.

[Figure 3 goes here]

To address the final part of our third research question, we asked whether students tend to prefer assessment methods which they see as better discriminators. We did this in two ways – dealing with the group as a whole and dealing with the preference-discrimination relationship on an individual basis. The first is achieved by looking at the correlation of the mean responses across the sample (Figure 4) which shows a strong correlation ($r = 0.79$).

[Figure 4 goes here]

The second approach calculates, for each student, a Spearman correlation between their individual responses for preference and those for discrimination and a mean of these correlation coefficients. The mean was significantly above zero ($t(103) = 15.65$, $p < 0.001$) and centred around a strong correlation ($M = 0.51$, $SD = 0.33$).

So, both considered as a homogenous group and as individuals, there is strong evidence that, in contrast to the findings of the general literature, students prefer assessments which are better discriminators, not those giving high marks irrespective of ability.

Qualitative Methods

Given the contrast with the existing literature on student preference and noting that it does not seem to be based on conservatism, we address our final research question: what factors do influence mathematics students' preferences. All students who completed the questionnaire were invited to attend an interview about assessment preferences. Twelve students agreed (seven from Uni1 and five from Uni2) and were interviewed individually with the same interviewer, with interviews lasting an average of 30 minutes. The characteristics of the interview sample did not vary strongly from the questionnaire sample except in being more heavily skewed towards year 2 students. The responses to the questions also showed a good level of agreement with the quantitative results.

The interviews were semi-structured: that is, the interviewer had a small set of key issues to explore (if they were not raised spontaneously by the participants) but the direction the conversation took was contingent on the views expressed. For example, all interviews began by asking the students to describe the types of assessment they had experienced and in all interviews students were asked about their ideal pattern of assessment in mathematics. Similarly, while a surprising number of students (10 out of 12) raised the issue of some form of oral assessment spontaneously, in the other two cases the interviewer asked students for their views about this explicitly.

The analysis of the interviews followed the thematic networks approach (Attride-Stirling, 2001). They were audio taped and fully transcribed and a coding framework devised inductively: both authors independently coded transcripts sequentially to uncover abstract themes, with their theme lists compared, refined and agreed after each

round of coding until the set of themes appeared saturated and no new ones emerged. Those then formed the networks of organising themes described below.

Results

Three core themes emerged from the analysis of the interviews which appear to underlie the students' patterns of preference for assessment methods. These are *fairness of assessment*; *assessment of abilities* and *the assessment diet.*

Fairness of assessment

Fairness of assessment emerged strongly in every interview. The students gave sophisticated interpretations of fairness which went beyond the simple idea of cheating. One common interpretation was that a fair assessment method is one that assesses students' individual cognitive processes. So, if the role of assessment is to test individual cognitive processes then any assessment that can be completed in groups or with access to external material may be deemed to be unfair:

> If you Google it [*a coursework question*], it'll come up, so you can basically cheat if that's used as an assessment method. (Tom*)

The assessment method considered by most of the students interviewed to be fairest in assessing individual cognitive skills was the closed book examination; it was seen to avoid most of the likely areas of unfairness inherent in other methods:

> I think it's [*closed book examination*] probably the fairest way of doing it because yeah, basically any other form of assessment you can… is open to plagiarism and

* Student responses are tagged with pseudonyms. Those with initial 'T' are from Uni1 and those with initial 'S' are from Uni2

> is open to people working together and stuff like that, which isn't necessarily wrong, but I think if you're going to base the results of an individual's degree on it, it needs to be a very individual form of assessment. (Trevor)

Fairness was not only discussed in terms of assessing one's own cognitive processes, but also was related to the issue of discrimination. Some students considered fairness to require a direct link between the demonstration of their knowledge and the marks, in which luck should play no part: so, for example, multiple choice questions were seen as unfair:

> With a written answer you see every step, with a multiple choice they could have just ticked that because they didn't know what the answer was, so they just chose 'b' just randomly. (Susan)

> The problem with multiple choice is it doesn't always show your reasoning, or how you got to that answer and that's often the important thing in maths to be able to produce rigorous proof that shows why things work and why you understand it. Just coming out with conclusive results, I mean, you could have arrived at that through proper deductive reasoning, or it could just have been a wild guess, or you might have reached that answer by accident, but your assumptions were faulty. (Tania)

Interestingly, some students felt that gaining marks without 'playing the game' to be unfair; particularly when some students were perceived to have obtained good marks without the appropriate amount or form of engagement and effort:

> It's just, for me I enjoy doing coursework, but the problem is for some people who don't go to lectures, it does seem a bit unfair that they can find all the notes [*on the university's Virtual Learning Environment*], they can get 90% on a piece of coursework without actually knowing anything about the subject. (Stewart)

Others go further and suggest that effort should be acknowledged directly in the assessment system:

> Perhaps, I'm not really sure how it would work, but it seems like a good idea to assess whether someone's actually putting effort in. (Sarah)

A small number of students did question the fairness of closed book exam. Trisha for example argued that an unfairness may arise from the differential opportunity to demonstrate ability that examinations can give to individuals:

> I think traditional assessment [*closed book exam*] isn't inappropriate, […] but it doesn't like, if some people are just not very good in that sort of environment and then it's not very fair, like if you don't use a variety of things. (Trisha)

A small number of others saw an unfairness in closed book examinations coming from their reliance on apparently less relevant skills, like memory:

> …but then when they look at your grades, they'd say your grades are low, you couldn't have done that very well. Well no, because that's a bad memory of mine. I haven't got the ability to sit there and learn things off by heart, I have got the ability to use the things I've got there and apply them properly. (Steve)

However, those concerns apart, it appears that the preference for traditional assessments is based in their being perceived as fair, in the sense used by the students: individual, with little opportunity for overt cheating, requiring direct expression of knowledge and hence having a high level of discrimination, allowing little room for luck and normally requiring the expenditure of appropriate effort.

Assessment of abilities

The results of the survey had shown a strong link between students' preferences and their perceptions of the power of assessment methods to discriminate between stronger and weaker mathematicians. Students in the interviews also emphasised this link, but they expressed different perceptions of what abilities and skills are important to be a

successful mathematician and, thus, what skills should form the basis of that discrimination.

Students tended to believe that the process of doing mathematics and of solving questions should be assessed alongside the product and the correctness (or otherwise) of the answer. We have seen an example of this in the earlier discussion on the fairness of multiple choice tests. Closed book examinations were perceived to assess both understanding and memory, but some students saw more biases towards memory because of the lack of access to external material. When comparing closed book exam with project work, some indicated how projects can help assess mathematical thinking:

> I think the [*closed book*] exams here specifically tend to be pretty well tailored. I mean they do seem to cover the whole subject and I'd say they were a pretty good indication of how good you are at understanding the subject, although a lot of the time it's memory tests, since it's closed book. I mean for example when I work on the project, I imagine I'm going to have some of the formulae there to remember, universal constants that I can't remember, that kind of thing with me, as long as I know how to manipulate them I would have thought that was more important thing. I understand you can't have all that sort of thing available to you, so I think it's about the best method you can have really. (Tania)

For Tania, being able to "manipulate" facts and formulae was more important in mathematics than remembering them, and other assessment methods, such as projects, also appear as valid ways of assessing this ability to manipulate without the need to rely on memory. However, for many other students, closed book examinations give the marker transparent access to their understanding provided the questions are well designed:

> Well, if they choose the right questions, like help them [*the lecturers*] understand, they can understand how much you understand the actual course because if you

> don't understand Maths, it's easy to see. Like you're just writing down things, but when you actually understand it, you can explain it better. So in an exam if they give you the opportunity to explain yourself, or do a proof or something, then… (Susan)

This sense of the importance of understanding mathematics over being able to perform long calculations and remembering formulae was shared by many and was seen as best assessed through the closed book examination:

> I assume if they're doing really well in the exam […] if you're going to get a really high mark, it's being able to really understand it, because they could throw any question at you and you have to be able to apply the knowledge to that question, […] So I think if someone's doing really well in maths exams, they're actually just got really, really good understanding. (Tina)

However, the emphasis on understanding was not shared by all: a small number of students saw a more substantial role for memory and calculations in mathematics, but again, this was best assessed by the closed book examination:

> It's literally learning facts and giving like stating definitions, getting the right answer, with maths exams it's very right or wrong, there's no kind of like question which are the right answer, you've either got it right, or you haven't. But then again, that's the way that maths is, it's that kind of subject, obviously it's got to be like that for the majority of it it's got to be like that. (Sarah)

In addition to understanding and memory, a number of students were concerned that problem solving abilities were recognised in assessments. For some, this was still achievable through the closed book system, while others felt alternative methods could be used:

> But then you'd also want to have some more extended things where you're given a problem of a type that you haven't seen before and you're expected to apply what

> you know and sort of fresh ideas to solve that. I think really examination is the best way of doing that, but maybe that doesn't work for all students. (Trevor)

> I think they [*the lecturers*] should, if they're going to assess you, they should assess … well if you can get the answer by using any method possible really, that's what a mathematician should do. They should problem solve and they solve their own problems, like my problem solving of not being a good memory, is by using resources and using the library and using books to get my answer. (Steve)

So the students in our interviews explored a number of different ways in which assessments might be better discriminators between stronger and weaker mathematicians, depending on what they felt were the bases for high ability in mathematics. Different students had conflicting ideas including memory vs. understanding, calculating vs. problem solving, process vs. product. However, in most cases, traditional methods were seen as a good way of assessing those abilities, with other methods having roles in some niches, just not having the general applicability of the traditional ones.

The assessment diet

The Assessment Preferences Inventory, used in the quantitative part of our study as well as much existing literature, asks students to consider different assessment methods in isolation from each other. In reality, the students encounter a mixture of different methods in greater or lesser portions across modules, years and the whole degree programme. Interestingly, in the interviews, discussions about the balance between assessment methods showed a more complex pattern than the questionnaire responses could show.

All the students talked about the importance of the closed book examinations, and nobody suggested removing them from the diet. However, every student in the interviews said they would welcome more variety.

> I would probably take an approach which used a variety of different systems and I would probably do it in a more equal manner to how it is done here and I imagine at most other universities. I would probably have a greater presentation based element. I would probably have some coursework stuff other than the project in the final year. (Ted)

Oral examinations play no role at all in the current assessment diet in either mathematics department. So we were surprised that most students suggested, unprompted, the introduction some variant of this type of assessment, in the form of project presentations followed by a question and answer session, in the form of an interview or in the form of assessment of interaction during tutorials:

> They'd give you a problem on the board and say not how to solve it, but how you would you go about solving it. That looks at how you think about things and how you would start off going at things. (Steve)

Such oral assessment, according to many students, could take the place of a coursework component, or even replace some of the weight that the closed book examination carries.

In contrast to the concerns about fairness outlined earlier, one other feature of oral assessment methods appreciated by some students is the potential to tailor the questions to the student and the opportunity to overcome initial problems:

> … if you have the teacher there who can just give you one relevant piece of information, if you're a bit stuck, because if there's one thing that you can't remember and you then mess up the whole, you can't do the whole problem.
>
> (Teresa)

Some students discussed variants of the oral examination which could be included in the assessment diet:

> I'd like to see some kind of way for them [*the lecturers, to assess you*] either with written pieces, or through actually talking to you. Maybe a meeting with an advisor, just a short meeting to discuss and ask questions and they could give you a mark, for like a small percent of the module, maybe only like five or ten percent, just for them to see how much you understand. (Sarah)

Our participants also suggested introducing more projects in their assessment diet:

> There would be a part of an exam and I think, like the exam does work and has been proven so it's been used so much, but I think you could have part of an exam, but then I think it's like doing a bit of a variety like get them to do a bit of everything like writing a report on a topic, then maybe do a bit of a presentation of something they've been taught before and they've gone to research it a bit more….
>
> (Trisha)

Other suggestions for a varied assessment diet include increasing the amount of credit for coursework and the introduction of more open book exams, though there was some difference of opinion about this latter method of assessment. Some, like Tania, thought that open book exams couldn't be used to assess mathematics as

> … you wouldn't be able to copy proofs verbatim, but there again in maths there's a set way of writing things, so it might, I mean it might not always be clear whether you've copied bits of it. I'm not entirely sure. I'm not sure if that would reflect someone's understanding of it, or whether it would just be part from the books, but it's a difficult one that. Maths isn't really the sort of thing you put into your own words, like an arts, or social science. (Tania)

Others thought that open book exams could be used successfully across topics:

> There'd be less pressure on learning the specifics of what a theorem says and more about how you can apply it and how to prove it and things like that. Because you could write down on your sheet what the theorem exactly says, so you wouldn't have to learn it and then you could just learn how to prove it and how to use it. (Stewart)

As part of a balanced assessment diet some students suggested that assessment should be tailored to different topics within the curriculum. The most commonly cited example was statistics, where some students saw an opportunity for open book exams.

> …last year's statistics exam was open book and I think the reason for that is that maybe sometimes it's not as much remembering all the formulae, it's actually being able to use them, because obviously if you're going to be a statistician then you want to be able to use the formulae, but you're never going to be required to remember them off by heart. (Tom)

So, despite the stark finding from the questionnaire, the interviews revealed a complexity about the mix of assessments. The trend was clearly to preserve the dominance of the traditional assessment methods, but there was a notable emphasis on introducing more variety of assessment, including some of the methods which had both been ranked poorly for preference and were seen as weak discriminators of ability in the survey.

Discussion

The aims of our study were to see whether mathematics students' assessment preferences followed the general trend in the literature in being less favourable towards traditional methods and to uncover the factors which might underpin those preferences.

The quantitative part of our mixed methods study suggests that, with the mathematics students from these research intensive universities at least, the preferences are in contrast to those reported in the general literature. Students tend to prefer traditional assessment methods over more innovative ones and rank multiple choice tests amongst their least preferred options. In our sample, exams were not "seen as a regrettable or second-rate situation" (Sambell et al. 1997, p357) nor were presentations or projects preferable or seen as more exacting. Moreover, there appeared to be no preference for assessment methods seen as superficial or bringing easy marks: the mathematics students link preferred assessment methods with those they see as the best at discriminating between people on the basis of ability; this, again, contrasts with the message from the existing assessment preference literature (Traub and MacRury 1990; Scouller 1998). They also do not appear to base those preferences on an unreflecting conservatism (focussing on assessment methods they know and are successful with).

Indeed, they had thoughtful reasons for the preferences. Three main factors emerged which seem to underpin them: a concern for fairness, the abilities being assessed and the proportion of a given method in a varied assessment diet.

The strong correlation between preference and discrimination from the questionnaire suggests that these students are most interested in being accurately assessed according to their abilities and, as the interviews imply, have deep-seated concerns about the unfairness of achieving marks through shared endeavour, with luck, without effort or without ability. While the students in this study shared a more complex sense of fairness with those in Sambell et al. (1997) - including notions of individual vs group work, access to external resources and the avoidance of luck - the mathematics students

differed by seeing closed book examinations as generally fairer than newer forms of assessment because they avoid these issues.

The interviews deepened our understanding of the link between preference and discrimination seen in the survey: students clearly maintained the importance of that link, even when they had quite different views of whether the basis for discriminating should be memory, understanding or problem solving ability. However, the interviews also uncovered a complexity which a reliance on the API alone masks: The students were concerned for the assessment diet as a whole, not simply the dominant method. All agreed that closed book examinations should play the majority role, but they were concerned to see a richer diet than they currently had. In particular and most surprisingly, a large number of the students interviewed spontaneously raised the possibility of a role for oral examinations, despite this method being ranked relatively poorly for preference and in the second group for discrimination. It may be that the low preference was related to concerns about the stress that oral examinations might cause amongst people who have not experienced them: Iannone and Simpson (2012) found that once mathematics students had experienced oral examinations, they saw considerable value in them for gauging understanding, albeit that they still saw closed book examinations as the "gold standard".

However, we argue that the contrast between our findings and the consistent message from existing, general assessment preference literature highlights the importance of subject matter. As with Joughin (2010) we believe the context of the subject material must be taken into account when interpreting results, and general rules should not be drawn from studies that reflect only a fraction of the academic disciplines. The

mathematics students in our research clearly tie some notions (such as fairness) to specific mathematical cognitive processes and even note themselves that this may not apply to other subjects (recall Tania's comment "maths isn't really the sort of thing you put into your own words, like an arts, or a social science").

Our own study itself is quite specific - while we set out to find if students of hard-pure subjects have similar or different views to others, we have looked at only one area, albeit the subject considered central to that quadrant of Biglan's classification. More work needs to be done to see if the distinct voice of mathematics students resonates more widely in hard-pure subjects.

Indeed, there could be many other ways of accounting for our results. For example, there may be issues with adapting the API to mathematics (though such concerns would apply to many other uses of the instrument), but the level of consistency between the results from the API and the interviews suggests good validity. It may be that mathematics, particularly at this level, attracts students with specific sets of preferences; though again such an argument would apply to all subjects – each discipline inevitably attracts students with particular skills and biases. It may be that the nature of mathematical thinking and performance is less visible than, for example, the humanities and social sciences and thus the methods for assessing performance are necessarily restricted.

None of these accounts, though, detracts from the central message we read in our results. Our findings strike a note of caution for practitioners interpreting research findings in their subject and institutional contexts – there is clearly need to take

disciplinary differences into account, whether those differences are in the type of student, the nature of knowledge, the form of performance or the pedagogical culture. The scholarly literature in university level mathematics which advocates poster presentations (Houston 2001) or multiple choice questions (Ramesh 2009) needs to be seen in the context of this research which suggests they are considered weak discriminators and are little preferred by students. Of course, our focus on students' preference comes not only from consideration of students' views in an increasingly marketised university provision, but mostly from consideration of students' motivation and engagement.

Moreover, our findings also suggest that we must be careful in interpreting research based on questionnaires like the API alone. Students do see a role for posters and projects in a rich and varied assessment diet, albeit one which is subservient to closed book examinations. A deeper understanding of the students' view of the complex diet of assessment may help assessors match methods to intentions in a way which is understandable for students.

Acknowledgment

We would like to thank the Maths, Stats & OR Network of the Higher Education Academy in the UK for funding the research reported in this paper.

Notes on Contributors:

Paola Iannone is senior lecturer in the School of Education and Lifelong Learning at the University of East Anglia. Her research has primarily focused on teaching and learning mathematics at university level, with particular attention to proof and proof production, and assessment of mathematics at university.

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

References

Amin T. T., Kaliyadan F. and Al-Muhaidib N. S. 2011. Medical students' assessment preferences at King Faisal University, Saudi Arabia. *Advances in Medical Education and Practice* 2: 95-103.

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

Becher,T. 1989. *Academic Tribes and Territories*. Milton Keanes, Open University Press.

Becher, T. 1994. The significance of disciplinary differences. *Studies in Higher Education.* 19(2): 151-161.

Berry, J. and Houston K. 1995. Students using posters as a means of communication and assessment. *Educational Studies in Mathematics* 29: 21-27.

Biglan, A. 1973. The characteristics of subject matter in different academic areas. *Journal of Applied Psychology* 57(3): 195-203.

Birenbaum, M. 1994. Towards adaptive assessment – the students' angle. *Studies in Educational Evaluation* 20: 239-255.

Birenbaum, M. 2007. Assessment and instruction preferences and their relationship with test anxiety and learning strategies. *Higher Education* 53(6): 749-768.

Boud, D., and Falchikov, N. 2007. *Rethinking Assessment in Higher Education Learning for the Longer Term*. Ney York: Routledge.

Burton L. and Haines C. 1997. Innovation in teaching and assessing mathematics at university level. *Teaching in Higher Education* 2: 273-294.

Dillman D. A. and Bowker, D. K. 2011. The Web Questionnaire Challenge to Survey Methodologists. In *Online,* eds. B. Batinic, Reips, U., Bosnjak M. and Werner A.

Gibbs, G. 2006. Why assessment is changing. In *Innovative Assessment in Higher Education*. Eds. C. Bryan and K. Clegg, London: Routledge.

Gielen, S., Dochy, F., and Dierick, S. 2003. Evaluating the consequential validity of new modes of assessment: The influence of assessment on learning, including pre-, post-, and true assessment effects. In *Optimising new modes of assessment: In search of qualities and standards.* Eds M. Segers, F. Dochy, and E. Cascallar 37–54. Dordrecht: Kluwer Academic Publishers.

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

Haines, C. and Crouch R. 2005. Applying mathematics: making multiple-choice questions work. *Teaching Mathematics and Its Applications* 24(2-3): 107-113.

Harlen, W. and Crick. R.D. 2003. Testing and motivation for learning. *Assessment in Education: Principles, Policy and Practice* 10: 169-207.

Houston, K. 2001. Assessing undergraduate mathematics students. In: *The Teaching and Learning of Mathematics at University Level: An ICMI Study*. Ed D. Holton 407-422. Dordrecht: Kluwer Academic Publishers.

Iannone, P. and Simpson, A. 2011. The summative assessment diet: how we assess in mathematics degrees. *Teaching Mathematics and its Applications*. 30(4), 186-196.

Iannone, P. and Simpson, A. 2012. Oral Assessment in mathematics: implementation and outcomes. *Teaching Mathematics and its Applications,* 31(4). 179-190.

Iannone, P. and Simpson, A. 2013. Students' perceptions of assessment in undergraduate mathematics. *Research in Mathematics Education Journal.* 15(1) 17-32.

Johnson, B and Turner, L.A. 2003. Data collection strategies in mixed methods research. In the *Handbook of Mixed Methods in Social & Behavioral Research*. Ed. A. Tashakkori, and C. Teddlie 297-319. Sage.

Joughin, G. 2010. The hidden curriculum revisited: a critical review of research into the influence of summative assessment on learning. *Assessment and Evaluation in Higher Education* 35(3): 335–345.

Kniveton, B. H. 1996. Student perceptions of assessment methods. *Assessment and Evaluation in Higher Education* 21(3): 229–238.

Kolb, D.A. 1981. Learning styles and disciplinary differences, In *The Modern American College*. Ed A. Chickering 232-255. San Francisco, CA, Jossey Bass.

Levesley, J. 2011. Taking control of the assessment agenda. In *Report of the HE Mathematics Curriculum Summit.* Ed. Rowlett, 21-23. P. York: Higher Education Academy.

LMS 2011. Comments on *Review of the UK Professional Framework for Higher Education, Available at* http://www.lms.ac.uk/sites/lms.ac.uk/files/Mathematics/Policy_repors/2011%20LMS_Response_to_PSF_Consultation.pdf last accessed 8th May 2013.

Povey, H. and Angier C. 2006. The Assessment of Undergraduate Mathematicians: Recrafting Assessment of Learning to Provide Opportunities for Assessment as Learning. *MSOR Connections* 6(4): 43-47.

Ramesh, N. 2009. Some issues on assessment methods and learning in mathematics and statistics. *MSOR Connections* 9(4): 28-31.

Sambell K., McDowell L. and Brown S. 1997. 'But is it fair?': an exploratory study of students' perception of the consequential validity of assessment. *Studies in Educational Evaluation* 23: 349-371.

Sax, L., Gilmartin, S. K. and Bryant, A. N. 2003. Assessing response rates and nonresponsive bias in web and paper surveys. *Research in Higher Education,* 44(4): 409- 432.

Schoenfeld, A.H. 1989. Explorations of students' mathematical beliefs and behavior. *Journal for Research in Mathematics Education* 20: 338-355.

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

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

Traub, R. E., and MacRury, K. 1990. Multiple choice vs. free response in the testing of scholastic achievement. In *Tests und Trends 8: Jahrbuch der Pädagogischen Diagnostik*. Eds K. Ingenkamp and R. S. Jager. 128–159. Weinheim und Basel: Beltz.

Wareing, S. 2009. Disciplines, discourse and Orientalism: the implications for postgraduate certificates in learning and teaching in higher education, *Studies in Higher Education,* 19: 917-928.

Zeidner, M. 1987. Essay versus multiple-choice type classroom exams: the student's perspective. *Journal of Educational Research* 80(6): 352–358.

APPENDIX

A: Preference

To what extent would you want your achievements in the course to be assessed by each of the following methods:

	Hardly at all	To a small extent	To a medium extent	To a large extent	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 examples 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 over the course of 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 over the course of a long period, such as a term or two)					

B: Ability

For each of these assessment methods decide how good it is at distinguishing those who are good mathematicians from those who are poor mathematicians.

	Very poor				Very good
	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 examples 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 over the course of 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 over the course of a long period, such as a term or two)					