

## Durham Research Online

---

### Deposited in DRO:

12 October 2015

### Version of attached file:

Accepted Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Iannone, P. and Simpson, A. (2017) 'University students' perceptions of summative assessment : the role of context.', *Journal of further and higher education.*, 41 (6). pp. 785-801.

### Further information on publisher's website:

<https://doi.org/10.1080/0309877X.2016.1177172>

### Publisher's copyright statement:

This is an Accepted Manuscript of an article published by Taylor Francis Group in *Journal of Further and Higher Education* on 17/05/2016, available online at: <http://www.tandfonline.com/10.1080/0309877X.2016.1177172>.

### Additional information:

## Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

## **University students' perceptions of summative assessment: the role of context**

We report on a mixed method study which compared students' perceptions of summative assessment across two distinct disciplines – education and mathematics, at two research-intensive institutions in the UK. The disciplines chosen represent opposing positions in Biglan's classification of academic disciplines, as well as having very different assessment practices. Results suggest that these education students prefer to be assessed by methods they perceive to discriminate for academic abilities. Moreover they perceive the traditional closed book examination as inadequate to assess the capabilities which are key to being successful in their subject, which fits some but not all of the general findings in the literature. However, comparing with an identical study with mathematics students, we find that the perceptions of summative assessment are very different. We account for that difference by suggesting that students' epistemic beliefs play a role in shaping these perceptions and conclude that, in designing summative assessment in higher education, generalised and centralised forces for change need to be tempered by contextual and disciplinary factors.

Keywords: Summative assessment, epistemic beliefs, academic disciplines, education, mathematics.

### **1 Introduction**

We previously reported on a study of mathematics students' assessment preferences which indicated that these were in marked contrast to the common message from most other studies in higher education (Iannone and Simpson 2015). That study suggested that a cause of this disparity lay with problematic sampling bias within that existing literature. Joughin (2010) raised concerns about the problems associated with decontextualising research on assessment and argued that both disciplinary and institutional contexts might have an important impact on assessment practices and students' perceptions of them. For example, an extensive review of previous literature on assessment preference (Struyven, Dochy and Janssens 2005) showed that samples in

Iannone, P. and Simpson, A. (in press) University students' perceptions of summative assessment: the role of context. To appear in the Journal of Further and Higher Education

that literature came from a very restricted range of disciplines. In particular, the voice of students from hard pure sciences (in the sense of Biglan [1973]) had gone unheard and, in general, the issue of how disciplinary context might play a role in assessment preference has not previously been addressed.

In order to explore this issue, we designed a study to replicate the mathematics study with samples from the same two institutions but in a very different discipline - education - which has a demonstrably different approach to assessment and is at the opposite end of mathematics in the pure/applied, hard/soft classification proposed by Biglan (1973). By comparing the two studies, this paper discusses the role which disciplines may play in shaping students' perceptions of assessment, using the differences in assessment between education and mathematics as an example. It reports, first, the key results of the education study and the messages from the analysis of the education students' views, then contrasts those results with the results in the study of mathematics students. We make a case for viewing these contrasts in the context of the nature of the discipline and suggest a link between views of assessment and epistemic belief which, if borne out more widely, should impact on where the control of assessment practices should lie within an institution.

## **2 Background and Literature Review**

Much attention has focused on students' perceptions of assessment and how these impact on students' engagement with their learning. Marton and Saljo (1997) were among the first to highlight a link between what students perceive as the requirements for success in (summative) assessment and their engagement with learning. For example, some research suggests that students perceiving assessment as requiring understanding tend to engage in deep learning (Harlen and Crick 2003). While this link

Iannone, P. and Simpson, A. (in press) University students' perceptions of summative assessment: the role of context. To appear in the Journal of Further and Higher Education

is not straightforward, and changing assessment may not be sufficient to change the way students learn (Gielen, Dochy and Dierick 2003; Gijbels and Dochy 2006), it is clearly important to understand what students think about the summative assessment they experience. Much of the research in this area has focused on the comparison between students' perceptions of assessment methods, focusing for example on multiple choice tests and essays (Scouller 1998; Zeidner 1987). Indeed Birenbaum (2007) raised the issue of the importance of research on students' perceptions of summative assessment and noted how the underlying issues shaping these preferences had not been investigated.

A comprehensive review (Struyven et al. 2005) noted that a common feature of existing literature is that students prefer to be assessed by innovative methods - that is, methods other than the closed book examination - and by assessment methods that allow them to accrue marks with less effort (see also Traub and MacRury [1990]). This review suggests that students perceive traditional assessment as unfair and damaging for learning as it encourages them to engage with the subject at a very procedural level. The same review also noted that students prefer to be assessed by innovative assessment methods (taking innovative to mean non-traditional, especially away from unseen timed exams) in which they recognise authentic value and which they think will prepare them for entering the workplace. Sambell, McDowell and Brown (1997) reported that 'many students perceived traditional assessment tasks as arbitrary and irrelevant' (359) while they believed innovative assessment to be fairer as it appear to measure capabilities relative to their studies and encourages them to engage in deep learning.

Iannone, P. and Simpson, A. (in press) University students' perceptions of summative assessment: the role of context. To appear in the Journal of Further and Higher Education

These studies have contributed to strengthening the calls for the need to reconceptualise summative assessment in higher education and to move away from the overuse of the traditional closed book exam. As Brown (2004) states:

Assessment methods and approaches need to be focused on evidence of achievement rather than the ability to regurgitate information. Inevitably this means a lesser concentration on traditional written assessments, particularly time-constrained unseen exams, and a greater emphasis on assessment instruments that measure not just recall of facts, but also the students' abilities to use the material they have learned in live situations. (2004, 82)

Brown argues that the use of innovative methods of assessment will help the students to engage with deep learning and will prepare them to enter the job market once they graduate. Indeed many authors have categorised assessment methods according to their purpose. In particular, there has been a long standing simple split of assessment into *summative* and *formative* which has shifted more recently into the apparent separation of *assessment of learning* from *assessment for learning*. The latter was described by Black et al. (2004) as

...any assessment for which the first priority in its design and practice is to serve the purpose of promoting pupils' learning. It thus differs from assessment designed primarily to serve the purposes of accountability, or of ranking, or of certifying competence. (p 10)

while the former has a predominant function of certification. Other authors (Birembaum et al. [2006]) place the emphasis on the richness of feedback associated to assessment for learning, the potential to develop students' autonomy and problem solving skills and the relation that such assessment has to real life situations. More recently Medland (2014) suggested that traditional assessment is inadequate and supports the current testing culture while innovative assessment, if supported by the right opportunities for

students to learn about the new methods, can lead to a true culture of assessment for learning.

However, these recommendations seem to propose change to assessment patterns regardless of context: sidelining the nature of the students, the institution in which they are studying, the subject they are studying and the aims of their study. This decontextualised approach to assessment research in higher education has led to criticism. Joughin (2010) noted that research which does not take into account the context may lead to these generalised recommendations which are not supported by data.

Moreover, there is research which does clearly demonstrate contextual influences on students' perceptions of assessment. Furnham and Chamorro-Premuzic (2005) and Furnham, Batey and Martin (2011) investigated the correlation between IQ, personality types and assessment preferences and found that there are strong correlations between some personality traits and preferences for certain assessment methods. Furnham and Chamorro-Premuzic (2005) report for example that extraverts are positively inclined towards oral examinations while conscientiousness is correlated with preference for continuous assessment. Findings such as these suggest that it is not feasible to treat the whole student body as one, and that there are well defined factors which shape assessment preferences. Furnham and Chamorro-Premuzic (2005) conclude:

Preference for examination method also no doubt related to particular academic discipline. Thus, arts, social sciences, and sciences might lend themselves to quite different assessment methods. Science students may favor multiple choice, while art students may prefer essays, and drama students may strongly favor performance based on oral exams (i.e., vivas). Indeed, disciplinary differences in the way

Iannone, P. and Simpson, A. (in press) University students' perceptions of summative assessment: the role of context. To appear in the Journal of Further and Higher Education

students are examined tend to vary across institutions. In this sense, what is taught and how it is taught may be the best predictor of how courses and methods are examined. (2005, 1989)

Indeed the assessment experiences of students across disciplines, at least in the UK, vary significantly. Figure 1 shows the key links between discipline and assessment.

Taken from the Higher Education Statistics Agency's (HESA) data on assessment practices on every degree course delivered in the UK, it summarises the proportion of written examinations across different disciplines. It shows that there are very clear differences between disciplines in their use of traditional written examinations with mathematical sciences and education as very significantly different (Mann-Whitney's  $U=534533$ ,  $p < 0.0001$ ,  $n_{math}=852$ ,  $n_{edu}=661$ ,  $CLES=94.7\%$ ).

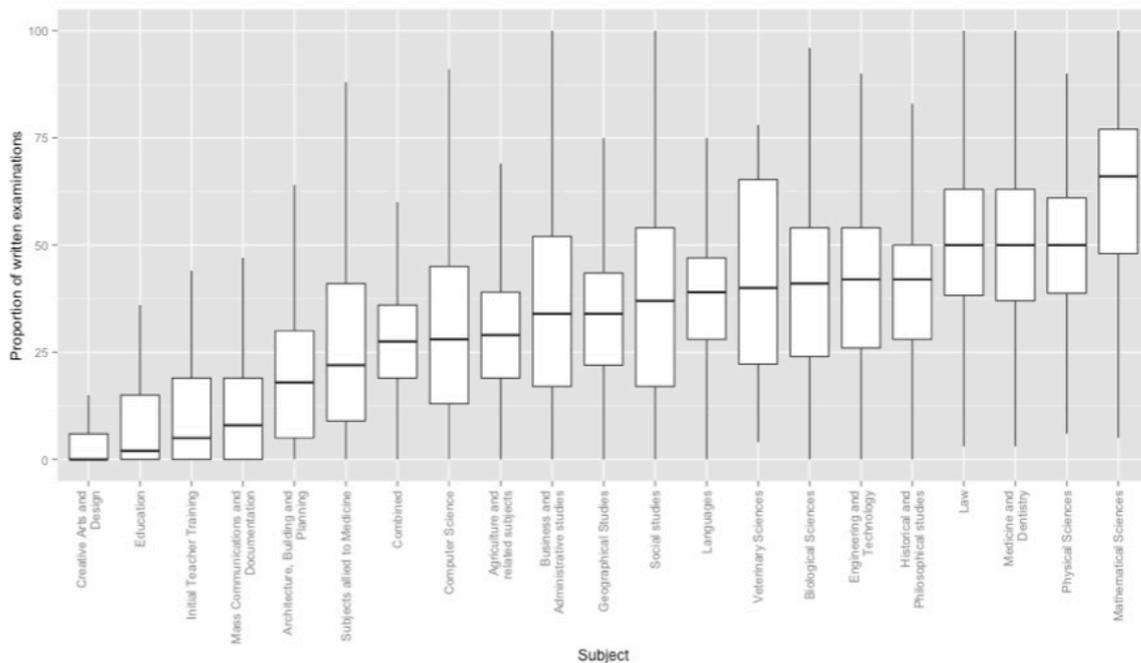


Figure 1: Spread of assessment methods across disciplines in the UK

This provides strong prima facie evidence that the assessment practices across disciplines (for example between mathematics and education) are very different.

Iannone, P. and Simpson, A. (in press) University students' perceptions of summative assessment: the role of context. To appear in the Journal of Further and Higher Education

However, Simpson (2015) has noted that disciplines and institutions have quite different influences on assessment outcomes. Disciplines have virtually no independent impact on the variation of degree scores, while institutions have a relatively large impact on that variation. However, disciplines and institutions have roughly equal impact on the variation in the *spread* of degree scores. This suggests it is important to examine the link between discipline and assessment preference and to do so across quite different disciplines.

Thus we addressed the following research questions:

RQ1: What forms of assessment do education students consider best at discriminating between students of different abilities?

RQ2: To what extent are education students' preferences related to their experience of assessment or to their views of how well methods discriminate?

RQ3: How do education students' perceptions of assessment differ from mathematics students' perceptions?

RQ4: On what factors are students' preferences based and how do these factors compare between education and mathematics?

Note that the first two research questions replicate those in Iannone and Simpson (2015), but are recontextualised to education, while the other two research questions compare the results of the previous mathematics study with the study reported here.

### **3 Methods**

To avoid conflating effects for institution and discipline, we conducted the study with students at the same institutions as Iannone and Simpson (2015), referred to as *Uni1* and *Uni2*, two research-intensive universities in the UK. As a recontextualised replication study, we used the same methods of the previous study except where highlighted below. That is, the study employed a mixed methods design (Johnson and Turner 2003) involving a survey and follow-up semi-structured interviews with a range of participants. The study was reviewed and approved by the Research Ethics Committee of the university where the first author works.

Education as an undergraduate degree programme in the UK has a somewhat complicated structure. While some degrees are professionally oriented (aligning academic study with obtaining “Qualified Teacher Status” providing, effectively, a license to teach) others do not have that extra accreditation focus and concentrate on education as an academic discipline. As with many other disciplines, education can be taken as one half of a “joint” degree (e.g. Education and History). For the purposes of this study, all the participants came from degree programmes (including joint degrees) which did *not* have the professional focus. The participants in this study had experienced a variety of assessment methods including closed and open book exams, essay coursework, multiple choice question tests in exam conditions, group and individual projects and associated presentations, pre-released exams and a dissertation for those students in their final year.

#### ***4.1 Quantitative methods: The survey***

All students enrolled on the chosen education degree programmes at both Uni1 and Uni2 were sent an email containing an invitation to take part in the study and the link to

Iannone, P. and Simpson, A. (in press) University students' perceptions of summative assessment: the role of context. To appear in the Journal of Further and Higher Education

an online survey of assessment preferences. After providing brief biographical information (such as gender and year of study), the survey asked participants to rate each item on a list of assessment methods against two criteria:

- (1) To what extent would you want your achievements in the course to be assessed by the given assessment method?
- (2) For each of these assessment methods, decide how good it is at distinguishing those who are good at academic studies in education from those who are poor at academic studies in education.

For each assessment method and criterion, the participants were given a brief illustrative example and a 5-point Likert scale (from 'Hardly at all' to 'Almost exclusively') against which to rate it. The phrasing of both of these questions was verified during earlier pilot interviews (see Iannone and Simpson [2013])

In keeping with the mathematics study, criterion 1 was taken as an indicator of preference and criterion 2 as an indicator of the students' view of the assessment method's ability to discriminate on the grounds of ability. While we accept that students may not be well equipped to decide what attributes make someone 'good' in a discipline, as with the previous study we were interested in the students' *perceptions* of which assessment methods best discriminate between students on the basis of ability.

The list of assessment methods was adapted from the Assessment Preference Inventory (Birenbaum 1994; Gijbels and Dochy 2006). The full survey is given in appendix A, and it replicates the previous study in all but one aspect. In mathematics, the phrase 'weekly example sheets' would commonly refer to the sets of questions handed out for

completion as consolidation, practice or expansion on the material covered in the lectures. This phrase would have little meaning for education students, but they would have experience of being given work to complete each week, so this item was replaced with 'weekly coursework'.

#### 4.1.1 Results

Table 1 gives a summary of the participants' demographics. We received 57 responses, but 13 participants omitted some demographic data and some other data and a further 3 omitted some non-demographic data and the analyses requiring those data has been adjusted for these omissions.

	Uni 1		Uni2		N(E)
	M	F	M	F	
Year 1	1	7	2	8	18
Year 2	0	1	1	7	9
Year 3	0	5	2	10	17
N(E)	1	13	5	25	44

Table 1: Demographic of the education participants

Note that, in general, undergraduate education degrees have few male students in the UK. Statistics from the Department for Education in the UK (DfE 2013) show that around 15% of students on a first degree in education are male which is closely reflected in the proportion (14%) of male students in our sample. Perhaps because of the small number of male respondents, no gender differences reached significance and this aspect was not investigated further (though we should note that in mathematics, where the male population far outweighs the female, and where there was a larger sample, we

also noted no significant gender differences). However, the lack of within group gender differences in this study should not be taken as meaning there would not be between group differences, or within group differences in different studies.

Considering students' preferences for assessment methods first, a three-way analysis of variance for assessment method, year group and university showed that there were main effects for assessment method ( $F(7, 313) = 11.9, p < 0.001$ ) and year group ( $F(2, 312) = 10.03, p < 0.001$ ), but not for university ( $F(1,312) = 0.23, p = 0.625$ ). The only significant interaction was assessment method by year ( $F(14, 312) = 2.14, p = 0.009$ ).

Table 2 gives the mean and standard deviation for the responses to the question about preference and the results of a Tukey HSD test for significance of the differences in means. From Table 2 we see that projects were significantly preferred to all other methods (all  $ps < 0.05$ ) and oral exams were significantly less preferred than open book exams ( $p = 0.003$ ), dissertations ( $p < 0.001$ ) and projects ( $p < 0.001$ ).

	Project	Dissertation	Open Book	Closed Book	Multiple Choice	Presentations	Weekly Coursework	Oral exams
<i>M</i>	3.66	3.00	2.81	2.57	2.49	2.47	2.36	1.94
<i>SD</i>	0.79	1.01	1.09	1.20	1.28	1.24	1.27	1.16
Project	--	--	--	--	--	--	--	--
Dissertations	0.047	--	--	--	--	--	--	--
Open Book	0.002	0.986	--	--	--	--	--	--
Closed Book	<0.001	0.509	0.964	--	--	--	--	--
Multiple Choice	<0.001	0.258	0.822	0.999	--	--	--	--
Presentation	<0.001	0.219	0.778	0.999	1.000	--	--	--
Weekly Coursework	<0.001	0.067	0.453	0.976	0.999	0.999	--	--
Oral Exams	<0.001	<0.001	0.003	0.083	0.219	0.259	0.566	--

Table 2: Analysis of education students' preferences responses

Figure 2 shows the interaction plot of mean preference for assessment method across first and third years and suggests that the interaction is the result of a general trend with

third year students to rate assessment methods lower for preference, except for closed book exams, which have a marked increase in preference.

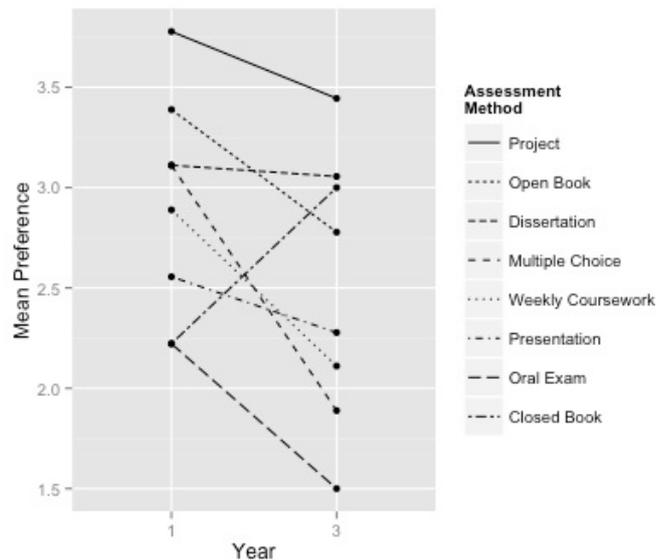


Figure 2: Education students' preference - interaction between years

Regarding students' views of the potential of different assessment methods to discriminate according to academic ability, a three way analysis of variance was conducted for assessment method, by year group and university. In this case, there were no significant interaction effects, just main effects for assessment method ( $F(7,312) = 12.4, p < 0.001$ ) and year ( $F(2, 312) = 4.7, p = 0.009$ ). Table 3 gives the mean and standard deviation for the responses concerning the discrimination of different assessment methods and the results of a Tukey HSD test for significance of the differences in means. Projects and dissertations were seen as significantly more discriminating than presentations, oral exams and multiple choice exams (all  $ps < 0.05$ ) and closed book exams were significantly more discriminating than oral exams ( $p = 0.006$ ) and multiple choice exams ( $p < 0.001$ ).

	Project	Dissertation	Closed Book	Weekly Coursework	Open Book	Presentations	Oral Exams	Multiple Choice
<i>M</i>	3.60	3.56	3.38	3.26	3.11	2.88	2.66	1.96
<i>SD</i>	0.97	1.04	1.11	1.16	0.96	1.07	1.22	0.99
Project	--	--	--	--	--	--	--	--
Dissertations	0.999	--	--	--	--	--	--	--
Closed Book	0.966	0.988	--	--	--	--	--	--
Weekly Coursework	0.731	0.834	0.999	--	--	--	--	--
Open Book	0.278	0.387	0.905	0.996	--	--	--	--
Presentation	0.014	0.028	0.255	0.611	0.957	--	--	--
Oral Exams	<0.001	<0.001	0.014	0.075	0.369	0.962	--	--
Multiple Choice	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.020	--

Table 3: Analysis of education students' discrimination responses

Figure 3 shows these results graphically, with the mean and standard error bars for preference on the *x*-axis and for discrimination on the *y*-axis. This picture suggests that students tend to prefer assessment methods which discriminate for academic ability, though with one outlier (that students tend to disproportionately prefer multiple choice exams given how little they think it discriminates for ability). An analysis of individual responses confirms this: the mean correlation between preference and discrimination is 0.54 and this is significantly above zero ( $t(46) = 9.8, p < 0.001$ ).

[Fig 3 goes here]

Summarising the findings: these education students generally prefer to be assessed by assessment methods they perceive to discriminate for ability in their subject. They do not rate preference for closed book examination highly, preferring to be assessed by

Iannone, P. and Simpson, A. (in press) University students' perceptions of summative assessment: the role of context. To appear in the Journal of Further and Higher Education

project and dissertation which they also perceive to be methods which are better at discriminating for academic ability in education.

#### ***4.2 Qualitative methods: the interviews***

All students who completed the survey were invited to attend an interview. Seven students were interviewed: two female students at Uni1 and four female and one male at Uni2. The two cohorts at the two sites were small which partly accounts for the small interview sample. There is some evidence however (Guest et al. 2006) that in a homogeneous sample, at least 80% of the main themes are likely to emerge from the first 6 interviews. Moreover, the proportion interviewed reflects the proportion interviewed for the mathematics study. The gender balance of the interview sample also reflects the gender balance of the two degree courses.

The aim of this part of the study was to answer our second research question as we set out to elicit what factors influence education students' perceptions of summative assessment.

As with the quantitative part of the study, the interview procedure was designed to replicate the previous study as closely as possible, but noting the changed disciplinary context. The interviews were semi structured: the first author was the interviewer in all cases and had a small list of key themes and initial questions prepared in advance (see appendix 2). However the discussion was allowed to take contingent directions according to participants' responses. The interviews lasted up to 45 minutes and were audio recorded and fully transcribed. The analysis of the interviews followed a thematic network approach (Attride-Stirling 2001). Each transcript was analysed sequentially by one author until a stable network was developed. This network was then used by the second author to code samples to check the confirmability and credibility of the scheme

(in the sense of Lincoln and Guba [1985]). The thematic network emerging from the analysis is shown in figure 4.

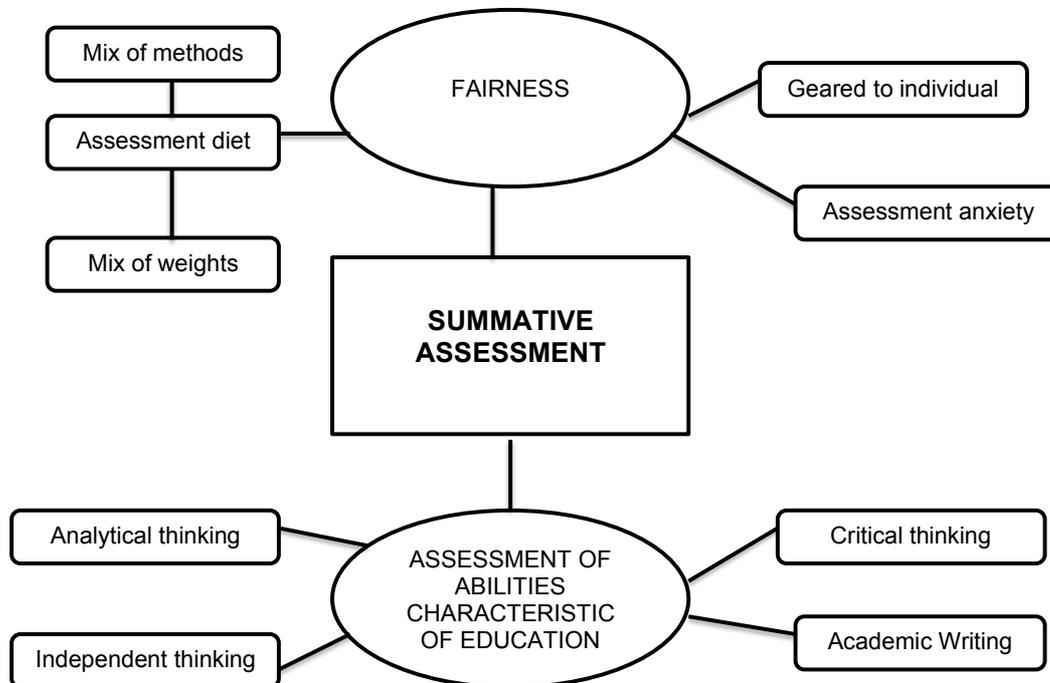


Figure 4: Thematic network

#### 4.2.1 Results

The thematic network shows two overarching themes: *fairness of assessment* and *assessment of abilities*, each comprising several sub-themes.

##### *Fairness of assessment - the assessment diet.*

As suggested by existing literature on summative assessment, fairness tended to dominate the students' focus. Fairness for the participants seemed to be defined in terms of the potential of the assessment for a module to account for individual differences.

Students believed that some perform at their best in exam conditions, while others need time and concentration to produce work at their own pace and, for the sake of fairness,

the assessment of a module should reflect this mix. The final mark should be obtained by the combined outcomes of several different assessment methods, as Elizabeth<sup>1</sup> explains here:

Yeah, I think people have different qualities, some people can remember things very well, so when they go into an exam they can do it very well. Some people get nerves, and they can't handle an exam. Other people don't find it quite so easy to write an essay or get it so that it flows very well, but they can do very well under exam conditions. I think that whole 50-50 kind of works for everybody. I would think that was fair anyway. (Elizabeth – Year 3)

Or as Charlotte remarks:

I think it's fair to have like a different range 'cos obviously some people would be better, some people would prefer this to coursework or exams so it gives people chance to . . . 'Cos its not like 100% of that module, so there's still like coursework for that module as well. So, it's fair I think. (Charlotte – Year 1)

Moreover, employing a variety of assessment methods away from the traditional closed book examination was also seen as 'fair' because it helps those students who might be anxious about being assessed in exam conditions

I think a 50-50 split is good [*between exams and another assessment method typically coursework*]. I quite like the way my course has been weighted. Yeah, I wouldn't like all exams, because I think there is too much pressure, and the amount of work you put into one exam, and if you are having a bit of an off day, then you have just one hour to show what you can do, I think that's... I think that's quite stressful. (Caroline – Year 3)

---

<sup>1</sup> All student names are pseudonyms, we also have indicated the respondent's year of study

Anxiety related to high-stakes summative assessment and its impact on students' performance was one of the most common themes in our interviews:

And the reason I have an issue with that [*exams*] is that . . . I know that I can work hard and I . . . I work hard to understand things and to go into a room and maybe have your nerves get the better of you and you can't remember anything doesn't really show off what I ...what I've been studying so I don't think it ...I just find it . . . I think it is an unfair method of . . . if it's a compulsory, I just feel it doesn't take into account different people's knowledge. It doesn't assess their knowledge in a fair way. I think its one thing to be able to memorise something, and it's another thing to actually apply what you've learnt. (Jane – Year 2)

One of the suggestions that students made in order to overcome the unfairness due to anxiety was to vary the assessment diet both at individual module level and at degree programme level. This they believed can cater for students' individual strengths, and our participants said this is highly desirable.

And then if you have lots of little things, that's better than just one exam because that um ...if you have an off day and it happens to be on the day of your exam, then you are completely screwed really. If it is sort of 20% every term or whatever, then you can have off days but then still thrive sort of thing. (Mary – Y1)

The possibility to produce assessment in the student's own time was also seen as increasing fairness, as in the case of coursework:

But I do think coursework it's more fair [*than exams*] because you can treat it in an exam situation or you can take it slowly, so I think coursework is just better for everyone. (Mary – Year 1)

Moreover, summative assessment, in order to be fair, should assess the students' own ability and progress. For this reason, and despite all the emphasis that is put on group work in education degrees in the UK (Reynolds and Michael 2013), assessment

Iannone, P. and Simpson, A. (in press) University students' perceptions of summative assessment: the role of context. To appear in the Journal of Further and Higher Education

methods asking students to work in groups were not considered fair for many as the student's mark was seen as impacted by the performance of others:

I like group work for formative assessment but I don't like relying on it for my own grade. I think it's . . . if you get a good group it's perfect, but a lot. . . too much depends on relying on other people and I'd rather have it relying on myself, yeah . . . (Caroline – Year 3)

Or as Charlotte explains:

But I do think that in some groups . . . just depends on who you are working with. Because I know some people are happy to let other people get on and do all the work. And then that's the only thing I probably wouldn't like to do a group work presentation where we get marked as a group, I prefer getting marked as individuals, but then sometimes I do think like even without meaning to sometimes you do get people who just like shine a lot better than you, so . . . don't know . . . it just depends who you are working with really, I guess. (Charlotte – Year 1)

We note here that these comments regarding group work, although do not refer to a single assessment method as such, were very common and addressed fairness of assessment in a more general way.

### *Assessment of abilities*

From the analysis of the Assessment Preferences Inventory we have seen that education students prefer to be assessed by methods they perceive to be good discriminators of academic ability. The link between assessment and ability was also a key overarching theme in the organising network diagram and, in focusing on this theme, we saw students discussing which abilities they see as important and how those particular abilities are best assessed. Students saw education as composed of many, often very different subjects with their own distinct characteristics:

Iannone, P. and Simpson, A. (in press) University students' perceptions of summative assessment: the role of context. To appear in the Journal of Further and Higher Education

For education you need a whole range of different subjects and different things that together make you an overall, . . . I don't think it is about that individual one essay, or one exam. It's got to be the combination of the whole. (Elizabeth – Year 3)

Success as a student of education is being able to master a variety of competencies such as analytical thinking and critical reading. Elizabeth again:

I think, [*with success in coursework essay*] you're probably better at critically analysing a wide range of medium, because you're doing books, other journals, you may even go and use newspaper articles. You're drawing on a massive field to kind of come up with that concise critical argument of what it is that you're discussing in the essay. . . (Elizabeth – Year 3)

and

. . . in an exam you know what the examiner wants in a way, you kind of know I need to put this, this and that in. While in an essay whilst there is an element of that, it is kind of more . . . you can take it in whatever direction you want to take it, so I think in a way an essay is a little bit more kind of independent academically than an exam. [ . . . ] Because I think education is such a kind of . . . Everyone's got a different opinion on it, and how it should be done. So I think that yeah, especially in education or a subject like that, it's important for you to kind of develop your own opinion on that, and your critical reflection and thinking. (Georgiana – Year 2)

In this case, being successful through being able to memorise a large number of facts was not seen as valuable in education:

. . . it is quite easy to kind of just memorise it for the exam and then, two weeks later you have probably forgotten most of it. But you know what you're going to write, you. . . . it's almost more formulaic, while an essay is much more . . . you can put much more of your personal style into it because you've got the time. (Georgiana – Year 2)

as these facts will be soon forgotten:

Iannone, P. and Simpson, A. (in press) University students' perceptions of summative assessment: the role of context. To appear in the Journal of Further and Higher Education

I think exams require good short term memory. I think you need to be concise in them, and I think exams are more the mark scheme and if you crack the mark scheme I think you've got it with an exam. (Caroline – Year 3)

Or, as Mary puts it:

So I think academically, exams don't actually really show the person's own intelligence. It's just how much you can remember something on a piece of paper, whereas coursework you can get out back and research, find your own personal experience [ . . . ] So I think academically, it shows off . . . academically exam shows off the person's memory, but coursework shows off the personal input and things like that. (Mary – Year 1)

Assessment which allows the students to bring their own perspective and their own voice was seen as the best assessment method and the pinnacle of the students' efforts in education. This is the case for the dissertation:

I think dissertation is really good because it allows the individual to . . . you've done all the subjects, you can pick which one you thrive in these and then carry that on. (Mary – Year 1)

Yes, I was just thinking what does it [*the dissertation*] assess. Because it is so heavily research based isn't it, it's about research really . . . Well. I suppose based on the idea that you choose what you are researching then it clearly ...you are being assessed on your ability to have a much deeper understanding of a specific area. (Jane – Year 2)

The pattern emerging from the analysis of the interviews is that education students prefer assessment methods such as coursework and projects/dissertations which are seen as better at assessing critical reading and critical thinking, and where students can better display their individual view of the debate they are asked to address. As Mary said in the earlier quotation, with these methods the personal contribution of the students can be heard. Our participants also tended to think that exams are not a good tool to assess

academic ability in education and success in exams is mostly attributable to memory and, in the words of Caroline, *cracking the mark scheme*. Indeed this pattern shows a close fit with the existing literature on students' preferences of summative assessment.

## 5 Comparing education and mathematics students

### 5.1 The surveys

As reported earlier, the study with the mathematics students adopted the same mixed-methods design and research tools of the one with education students and it was carried out at the same two institutions. Table 4 gives the summary of the survey sample for the mathematics students (which is detailed by Iannone and Simpson [2015]). The mathematics cohorts are much larger than the education cohorts at the two participant institutions and the larger sample represents a similar proportion of respondents. The gender distribution reflects the gender distribution in the two courses.

	Uni 1		Uni2		
	M	F	M	F	N(M)
Year 1	9	13	12	10	44
Year 2	15	13	7	5	40
Year 3/4	15	7	5	3	30
N(M)	39	33	24	18	114

Table 4: Demographic of the mathematics participants

A detailed description of the statistical analysis of the survey can be found in Iannone and Simpson (2015). Here we report the main results compared to the result of the education students study.

Fig 5 is the equivalent of Fig 3, but for the mathematics students. That is, it shows the relationship between mean preferences and discrimination for mathematics students with the mean and standard error bars for preference on the x-axis and for discrimination on the y-axis (reproduced from Iannone and Simpson 2015).

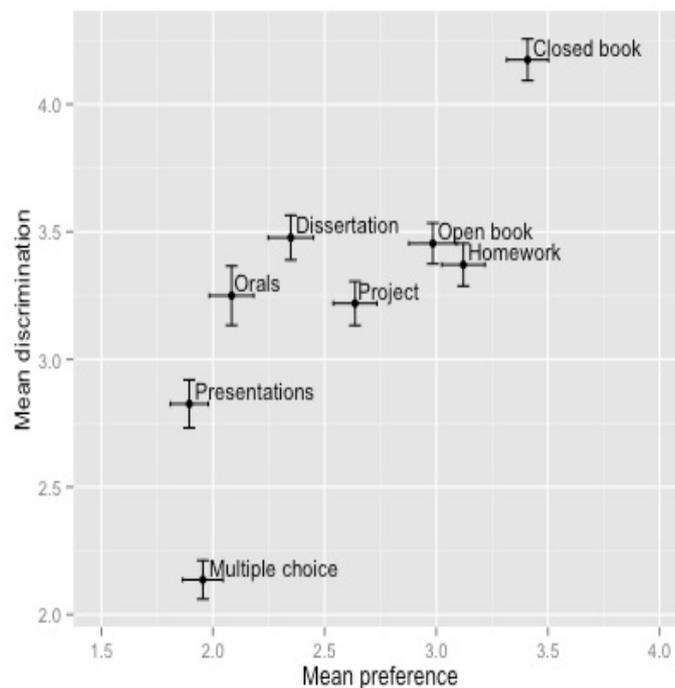


Figure 5: Mathematics students' discrimination and preference responses - with error bars

Comparing figure 3 and figure 5 we see that both mathematics and education students prefer assessment methods that discriminate according to academic ability. What these

assessment methods are however varies considerably. While mathematics students significantly prefer closed book exams over more innovative assessment methods such as projects and this method is also perceived to be the best discriminator of ability, for education students projects and dissertations are significantly preferred over more traditional methods, but they are also considered the best discriminator of ability in education. In both cases multiple choice exams are perceived to be a particularly poor method for discriminating for academic ability, even given its preference level. These findings are summarised in figure 6 and figure 7.

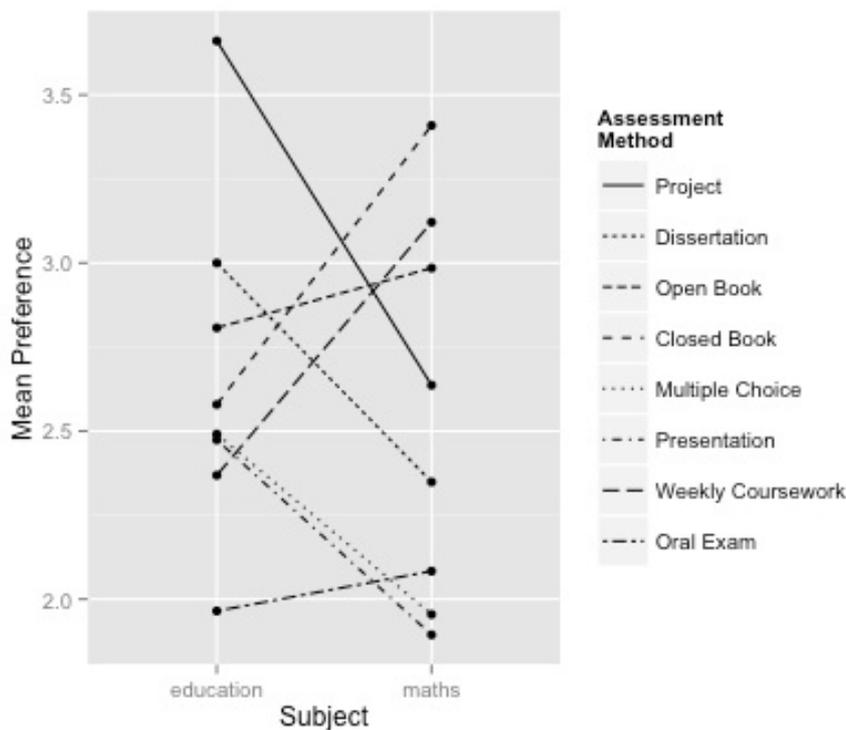


Figure 6: Interaction plot for assessment method preference between subjects

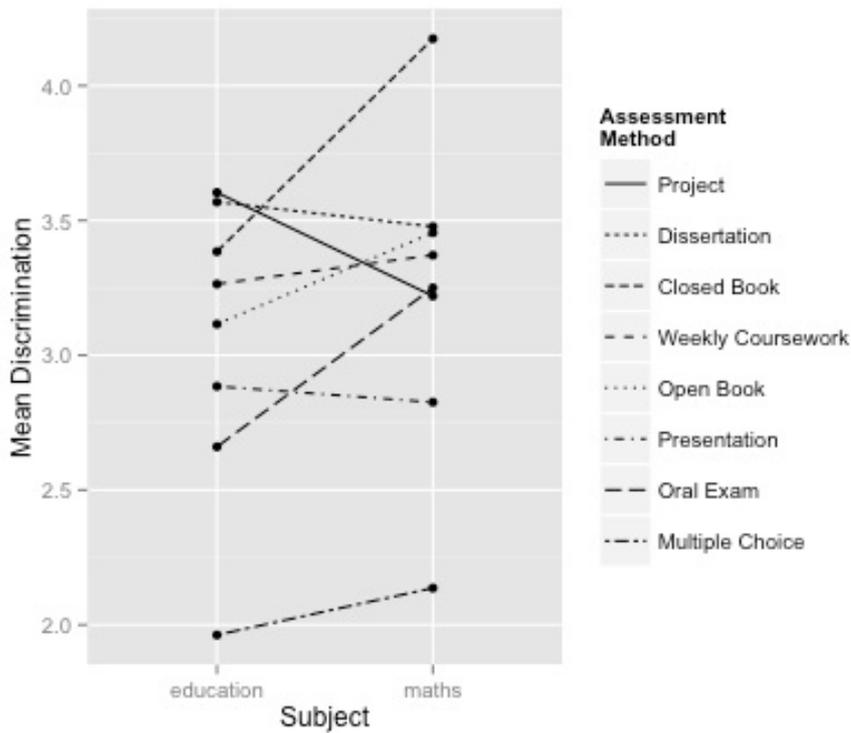


Figure 7: Discrimination interaction between subjects

### 5.2 The interviews

Twelve students took part in the semi-structured interviews for the study with the mathematics students, seven from Uni1 and five from Uni2, representing a little over 6% of the total number of mathematics students at both universities. The interviews were audio recorded, the data was fully transcribed and analysed following the thematic analysis approach (Attride-Stirling 2001). The originating thematic network comprised the following three umbrella themes: *fairness of assessment*; *assessment of abilities and the assessment diet*.

Both education and mathematics students had a focus on *fairness*. However, the centre of that focus appears quite different. Only a minority of mathematics students focused on the issue of stress leading to unfairness by impacting on candidates' ability to perform at their best, but this was key for most of the education students in our sample. It is possible that this is related to between group differences – for example that the education students are disproportionately female and the mathematics students male. The mathematics students were primarily concerned that marks should be awarded for individual performance and, in particular, fairness was achieved by avoiding marks they perceived as having a random element (e.g. in multiple choice questions where a candidate might guess a correct answer) or inappropriately (e.g. in coursework which might be open to collusion or plagiarism). The education students showed a concern for fairness through being able to demonstrate ability in a variety of ways. One might argue that the education students were interested in enhancing opportunities to gain marks fairly, while the mathematics students were interested in eliminating opportunities to gain marks unfairly. Both groups, however, were similar in their views of group work. They saw its usefulness especially as a formative method of assessment not bearing credit, but were concerned about marks rewarding individuals, rather than groups. The theme *assessment of abilities* came from both the mathematics and the education students, but again, underlying this theme there were some key differences. The core of those differences is in what the students think are the abilities on which they should be assessed: the Assessment Preferences Inventory showed that mathematics and education students have very different perceptions of which assessment methods discriminate for ability in their discipline. Mathematics students polarise the discussion regarding abilities in mathematics in terms of 'understanding, 'memory' and 'problem solving and process'. They appear to believe

that ability in mathematics is characterised by being able to apply their knowledge to a very wide range of problems. For them, the process and their problem solving abilities are also important in doing mathematics, and these abilities are best assessed in exam conditions. Education students on the other hand describe abilities in their discipline in relation to analytical thinking, critical reading and academic writing. For them, being successful in their academic subject means being able to master many different abilities and forming their own argument in support of their own thesis, as long as this argument is based on the critical elaboration and synthesis of the literature. These abilities are better assessed by work which is carried out in the students' own time, as it requires extensive literature searches and document reading. For education students, success in closed book examinations is equated with the ability to understand the mark scheme and having a good short-term memory, not necessarily that the student has a high ability in their academic discipline.

## **6 Discussion and concluding remarks**

The first unexpected finding of these studies is that both groups of students tended to prefer assessment methods which they perceive as discriminating for academic ability. This is in contrast to the suggestions of Scouller (1998) and Traub and MacRury (1990) which may imply that students could prefer assessment methods which enable them to obtain easy marks. Indeed, both groups perceived assessment methods which allowed students to obtain marks without necessarily assessing ability as unfair, albeit for slightly different reasons. One such example was the discussion in both sets of interviews around multiple choice assessments where both education and mathematics students deemed multiple choices to be unfair and where among the least preferred assessment methods. It is possible that this discrepancy with what is reported in the literature could be linked to the institutional context: in both the current study and the

study with mathematics students, the participants were enrolled at high ranking, research intensive UK universities, accepting students with high entry requirements. Following the suggestion made by Joughin (2010) there is certainly some evidence that the institutional context may influence students' perceptions of assessment. However, we did note in the quantitative analysis, for both sets of students, multiple choice was rated higher for preference than would be expected for their view of its ability to discriminate, so there may still be some impact of perceived ease of assessment method. The analysis of the Assessment Preferences Inventory for the education students has however highlighted a quite different pattern of assessment methods for both preference and discrimination to those for the mathematics students. The most notable difference which emerges from the comparison of the findings is what each group views as the abilities which are important in their subject. For mathematics students, ability in mathematics is characterised by understanding specific topics, being able to apply existing knowledge to new situations through effective problem solving and (for a minority of students) memory. Mathematics has a set way of being presented that needs to be followed to be successful and it is not the kind of subject where the student's voice and input are immediately recognisable, especially in the way in which is communicated, as in mathematics it can appear that there is only one right answer (see also the investigation of mathematics students' epistemic beliefs in Op't, De Corte and Verschaffel [2006]). Moreover in mathematics, there is often a clear sense of whether something is true or false and hence, for these students, truth may be seen as stable and, once something has been proved to be correct, it remains correct. For education students, abilities which are crucial to become successful in their subject are critical analysis, critical thinking and drawing together a wide range of issues to construct a research-based argument. In this discipline, the 'truth' and the arguments people

construct to support their position are legitimate subjects of debate and those arguments need to be constructed with the support of empirical and/or theoretical evidence. So, even though mathematics and education students concur in their preference for being assessed in ways which best differentiate on ability, the nature of these abilities is very different in each set of student. Thus they will require different ways of being assessed. It appears evident from this discussion that at the centre of the discrepancy we have observed is the way in which the students perceive their subject, what is important in that subject and how knowledge is accumulated. In short, students in our studies have distinct epistemic beliefs and these appear to be one of the factors influencing their perceptions of assessment, and these are grounded in disciplinary differences. We believe these findings are important to understand how to assess students. As early as the 1990s scholars such as Becher were warning against uniform recommendation for curriculum design:

Similarly, courses with a high factual content - as in certain areas of law - may appropriately be assessed by multiple choice tests; but the latter are entirely unsuitable for subjects such as sociology, where the emphasis tends to be on the need to decide between competing theories and to justify that decision. Again, pure and applied research on the topic tends to overlook these distinctions. (1994, 158)

However, much assessment literature continues to overlook these distinctions and make general recommendations which are taken to apply to all disciplines (Brown, 2004). Simpson (2015) appears to imply that increasing centralization of regulatory forces within institutions is outweighing disciplinary differences. Our study suggests that the students themselves understand the need for disciplines to be the key agent in determining patterns of assessment and the on-going homogenization of patterns of assessment will certainly leave some disciplines with practices which are not justified by their epistemologies.

Iannone, P. and Simpson, A. (in press) University students' perceptions of summative assessment: the role of context. To appear in the Journal of Further and Higher Education

### **Acknowledgements**

We would like to thank the School of Education and Lifelong Learning at the University of East Anglia for funding the research reported in this paper through the EDU Pump Priming Research Fund, 2014.

### **References**

- Attride-Stirling, J. 2001. Thematic networks: an analytic tool for qualitative research. *Qualitative Research* 1(3): 385-405.
- Becher, T. 1994. The significance of disciplinary differences. *Studies in Higher Education* 19(2): 151-161.
- Biglan, A. 1973. The characteristics of subject matter in different academic areas. *Journal of Applied Psychology* 57(3): 195-203.
- Birenbaum, M. 2007. Assessment and instruction preferences and their relationship with test anxiety and learning strategies. *Higher Education* 53(6): 749-768.
- Birenbaum, M., K. Breuer, E. Cascallar, F. Dochy, Y. Dori, J. Ridgway, R. Wiesemes, and G. Nickmans. 2006. "A Learning Integrated Assessment System." *Educational Research Review* 1 (1): 61-67.
- Birenbaum, M. 1994. Towards adaptive assessment - the students' angle. *Studies in Educational Evaluation* 20: 239-255.
- Black, P., C. Harrison, C. Lee, B. Marshall, and D. Wiliam. 2004. "Working Inside the Black Box: Assessment for Learning in the Classroom." *Phi Delta Kappan* 86 (1): 8-21.
- Brown. S. 2004. Assessment for Learning. *Learning and Teaching in Higher Education* 1: 81-89.
- DfE (2013) *Education and Training Statistics for the UK:2013* London: Department for Education.
- Furnham, A., and Chamorro-Premuzic, T. 2005. Individual differences and beliefs concerning preferences for university assessment methods. *Journal of Applied Social Psychology* 35: 1986-1994.

Iannone, P. and Simpson, A. (in press) University students' perceptions of summative assessment: the role of context. To appear in the *Journal of Further and Higher Education*

Furnham, A. Batey M. and Martin N. 2011. How would you like to be evaluated? The correlates of students' preferences for assessment methods. *Personality and Individual Differences* 50: 259-263.

Gielen, S., F. Dochy, and S. Dierick. 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.

Guest, G., Bunce, A. and Johnson, L. 2006. How Many Interviews Are Enough? An Experiment with Data Saturation and Variability. *Field Methods*. 18(1): 59-82.

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

Iannone, P., Simpson, A. 2013. Students' perceptions of assessment in undergraduate mathematics. *Research in Mathematics Education Journal* 15, 17-33

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

Johnson, B., and L.A. Turner. 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. London: Sage.

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

Lincoln, Y. S. and Guba, E. G. 1985. *Naturalistic Inquiry*. Newbury Park, CA: Sage Publications.

Iannone, P. and Simpson, A. (in press) University students' perceptions of summative assessment: the role of context. To appear in the *Journal of Further and Higher Education*

Marton F., and R. Saljo. 1997. Approaches to learning. In *The experience of learning. Implications for teaching and studying in higher education*, edited by F. Marton, D. Hounsell and N. Entwistle, 36-55. Edinburgh: Scottish Academic Press.

Medland, E. 2014. Assessment in higher education: drivers, barriers and directions for change in the UK. *Assessment & Evaluation in Higher Education* DOI: 10.1080/02602938.2014.982072.

Op 't Eynde P., De Corte E. and Verschaffel, L. 2006. Epistemic dimensions of students' mathematics-related belief systems. *International Journal of Educational Research* 45: 57- 70.

Reynolds, M. and Michael, R. 2013. *Group Work in Education and Training*. Routledge.

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

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

Simpson, A. 2015. Assessment and its outcomes: the influence of disciplines and institutions. *Assessment and Evaluation in Higher Education*. Ahead-of-print, DOI:10.1080/02602938.2015.1052369

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

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

Iannone, P. and Simpson, A. (in press) University students' perceptions of summative assessment: the role of context. To appear in the Journal of Further and Higher Education

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

## **Appendix**

### ***Appendix 1 - The API***

List of assessment methods - Education

<b>Multiple choice examination</b>	(e.g. a test taken in an exam room, where for each question you have to select one response from five possible choices)
<b>Written examination with no support materials</b>	(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)
<b>Written examination with support materials</b>	(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 module)
<b>Weekly coursework</b>	(e.g. coursework which you complete in your own time over the course of a week, based on the material covered in the module over that week)
<b>Project coursework</b>	(e.g. a piece of written work submitted in response to a question or problem, undertaken over the course of a number of weeks)
<b>Project presentation</b>	(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 )
<b>Oral examination</b>	(e.g. a Q&A session with a tutor present who can provide suggestions as you answer the question asked)
<b>Dissertation</b>	(e.g. a substantial piece of written work, on a set topic, undertaken over the course of a long period, such as a term or two)

### *Appendix 2 - The interview schedule*

1. What is your background? What subjects did you take before entering university?

Iannone, P. and Simpson, A. (in press) University students' perceptions of summative assessment: the role of context. To appear in the Journal of Further and Higher Education

2. Which assessment methods have you experienced as a student (at school? At university)
3. With which would you like your academic efforts to be assessed by? Why?
4. What do you think are the characteristics of a student who is successful in each one of these assessment methods?
5. Which ones are the characteristics (in terms of assessment) of a student who is successful in their degree in education?
6. If you were to design an assessment schedule for education students in your year at uni- versity (comprising more than 1 method but less than 4) which assessment methods would you include? Why?
7. A scholar who carried out research on students' perception of assessment methods at university has written:

Normal assessment was seen as a necessary evil that allowed them [*the students*] to accrue marks. The accompanying activities are described in terms of routine, dull artificial behaviour. Traditional assessment is believed to be inappropriate as a measure, because it appeared simply to measure memory, or in case of essay-writing tasks, to measure ability to marshal lists of facts and details. Alternative assessment was believed to be fairer, because by contrast, it appeared to measure qualities, skills and competences which would be valuable in contexts other than the immediate context of assessment. (Sambell et al., 1997)

Do you agree with this statement? Why?

Iannone, P. and Simpson, A. (in press) University students' perceptions of summative assessment: the role of context. To appear in the Journal of Further and Higher Education

Table 1: Demographic of the education participants

Table 2: Analysis of education students' preferences responses

Table 3: Analysis of education students' discrimination responses

Table 4: Demographic of the mathematics participants

Figure 1: Spread of assessment methods across disciplines in the UK

Figure 2: Education students' preference - interaction between years

Figure 3: Education students' discrimination and preference responses - with error bars

Figure 4: Thematic network

Figure 5: Mathematics students' discrimination and preference responses - with error bars

Figure 6: Interaction plot for assessment method preference between subjects

Figure 7: Discrimination interaction between subjects