

Students' perceptions of assessment in undergraduate mathematics

P. Iannone^{a*} and A. Simpson^b

^a *School of Education and Lifelong Learning, University of East Anglia, Norwich, UK*

^b *School of Education, Durham University, UK*

A consistent message emerges from research on undergraduate students' perceptions of assessment which describes traditional assessment as detrimental to learning. However this literature has not included students in the pure sciences. Mathematics education literature advocates the introduction of innovative assessment at university. In this literature however students' voices tend to be unheard. We investigate mathematics students' perceptions of assessment at one high-ranking UK University. We find that, in contrast to the message from general the literature, students perceive traditional assessment as the best discriminator of ability. We suggest that the drive towards the introduction of innovative assessment in mathematics should be considered in the light of students' perceptions and that further research is needed to investigate the origins of those perceptions as well as how these perceptions may be influenced by the subject of study.

Keywords: assessment of mathematics; higher education; students' perceptions

Introduction

Research on assessment in higher education has become central to the development of policies and practices in higher education for decades (Elton and Laurillard 1979; Ramsden 1988; Brown and Glasner 1999; Boud and Falchikov 2007). Some of this research has focused on the effects of assessment on student learning (Harlen and Crick 2003), on the validity, fairness and value of traditional and innovative assessment forms (Struyven, Dochy and Janssens 2005), and on the impact of assessment on the hidden curriculum at university (Sambell, McDowell, and Brown 1997).

*Corresponding author. Email: p.iannone@uea.ac.uk

There is strong empirical evidence that students' perceptions of the value and validity of assessment affect their learning (Scouller 1998), albeit that the interaction between those perceptions and the students' approaches to learning are far from straightforward. In a systematic review of the literature, Harlen and Crick (2003) highlight the complex ways in which students' views of upcoming assessments influence their motivations, and Baeten, Dochy and Struyven (2008) give an example of assessments which might be intended to encourage a deep approach to learning but actually gave rise to higher levels of surface approaches.

Segers, Dochy, and Cascallar (2003) note that the 'pre-assessment effect', which influences how student learning is affected by an assessment task, must be mediated by how students perceive that task, its fairness, validity and the values it embodies. Thus, while academics may have a wide variety of reasons for selecting particular assessment methods, they need to be aware of their students' perceptions of these methods and how these perceptions influence their students' learning.

Struyven et al. (2005) produced a comprehensive review of the literature on students' perceptions of assessment in higher education. The rationale behind this review lies in the importance of students' perceptions of reality. As Entwistle (1991, 202) notices "Thus, it is students' perceptions of the learning environment that influence how a student learns, not necessarily the context in itself." The diet of different methods, forms and impacts of assessment faced by students is clearly a dominant factor of this learning environment. Therefore, the investigation of these perceptions of assessment is important in understanding how students' learning is influenced by the assessment they are assigned. Struyven et al. (2005) suggested that the substantial body of research they reviewed contains a consistent message. It supports the idea that students have clear perceptions of what cognitive skills are assessed by different

assessment methods, and that they are also often conscious of the ways they adapt their learning to their perceptions of assessment. Across a wide range of literature reviewed, it reinforces the view, also noted by Sambell et al. (1997), that students perceive traditional assessment as inadequate to assess the cognitive skills relevant to their subject, and they perceive success in traditional assessment as relying mostly on memory recall. However, the general literature on assessment comes from research with particular groups of students. For example, Struyven et al. (2005) list 36 empirical studies, 28 of which were with participants from four academic subject areas (psychology, medicine, engineering and education), and none of which included participants on degrees in mathematics.

Assessment of mathematics at university

There appears to be little research on students' perceptions of assessment in university mathematics. By contrast, there is a large body of literature devoted to the support of 'novel' forms of assessment such as projects and presentations (Berry and Houston 1995; Houston and Lazenbatt 1996; Challis, Houston and Stirling 2004; Steen 2006). Most of this work tends to downplay the role of closed-book examination in favour of innovative assessment. Hirst and Biggs (1968) were amongst the first to support the introduction of projects in mathematics. They argue that mathematics students lack insight into the messy process of doing mathematics, and that tackling a project would show them just how the process of doing mathematics is different from the linear exposition of mathematics in textbooks. In recent years, many authors have advocated the introduction of other innovative methods of assessment, such as poster presentations (Houston, 2001), a combination of projects, posters and oral presentations (Povey & Angier 2006), multiple choice questions (Haines and Crouch 2005; Ramesh 2009) or

mathematical writing (McConlogue, Mitchell, and Vivaldi 2010). Amongst the reasons for supporting the introduction of innovative assessment methods is the perception that traditional methods such as closed-book examinations are not suitable for assessing the variety of skills required in the workplace (Challis et al. 2004). The prevalence of traditional assessments has also been suggested as a source of the apparent link students make between being good at mathematics and having a good memory, and their apparent lack of appreciation of conceptual understanding in mathematics (Schoenfeld 1989). Berry and Houston (1995, 22), for example, argue that more innovative forms of assessment focused on communication, such as project and poster presentations, support “exposing and confronting misconceptions” and put “emphasis on concept as well as procedure”. Others (Burton and Haines 1997) call for a change in the assessment diet of mathematics students on the basis that it is necessary to move away from “the dominant epistemological paradigm which is one of absolutism of knowledge ...[*which is*]... reproductive and content dominated” (Burton and Haines 1997, 275), and that it is “the extremely narrow view held by many mathematicians about their discipline and its teaching and learning” (ibid., 287) which holds back change to innovate away from traditional forms.

Despite the repeated calls for innovation, a recent survey of assessment in UK mathematics departments (Iannone and Simpson 2011) shows that assessment is highly uniform across institutions, and dominated by closed-book examination. While some novel forms of assessment have found use in the mathematics community (all but one of the department surveyed had at least one module assessed by a project), other forms are not yet as widespread. There appears to be a tension between professional bodies and individual academics who point to the ‘special’ nature of mathematics as a reason to challenge the prevailing direction of change of current practice (LMS 2010), and the

voices of pedagogues who argue strongly in favour of innovative assessment. The majority of the mathematics education literature, however, consists of scholarly opinions rather than empirical research into students' views. Thus, just as the literature into students' views of assessment lacks the views of mathematics students, the literature in mathematics education lacks research on students' views of assessment.

This paper investigates mathematics students' perceptions of assessment, what cognitive processes students perceive as being assessed by particular assessment methods, and which assessment methods are better discriminators of mathematical ability. To do this, we use an approach and methods similar to the general assessment literature. Specifically, we ask:

- What cognitive processes do students perceive being assessed by particular assessment methods in mathematics?
- What assessment methods do students perceive to be valid discriminators of mathematical ability?

Although we are aware that all mathematics departments use formative assessment to guide their students' learning, with feedback on their work, there is strong evidence in the research literature (Scouller 1998) that the higher the stakes, the more assessment influences students' approaches to learning. Hence, for this research we focus on summative assessment. We also focus on distinct assessment types rather than on the different ways in which an assessment type can be delivered. For example, a multiple choice test can be taken as a pen and paper test in a lecture hall or as an online computer-marked test. In this paper, we will consider all these variants as "multiple choice tests".

Methods

The pilot study

For the pilot study, we interviewed nine mathematics students from a research-intensive university in the UK to pilot aspects of the questionnaire for our main study. The questionnaire used was adapted from the Assessment Preferences Inventory (API) developed by Birenbaum (1994). The original version of the API consists of 67 items measuring seven different areas of assessment (including preparation, cognitive processes and conative aspects). While this is a very comprehensive approach for assessing students' perceptions of assessment methods, it is too cumbersome for practical use. For this reason we adopted van der Watering, Gijbels, Dochy, and van der Rijt's (2008) modification of the API, and in this version we identified those assessment methods (along with an explanation in the form of an example) which are in widespread use, or which feature in the literature as potential forms of assessment in university mathematics in the UK (Table 1). We excluded those methods which have been ignored in the literature, or considered unsuitable to assess mathematics. In order to simplify the idea of 'cognitive processes' we focused on just two terms used commonly in discussions about mathematics and its assessment (Bergqvist 2007), and which feature in the API developed by van der Watering et al. (2008): *memory* and *understanding*.

[Table 1 goes here]

The questionnaire had two sections (which are reproduced in the Appendix). One section asked the students to place the eight assessment methods listed in Table 1 in rank order according to their perception of the method as:

a good measure of mathematical ability...so that students who are likely to be good mathematicians are likely to score most highly and those who are likely to be poor mathematicians are likely to score most poorly.

This section aimed at assessing students' perceptions of assessment methods as better or worse discriminators of ability. The other section asked students to assess how well the assessment methods in the questionnaire might measure someone's memory ("e.g. ability to recall a definition from the course") or someone's understanding ("e.g. being able to think about and use a mathematical idea from the course"), with answer boxes allowing a response on a five point Likert scale from 'poor measure' to 'excellent measure'. Given that summative assessments are purportedly designed to distinguish mathematical ability (whatever characteristics the assessment designers believe embody the notion of 'ability'), we wished the students to make clear choices regarding how well a method discriminates between candidates. On the other hand, it is reasonable for someone to believe that some assessment methods share similar characteristics, such as how much they rely on the candidate's memory and how much they rely on their understanding. Therefore, it was felt there was no need for students to make an absolute choice between methods on these characteristics (Russell and Gray 1994). Note that, throughout the questionnaire, the name of the assessment method was consistently accompanied by an explanatory example of that method in Table 1.

The aim of this pilot study was to check that students interpreted these definitions of the assessment methods clearly, and that their interpretation was consistent with the explanatory examples included in the questionnaire. The students who took part in the pilot represent an opportunistic, but relevant, sample; they were volunteers who responded to an invitation to take part in the interview study, and were at various stages of their three-year course at an institution of similar ranking and assessment pattern to that in the main study. The students were asked to fill in the

section of the questionnaire which asked them to rank order the assessment methods as discriminators of ability, and to explain their thinking aloud as they did so. The interviewer intervened only to ask for clarification about the choices students made when these were unclear. These interviews lasted between 6 and 13 minutes, and were audio recorded and fully transcribed. Thematic analysis of the interviews revealed students' interpretations of the assessment methods they were asked to rank.

We found that, across all methods and all students, there was indeed a clear and consistent understanding of the meaning of the methods drawn from the name and explanatory example. For instance, while oral examinations in mathematics are rare, students showed a good understanding of what we intended by this phrase. One student described the oral examinations as:

If it is just one-on-one and it's someone working through a problem on a chalkboard with someone just pointing out a little bit along the way, I think that's, that would be a good assessment, because it is different problems but you can see how they work through it and you can gauge what they're doing, why they're doing it, as they go through and if they can explain what they're doing well, you know they're a good mathematician. (Student 2)

Similarly, project presentations are relatively rare, but again students showed a clear and consistent understanding:

A presentation might be better because you'd have to, you can't just get the answers off the Internet, you have to go and stand to make your, to talk about it, so [...] This is probably the best, examination, I would say. (Student 1)

In all eight methods, we found the same level of clarity.

Main study

Our participants were mathematics students at a high ranking research-intensive

university in the UK. The institution where we collected our data demands the highest entrance examination grades although it does not require prospective students to attend a subject-specific entrance interview. In this institution, Year 1 modules are assessed mostly by closed-book examination (with some statistics modules assessed by open book examination) and a small coursework component, predominantly in the form of weekly exercise sheets. All Year 2 modules are assessed by closed-book examination, with the exception of one module which has a 20% coursework component in the form of e-assessment. All Year 3 modules are assessed by closed-book examination, with the exception of one option module in mathematics education which has a large coursework component, and the project modules which are assessed by a written report and a poster presentation. If we compare this pattern with a recent survey of assessment in UK mathematics departments (Iannone and Simpson 2011), it is clear that this assessment diet is representative of assessment at research-intensive mathematics departments.

The study was conducted with 48 undergraduate students studying for a mathematics degree, representing about one third of the first year cohort. The modal student in the research reviewed by Struyven et al. (2005) is in their first year and, while students' perceptions of assessment will be influenced by their previous experience, choosing students at the start of their first year mitigates the influence of previous success or failure, at least in the university context. The students were chosen from two out of six, randomly-assigned, seminar groups for a core first year lecture course on mathematical problem solving. They were asked to complete the full questionnaire (with both the section asking for the rank ordering of methods according to how well they discriminate according to ability, and the section asking them to rate the methods as measures of memory and understanding) at the beginning of their seminar session. To

avoid bias caused by the order of the sections, they were presented in random order. The data were collected in Week 8 of the first term.

Analysis

The analysis was performed using the statistics package R, and four questionnaires were discarded as they were filled in incorrectly or incompletely. While there is much discussion about the analysis of Likert scales with parametric methods, we note that Rasmussen (1989) demonstrates that they result in increased type I or II error rates only in the most non-normal of situations.

A two-way analysis of variance was conducted to examine the ratings given to the cognitive characteristics (memory and understanding) for each type of assessment. We found significant main effects for characteristics ($F[1,688]=267.9, p<0.001$) and assessment type ($F[7,688]=28.2, p<0.001$) and a significant interaction of characteristic and assessment type ($F[7,688]=20.3, p<0.001$)

Figure 1 shows the mean rating for each type of assessment as a test of memory (with standard error bars). Post-hoc pairwise *t*-tests, with Bonferroni adjustment, showed that closed-book examinations were rated significantly higher than all other types of assessment, and that oral examinations and multiple choice questions were significantly higher than the remaining types of assessment (*p* values shown in table 2).

[Figure 1 goes here]

[Table 2 goes here]

Figure 2 shows the mean rating for each type of assessment as a test of understanding (with standard error bars). Post-hoc pairwise *t*-tests, with Bonferroni adjustment, showed that multiple choice was rated significantly lower than all types of assessment apart from open book and presentations, and open book examinations were rated

significantly lower than the remaining methods apart from example sheets, projects and presentations (p values shown in table 3).

[Figure 2 goes here]

[Table 3 goes here]

Figure 3 shows box and whisker diagrams summarising the responses given to ranking the types of assessment according to how good they are as discriminators of mathematical ability. A Friedman test showed that the students had differentially ranked the types of assessment according to how well they discriminate ($\chi^2(7)=117.4$, $p<0.001$). A post-hoc pairwise comparison indicates that types of assessment are ranked significantly differently if their ranks sums differ by more than 48 (Conover 1980).

[Figure 3 goes here]

To examine the interaction between the cognitive characteristics and the type of assessment, Bonferroni-adjusted paired t -tests were conducted for each assessment method to give a measure of the extent to which the students perceived the assessment as more of a measure of understanding than of memory. As illustrated in the interaction plot in Figure 4, there were no significant differences between memory and understanding for closed-book examinations or multiple choice ($t(43)=1.94$, $p=0.227$; $t(43)=0.96$, $p=1.000$), and for all other types of assessment, understanding dominated memory ($t(43)=4.88$, 6.91 , 12.83 , 13.49 , 7.85 , 10.67 ; all $ps<0.001$).

[Figure 4 goes here]

Discussion and concluding remarks

The aim of this paper was to investigate mathematics undergraduate students' perceptions of assessment methods in terms of how much each method acts as a test of two cognitive characteristics associated with mathematics – memory and understanding – and the extent to which each method is seen as a discriminator of mathematical ability.

Our findings show that the stereotypical assessment method in mathematics - closed-book examination - is perceived to be the best discriminator of mathematical ability. Although students perceive assessment of memory to be dominant over assessment of understanding for closed-book examinations, this method was rated highly for both cognitive factors, and indeed, the difference was not statistically significant. One might also consider this to be an indication that students see closed-book examinations as a well-balanced form of assessment. More innovative forms of assessment which have come into use only recently in mathematics, such as projects and dissertations, occupy a middle rank for assessing mathematical ability, while presentations are perceived to be the weakest discriminator of ability (Fig. 3). It is interesting to note that this finding seems to conflict with findings in the general assessment literature. The review by Struyven et al. (2005) reports that, on the whole, students perceive innovative assessment methods as better for assessing key cognitive skills in their subject, and Sambell et al. (1997) discuss how students in their study (which involved thirteen different subjects but did not include students from a pure, numerically-based science) held very negative views of traditional assessment: "... In their [*the students*'] view, exams had little to do with the more challenging task of trying to make sense and understand their subject." (p. 357)

Moreover, Sambell et al. (1997) report how these students perceive traditional assessment (and in particular closed-book examination) as detrimental to their learning, while innovative forms of assessment (such as projects and presentations) are perceived to be fostering “high quality of learning”. (p. 358)

From Figure 1, we see that assessment methods which allow access to external materials are perceived to have a less dominant memory factor than those which do not (such as closed-book examinations and multiple choice questions). This is perhaps not surprising, as it is reasonable to believe that access to external materials will reduce the amount of facts that need to be committed to memory. As for assessment methods which are perceived to test understanding over memory, oral examination is perceived by our students to be the method which most dominantly assesses understanding, followed by closed-book examination. Interestingly, this finding resonates with some of the research on oral assessment where this method is found to be beneficial for students in terms of engagement with the subject being studied (Huxham, Campbell and Westwood 2010), and where it is perceived by students to require deep understanding (Joughin 2007). The assessment method perceived to test understanding the least is multiple choice questions. If we look at interactions between cognitive characteristics of assessment methods (Figure 4), students’ perceptions seem to indicate that all assessment methods included in the API assess understanding over memory, apart from closed-book examination and multiple choice questions. The students appear to have seen quite large differences between the methods as measures of memory compared to the differences between the methods as measures of understanding.

In the institution where we carried out our data collection, as in any learning environment, students’ perceptions of assessment methods as a discriminator of mathematical ability will be linked to their enculturation into mathematics (Schoenfeld

1992), and a part of this will consist of enculturation into the assessment of mathematics. Skovsmose and Nielsen (1996) point out that even when subject values are not explicitly discussed by lecturers, students can infer them from their actions. Perrenet and Taconis (2009) show how mathematics students' beliefs and attitudes towards mathematics (in the context of problem solving) tend to move towards their lecturers' beliefs and attitudes during the course of their degree. In the light of this and other research evidence, it is not unreasonable to assume that messages about mathematics conveyed by lecturers are likely to be received by students as they progress in their studies. The messages students receive from mathematicians in the form of assessment tasks form part of their enculturation into mathematics. In this context we can infer, for example, that the predominance of the closed-book examination in the students' assessment diet may convey the message that success in this assessment mode implies success in mathematics. Implicit messages regarding enculturation into the assessment of mathematics may, of course, be different (or differently received) in less research-intensive institutions (where assessment by closed-book examination may be less dominant) or in other higher education cultures (where other assessment methods may be dominant). Thus, further research might usefully investigate the relationship between the messages that mathematics lecturers intend to convey with their assessment methods and the messages construed by the students, across different types of institutions.

It can also be argued that successful mathematics students hold their views because of their own past success with particular assessment methods, and that the more traditional the subject and the university, the more likely it is that students hold very traditional views. As the students in our sample achieved the highest grades in school examinations – which in the UK in mathematics consist almost solely of closed-book

examinations – our students may think of closed-book examinations as the best assessment of mathematical ability. However, the traditionalism and past experience of the participants in our study cannot account for other findings emerging from the data. It is surprising to see that oral examinations, while perceived as relatively reliant on memory, are not perceived to have memory as the dominant cognitive factor. This suggests that our findings may not simply be a reflection of inherent conservatism, reflecting the students' own successful experience with assessment methods, but may indicate genuinely held beliefs about what kinds of assessment focus on what they deem to be important and what methods best discriminate on the basis of ability in their subject. After all, even though oral examinations remain the dominant form of assessment of mathematics in many European countries, they have disappeared in the UK.

We have noted how little of the generalist research on assessment focuses on students of pure numerically-based subjects. It could be that it is the nature of mathematics that makes assessment patterns different from those in other academic disciplines, as professional bodies believe (LMS 2010), rather than the resistance to change supposed to be typical of many mathematicians (as claimed by Burton and Haines 1997). We acknowledge, however, that more research in different institutions and with larger samples of students is needed to substantiate this hypothesis.

In the light of our findings, it appears that more attention needs to be given to students' perceptions of assessment, and that the calls to innovate that are so common in the mathematics education literature need to be considered in the light of how these new assessment methods will influence students' learning. Our findings also suggest the need to be cautious in adapting the recommendations from the general research literature to the practices of specific subjects.

References

- Baeten, M., F. Dochy, and K. Struyven. 2008. Students' approaches to learning and assessment preferences in a portfolio-based learning environment. *Instructional Science* 36: 359–74.
- Bergqvist, E. 2007. Types of reasoning required in university exams in mathematics. *Journal of Mathematical Behavior* 26: 348–70.
- Berry, J., and K. Houston. 1995. Students using posters as a means of communication and assessment. *Educational Studies in Mathematics* 29: 21–7.
- Birenbaum, M. 1994. Towards adaptive assessment – the students' angle. *Studies in Educational Evaluation* 20: 239–55.
- Boud, D., and N. Falchikov. 2007. *Rethinking assessment in higher education learning for the longer term*. London: Routledge.
- Brown, S., and A. Glasner, eds. 1999. *Assessment matters in higher education*. Buckingham: Open University Press.
- Burton L., and C. Haines. 1997. Innovation in teaching and assessing mathematics at university level. *Teaching in Higher Education* 2: 273–94.
- Challis, N., K. Houston, and D. Stirling. 2004. *Supporting good practice in assessment in mathematics, statistics and operational research: Briefings and guides*. Birmingham: MSOR (Maths, Stats & OR Network).
- Conover, W.J. 1980. *Practical nonparametric statistics*. New York: Wiley.
- Elton, L., and D. Laurillard. 1979. Trends in research on student learning. *Studies in Higher Education* 4: 87–102.
- Entwistle, N.J. 1991. Approaches to learning and perceptions of the learning environment. *Higher Education* 22: 201–4.
- Haines, C., and R. Crouch. 2005. Applying mathematics: Making multiple-choice questions work. *Teaching Mathematics and its Applications* 24, nos. 2–3: 107–13.
- Harlen, W., and R.D. Crick. 2003. Testing and motivation for learning. *Assessment in Education: Principles, Policy and Practice* 10: 169–207.
- Hirst K., and N. Biggs. 1968. Undergraduate projects in mathematics. *Educational Studies in Mathematics* 1, no. 3: 252–61.

- Houston, K. 2001. Assessing undergraduate mathematics students. In *The Teaching and Learning of Mathematics at University Level: An ICMI Study*, ed. D. Holton, 407–22. NL: Kluwer Academic Publishers.
- Houston, K., and A. Lazenbatt. 1996. A peer-tutoring scheme to support independent learning and group project work in mathematics. *Assessment and Evaluation in Higher Education* 21: 251–66.
- Huxham, M., F. Campbell, and J. Westwood. 2010. Oral versus written assessments: A test of student performance and attitudes. *Assessment and Evaluation in Higher Education* 37, no. 1: 125–36.
- Iannone, P., and A. Simpson. 2011. The summative assessment diet: How we assess in mathematics degrees. *Teaching Mathematics and its Applications* 30, no. 4: 186–96.
- Joughin, G. 2007. Student conceptions of oral presentations. *Studies in Higher Education* 32, no. 3: 323–36.
- LMS. 2010. *Mathematics degrees, their teaching and assessment*, London Mathematical Society.
- McConlogue, T., S. Mitchell, and F. Vivaldi. 2010. Beyond templates 2: Exploring students' approaches to learning on a mathematical writing course. *MSOR Connections* 10, no. 2: 12–5.
- Perrenet J., and R. Taconis. 2009. Mathematical enculturation from the students' perspective: Shifts in problem-solving beliefs and behaviour during the bachelor programme. *Educational Studies in Mathematics* 71 :181–98.
- Povey, H., and C. Angier. 2006. The assessment of undergraduate mathematicians: Recrafting assessment of learning to provide opportunities for assessment as learning. *MSOR Connections* 6, no. 4: 43–7.
- Ramesh, N. 2009. Some issues on assessment methods and learning in mathematics and statistics. *MSOR Connections* 9, no. 4: 28–31.
- Ramsden, P. 1988. Studying learning, improving teaching. In *Improving Learning: New Perspectives*, ed. P. Ramsden, 13–31, London: Kogan Page.
- Rasmussen, J. 1989. Analysis of Likert-scale data: A reinterpretation of Gregoire and Driver, *Psychological Bulletin* 105, no. 1: 167–70.
- Russell, P.A., and C.D. Gray. 1994. Ranking or rating? Some data and their implications for the measurement of evaluative response. *British Journal of Psychology* 85: 79–92.

- 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: 349–71.
- Schoenfeld, A.H. 1989. Explorations of students' mathematical beliefs and behavior. *Journal for Research in Mathematics Education* 20: 338–55.
- Schoenfeld, A.H. 1992. Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In *Handbook of Research on Mathematics Teaching and Learning*, ed. D.A. Grouws, 334–70, New York: Macmillan.
- Scouller, K. 1998. The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education* 35: 453–72.
- Segers, M., F. Dochy, and E. Cascallar. 2003. *Optimizing new modes of assessment: In search of qualities and standards*. Dordrecht: Kluwer.
- Skovsmose, O., and L. Nielsen. 1996. Critical mathematics education. In *International Handbook of Mathematics Education*, ed. A.J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, and C. Laborde, 1257–88, Dordrecht: Kluwer.
- Steen, L.A., ed. 2006. *Supporting assessment in undergraduate mathematics*. Washington, DC: Mathematical Association of America.
- Struyven, K., F. Dochy, and S. Janssens. 2005. Students' perceptions about evaluation and assessment in higher education: A review. *Assessment and Evaluation in Higher Education* 30: 331–47.
- van de Watering, G., D. Gijbels, F. Dochy, and J. van der Rijt. 2008. Students' assessment preferences, perceptions of assessment and their relationship to study results. *Higher Education* 56: 645–58.

[illegible]

Section B

Consider each of the following ways in which you might assess mathematical ability. For each, tick the box which most accurately describes how well you feel the assessment method might measure someone's

- a) Memory (e.g. ability to recall a definition from the course)
- b) Understanding (e.g. being able to think about and use a mathematical idea on the course)

poor measure					Excellent measure	
		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)

a) Memory					
b) Understanding					

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)

a) Memory					
b) Understanding					

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)

a) Memory					
b) Understanding					

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)

a) Memory					
b) Understanding					

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)

a) Memory					
b) Understanding					

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)

a) Memory					
b) Understanding					

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)

a) Memory					
b) Understanding					

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)

a) Memory					
b) Understanding					

Table 1. Taxonomy of assessment methods.

Assessment methods	Explanatory example
Multiple-choice examination	Test taken in an exam room, where for each question the student can select one response from five possible choices
Written examination with no support materials	Test taken in an exam room, with a separate booklet in which the student writes solutions, but no support material is allowed
Written examination with support materials	Test taken in an exam room, with a separate booklet in which the student writes solutions, but support material is allowed
Weekly examples sheets	Test completed in the students' own time over the course of a week
Project coursework	A piece of written work submitted in response to a question or problem, undertaken over the course of a number of weeks
Project presentation	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	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	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

Table 2. Means, standard errors and p -values for post-hoc t-tests (with Bonferroni correction) comparing assessment types as measures of memory.

	Closed-book $\sigma=4.45$ s.e.=0.13	Multiple Choice $\sigma=3.45$ s.e.=0.20	Oral Exam $\sigma=3.41$ s.e.=0.17	Presentations $\sigma=2.36$ s.e.=0.16	Example Sheets $\sigma=2.02$ s.e.=0.14	Dissertations $\sigma=2.00$ s.e.=0.15	Open Book $\sigma=1.98$ s.e.=0.14	Projects $\sigma=1.98$ s.e.=0.14
Closed-book Exam	-	-	-	-	-	-	-	-
Multiple Choice	0.0012	-	-	-	-	-	-	-
Oral Exam	<0.001	1.000	-	-	-	-	-	-
Presentations	<0.001	<0.001	<0.001	-	-	-	-	-
Example Sheets	<0.001	<0.001	<0.001	1.000	-	-	-	-
Dissertations	<0.001	<0.001	<0.001	1.000	1.000	-	-	-
Open Book Exam	<0.001	<0.001	<0.001	1.000	1.000	1.000	-	-
Projects	<0.001	<0.001	<0.001	1.000	1.000	1.000	1.000	-

Table 3. Means, standard errors and p -values for post-hoc t-tests (with Bonferroni correction) comparing assessment types as measures of understanding.

	Oral Exam $\sigma=4.30$ s.e.=0.10	Closed-book $\sigma=4.16$ s.e.=0.10	Dissertations $\sigma=4.11$ s.e.=0.13	Example Sheets $\sigma=3.98$ s.e.=0.14	Projects $\sigma=3.95$ s.e.=0.13	Presentations $\sigma=3.77$ s.e.=0.10	Open Book $\sigma=3.45$ s.e.=0.17	Multiple Choice $\sigma=3.32$ s.e.=0.12
Oral Exam	-	-	-	-	-	-	-	-
Closed-book Exam	1.000	-	-	-	-	-	-	-
Dissertations	1.000	1.000	-	-	-	-	-	-
Example Sheets	1.000	1.000	1.000	-	-	-	-	-
Projects	1.000	1.000	1.000	1.000	-	-	-	-
Presentations	0.108	0.904	1.000	1.000	1.000	-	-	-
Open Book Exam	<0.001	0.003	0.008	0.108	0.160	1.000	-	-
Multiple Choice	<0.001	<0.001	<0.001	0.008	0.013	0.333	1.000	-

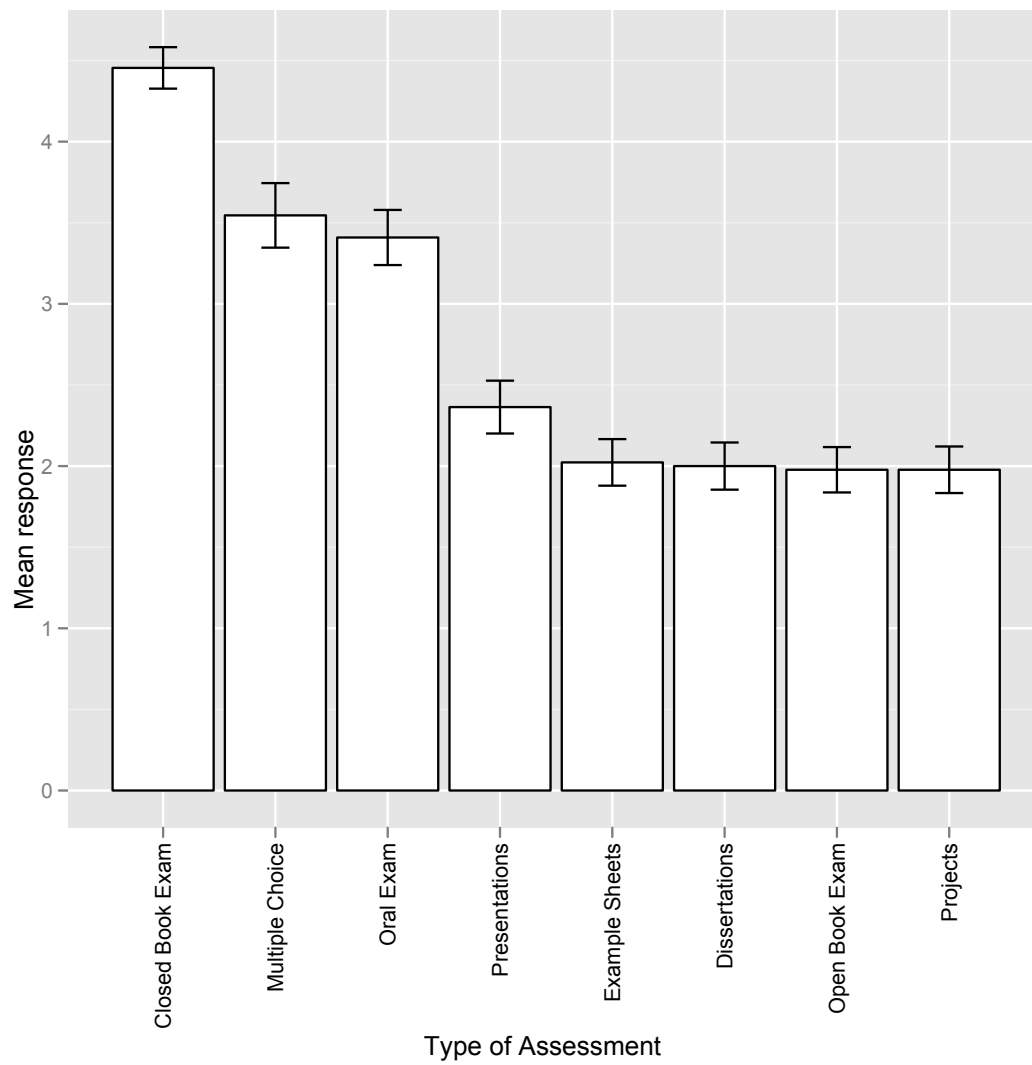


Figure 1. Mean rating for each type of assessment as a test of memory (with standard error bars).

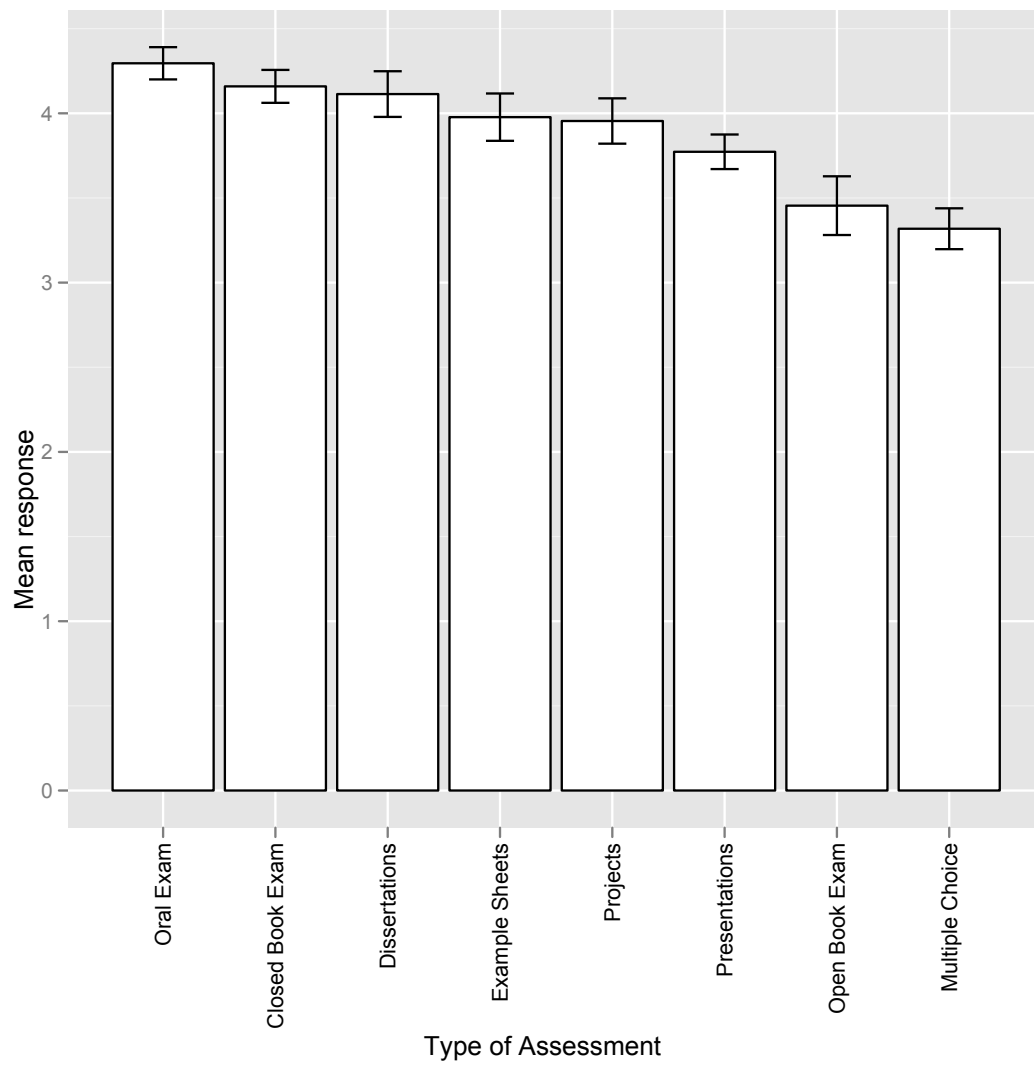


Figure 2. Mean rating for each type of assessment as a test of understanding (with standard error bars).

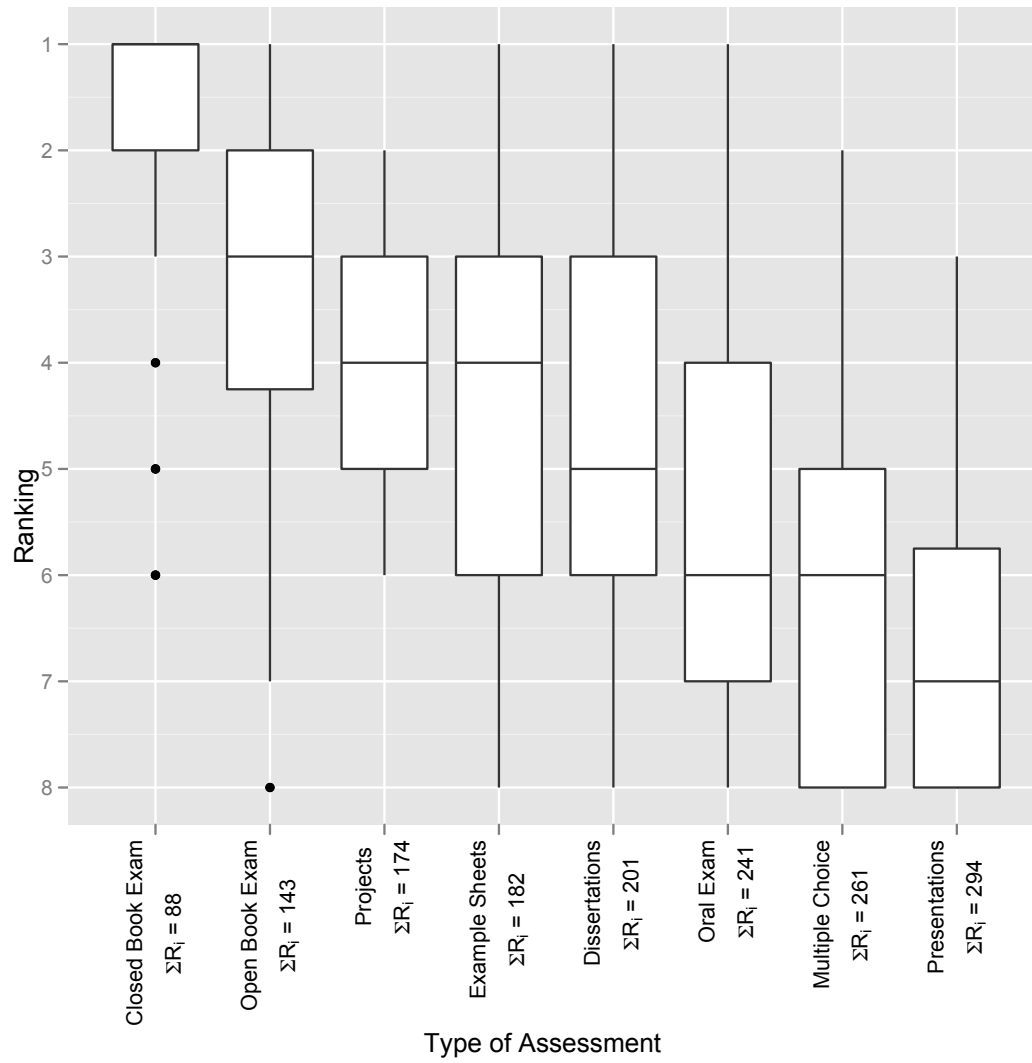


Figure 3. Ranking of the types of assessment according to how well they distinguish good mathematicians from poor ones, with rank sums for each assessment type and with outliers marked.

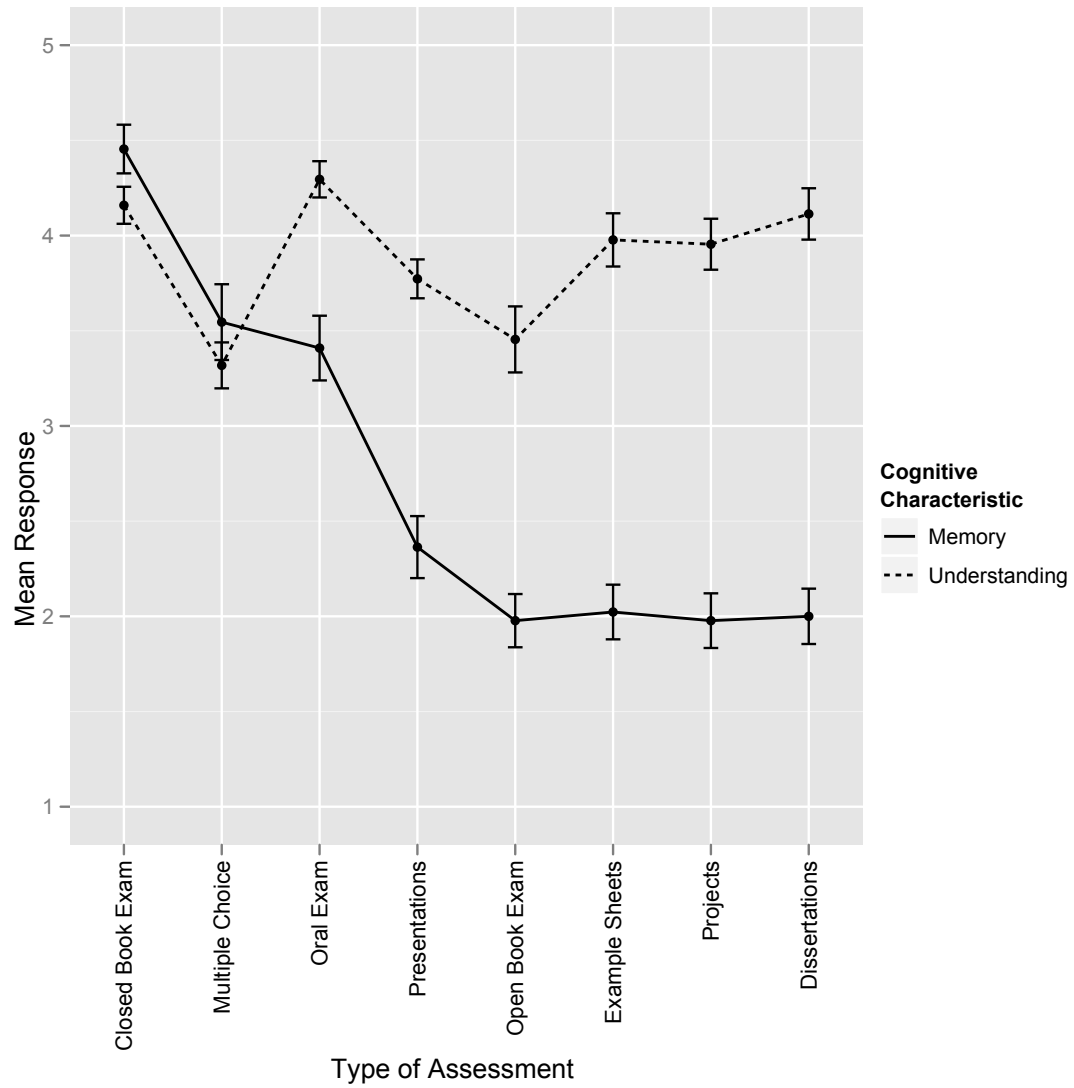


Figure 4. Interaction plot between the cognitive characteristics (memory and understanding) and the type of assessment.