# Students' views of value and validity in undergraduate mathematics assessment

P. Iannone[a] and A. Simpson[b]

[a] *School of Education and Lifelong Learning, University of East Anglia, Norwich, United Kingdom*

[b] *School of Education, Durham University, Durham, United Kingdom*

**Students' views of value and validity in undergraduate mathematics assessment**

There is a considerable body of research on students' perceptions of the value and validity of different modes of assessment, with a consistent message about those perceptions. However, an examination of the key empirical works shows that the main findings are based on views from a restricted range of students. Despite this, those findings are used to suggest changes in policy beyond this range, on the apparent assumption of the wide generalizability of these results. Notably, the research has tended not to hear the voices of students of the hard sciences. This paper examines the perceptions of students in one such area – mathematics at a high ranking UK University. Our findings suggest that, at least in the particular context of our research, the students' views are quite different. We suggest that this finding should lead researchers, academics and policy makers to exercise caution when basing specific recommendations on general findings.

Keywords: Assessment; Higher Education; 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). Much 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).

There is strong empirical evidence that students' perceptions of the value and validity of the assessment affect their learning (Scouller 1998), albeit that the interaction between those perceptions and the approaches to learning are far from straightforward. In their 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.

However, Segers, Dochy and Cascallar (2003) point out 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 the students' perceptions of these methods.

Struyven et al. (2005) produced a comprehensive review of the literature on students' perceptions of assessment in higher education. They suggested that the substantial body of research they reviewed contains a consistent message supporting the ideas that students have clear preferences in terms of assessment methods (for example, favouring multiple-choice tests over essay tasks) and that they are also often conscious of the ways they adapt their learning to their perceptions of the value and validity of the assessment.

The message appears to be that students view traditional forms of assessment (such as closed book examinations) as "arbitrary or irrelevant" (Struyven et al., 2005, p338) whereas innovative forms of assessment, such as projects and presentations, are perceived as fairer, because they "measure qualities, skills and competences which would be valuable in contexts other than the immediate context of assessment" (ibid, p. 339). This apparently consistent message from students has been used as a driver of change (Falchikov and Thompson, 2008; Meyer et al., 2010).

However, an examination of the literature underlying this message reveals that it comes from a particular range of voices. Table 1 shows the distribution of academic subjects (by paper and sample size) and year groups in the original research articles included in Struyven et al.'s review. Given the description of how the literature was selected and analysed, this review appears to be a comparatively unbiased and comprehensive reflection of the state of the art. It is clear that the major source of data in the research has been students on particular courses from arts, humanities and social sciences, with large science samples focused on applied subjects such as engineering and medicine, and that many subjects are unrepresented. Indeed, over 80% of the students surveyed across this literature come from just five subject areas (psychology, medicine, economics, education and engineering). In particular, only one study involved students from a pure, numerically-based science (physics), and that only with a small group of students.

[Table 1 goes here]

Thus, the message from the existing research comes from a restricted range of student voices, but the impact of the messages from that research has influenced subject areas not included in the research.

For example, in university mathematics in the UK, there has been a consistent push from those focused on pedagogy to move assessment in the direction advocated by this research – downplaying the role of closed book examinations in favour of innovations such as projects and presentations (Berry and Houston 1995; Houston and Lazenbatt 1996; Challis, Houston and Stirling 2004; Steen 2006). The prevalence of traditional assessments has been suggested as one source of the apparent link students make between being good at mathematics with having a good memory and their apparent lack of appreciation of conceptual understanding in mathematics (Schoenfeld, 1989).

Despite this push to innovate, professional bodies and individual academics point to the "special" nature of mathematics as reasons to challenge the prevailing direction of change (LMS, 2010) and a recent survey of assessment methods indicates that the traditional, closed book examination remains as the sole assessment method in the overwhelming majority of mathematics modules in almost all universities (Iannone and Simpson, 2011). Thus there appears to be a tension between current practice and the professional voice of the academic and the voices of the pedagogues and the research evidence about student views of assessment.

Is this, as Burton and Haines (1997) suggest, intransigence on the part of mathematics academics, or are there differences between the general message and the specific situation of mathematics?

There appears to be no existing empirical research into mathematics students' perceptions about assessment which might address this question. This paper thus seeks to examine whether mathematics students' perceptions of assessment methods follow the same pattern as that repeatedly highlighted in the general assessment literature and repeatedly advocated in works giving advice on university mathematics teaching and learning. To do this, we take a similar approach to the general assessment literature in this area and focus on the students' views of assessment as validly measuring mathematical ability and as embodying the values of the subject.

## Validity and Value in Higher Education Assessment

At its simplest level, we can define a form of assessment as valid if it 'measures what it set out to measure'. In many situations operationalising this can be quite complex and controversial (Messick, 1989), but since our aim is to investigate students' *perceptions* of validity, we adopt Borsboom, Mellenbergh and van Heerden's (2004) basic interpretation.

Each assessment task is intended to test some particular aspects of students' performance and understanding of the subject. For example, in mathematics, asking the candidate to state a named theorem may be intended to test declarative memory while asking for a proof of a statement which the student has not seen before may be intended to test conceptual understanding and logical reasoning. Thus we will use 'validity' in this paper to refer to whether the given assessment does indeed measure the aspects of learning that it was intended to measure.

However, assessment also reflects 'value': testing some characteristics of learning rather than others, differentially testing some characteristics more often than others, or testing characteristics using different methods will convey messages to students about what is deemed to be more or less important and thus reflects the assessors' perceptions of *value* within the subject and is likely to have a notable pre-assessment effect.

For example, Jankvist (2009) argues for the explicit teaching of the history of mathematics within mathematics modules at the university level. However, Tzanakis and Arcavi (2000) argue that it may be difficult to integrate assessment of history into modules and if there are no tasks focussed on the historical in that assessment then the possession of knowledge or cognitive skills associated with history will not play a part in distinguishing a good mathematician from a poor one. That is, the assessment would demonstrate that the assessor does not *value* the historical in mathematics assessment.

Using these notions of validity and value in mathematics assessment, this paper addresses the following research questions to explore the overarching aim of seeing whether the general trends in the higher education assessment literature apply in mathematics. Specifically, we ask:

- What do students perceive as the most valid forms of assessment of mathematical cognitive processes?
- What forms of assessment do students perceive as most embodying the values of mathematics?
- What is the relationship between students' perceptions of validity and value of different assessment forms?

**Methods**

Given that our overarching aim is to 'sense check' the findings of the general literature in the previously unexamined area of university mathematics, we adopt methods and sampling common to that general literature, but we focus on mathematics students at a high ranking university in the UK. This is a context which, as shown earlier, has a highly restricted assessment diet (Iannone and Simpson, 2011).

The study was conducted with 48 students studying for a mathematics degree. Noting that the modal student in existing research is in their first year (see Table 1), the students were chosen from two, randomly assigned seminar groups out of six for a core first year lecture course focused on mathematical problem solving. They completed a questionnaire looking at validity and value in different forms of assessment at the beginning of a seminar session.

The instrument 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). Whilst this is a very comprehensive approach to assess students' preferences 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 which are in widespread use or feature in the literature as potential forms of assessment in university mathematics in the UK (Table 2). In order to explore students' notions of validity, we chose two key cognitive processes involved in the learning of mathematics (Bergqvist 2007) and which feature in the design of the API developed by van der Watering et al. (2008): *memory* and *understanding*.

[Table 2 goes here]

The questionnaire had two sections. One section asked students to place in rank order the eight assessment methods listed in Table 2 according to the value 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'. The deliberate use of the value words 'good' and 'poor' without further definitions follows Messick's (1989) view of how assessment embodies the subject's values, without pre-determining what those values are. Note that, throughout the questionnaire, the name of the assessment methods was consistently accompanied with the description of the method in Table 2 to ensure that students understood the meanings of the terms.

The second section asked students to assess the validity of each method for how well they 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'.

The two sections of the questionnaire focused on value and validity were presented in random order on the questionnaires to address any potential bias for the order of the questions, and are presented in the Appendix.

## Analysis

### *Validity*

A Friedman Test was conducted to determine if the students had differentially rank ordered the validity they attributed to each assessment method as a test of memory. Results of the test indicated there was a differential rank ordering, $\chi^2(7)=165.19$, $p<0.001$ with critical difference of 0.66 (Conover 1980). We thus obtain a partial ordering of methods on mean ranks as shown in Figure 1, with the groups indicating where the rank orderings are significantly different.

[Figure 1 goes here]

Closed book examinations are clearly dominant in the students' views of valid assessments of memory. The ordering shown in Figure 1 is probably not surprising: those rated highly as tests of memory are assessments where one performs without explicit access to external information sources or assistance, while those rated lowest (such as dissertations, project coursework and open book examinations) are those which allow students time and access to resources.

An identical analysis was undertaken for students' views about valid measures of understanding. The Freidman test again showed a significant differential rank ordering, $\chi^2(7)=54.19$, $p<0.001$, with critical rank difference of 0.78. The partial ordering according to ranks is shown in Figure 2.

[Figure 2 goes here]

Two issues stand out from Figure 2. It is surprising to see a method which few, if any, students will have experienced in mathematics - oral examinations - at the top of the first group in this ranking and to see many more recently advocated innovations - such as presentations and projects - ranked by the students significantly lower as valid measures of understanding.

The link between students' views of the methods' validity as measures of understanding and memory was explored. A Wilcoxon signed rank test was conducted for each assessment form to give a measure of the extent to which the students perceived understanding as dominating memory for each method.

[Figure 3 goes here]

The only method in which memory significantly dominated understanding was closed book examinations. There was no significantly dominant cognitive process for multiple choice examinations and for all other methods, understanding significantly dominated memory in the view of the students. Note that this does not suggest that closed book examinations are a poor assessment of understanding (after all, they were ranked second as a valid measure of understanding) just that, across the sample, memory was seen as a *more* dominant aspect. In general, this does seem to suggest that students see a larger role for understanding in assessment than they do for memory, apart from in the traditional closed book examinations.

### *Value*

A Friedman Test was conducted to determine if the students had differentially rank ordered the extent to which they felt each assessment method embodied the values of mathematics. Results of the test indicated that there was a differential rank ordering, $\chi^2(7)=97.918$, $p<0.001$ with critical value 0.84, the partial rank ordering is given in Figure 4.

[Figure 4 goes here]

Again, it is surprising to see that an assessment method that has such widespread support in the mathematics assessment literature as an innovation – presentations – embodies so little of the value of mathematics in the views of the students. It is not simply the more traditional forms of assessment, however, which they think embody the values that differentiate good mathematicians from poor ones – projects and open book examinations rank quite highly.

### *Validity and value*

A further question we were exploring concerned the relationship *between* the validity and value of assessment methods for students, thus linking together measures of cognitive processes with the values of the subject. In particular, do students consider those assessments they view as testing understanding more than memory to be more reflective of mathematical values than those they see as assessing memory more than understanding? This entails considering the link between the validity measure of each assessment method for each cognitive skill (memory and understanding) with the value measure. For each student we calculated a Kendall $\tau_b$ as a measure of the strength of the relationship of the *difference* between the responses for memory-validity and understanding-validity with the ranking given for the each assessment method. We discarded data from four students who had given tied ranks for some assessment methods.

A one sample t-test showed that the mean of these relationships was significantly above zero, $t(43)=2.196$, $p<0.05$, representing a medium to large effect size ($d=0.67$). That is, across the sample, students tended to ascribe mathematical value to those assessment methods which they saw as more valid measures of understanding than those which were seen as more valid measures of memory.

**Discussion**

The overarching aim of our research was to ‘sense check’ the general literature on students’ perceptions of assessment to an area which had not previously been investigated – mathematics - to see whether the results did indeed generalize here or whether, as mathematicians and professional bodies seem to suggest, mathematics is somehow different. Our analysis suggests there are some notably different views amongst our sample from those in the general literature. For example, our participants ascribed considerable mathematical value to the stereotypical traditional assessment form, closed book examinations. While our analysis also shows that they consider memory to be a more important factor than understanding in this form of assessment, it was rated highly by the students for both of these cognitive factors and one might consider this to be an indication that they see closed book examinations as a well balanced form of assessment. In general, however, the analysis suggests that students tend to ascribe mathematical value to those forms of assessment which they perceive to be more validly measuring understanding rather than memory. This indicates that students do indeed appreciate the importance of conceptual understanding in mathematics. This finding seems to be at odds with much literature in mathematics education that suggests students do not have an appreciation of the importance of conceptual understanding and that they equate being good at mathematics with having a good memory (Schoenfeld 1989).

Similarly, what little research has been published in university level mathematics education has tended to be focused on promoting innovative forms of assessment (e.g. Berry and Houston 1995) and yet these are seen by the students in our sample to have little validity and seem to embody little of the value of mathematics. Projects and, particularly, presentations tended to score very poorly for both value in differentiating good from poor mathematicians and as valid forms of assessing mathematical understanding.

Berry and Houston (1995) argue that more innovative forms of assessment focused on communications such as project and poster presentations support "exposing and confronting misconceptions" and put "emphasis on concept as well as procedure" (p 22). Burton and Haines (1997) suggested that traditional forms of assessment are "reproductive and content dominated", that closed book examinations do not appear to "allow student to demonstrate … selection and use of mathematical facts, concepts and techniques" (p. 280) and that it is "the extremely narrow view held by many mathematicians about their discipline and its teaching and learning" (p. 287) which holds back change to innovate away from traditional forms. Our analysis suggests that our students do not see the assessment methods in this way: they do not see closed book examinations as narrow and reproductive, but as having a relatively good balance in validly assessing both understanding and memory and of having good value in differentiating good from poor mathematicians.

Equally surprising, though, is that our analysis suggests that there is one form of university mathematical assessment which is not used in the UK which does rank very highly as a valid assessment of mathematical understanding. Oral examinations were considered as the most valid form in this respect and, while seen as a relatively good assessment of memory, were not taken to have a dominant memory factor. Again, this stands in contrast to existing research: Birenbaum (1997) has both education and engineering students ranking oral examinations as the least favoured out of six categories of assessment types. While they have almost no role in assessment in mathematics in the UK, oral examinations are a dominant form of assessment in many other countries, such as Germany, Hungary, the Czech Republic and Italy and are valued as providing direct 'evidence of a candidate's ability for critical and reflective thinking.' (Kehm 2001 p. 27).

Taken as a whole, these results are quite distinct from those summarized in Struyven et al. (2005). Sambell et al. (1997), for example, argue that students have negative views of traditional forms of assessment, such as closed book examinations, and that they are "detrimental to the learning process" (p. 357). It is notable, however, that Sambell et al. did not include mathematics (or other pure sciences) students in their study, nor did they make clear the nature of the institution from which their sample came. Similarly, Struyven et al.'s (2005) review suggests that students do not value traditional assessments because of their perceived reliance on memory and "ability to marshal lists of facts and details" (p. 339). As previously noted, the 35 different studies in Struyven et al.'s review did not include any with a substantial mathematics or pure science focus (Table 1). Given the extent of their review and the clear exposition of their selection methods, it is unlikely that Struyven et al.'s analysis is simply mistaken or their selection biased from amongst the existing literature. Instead, it is more likely that higher ability students on mathematics courses have quite different views of value and validity, which may, of course, be related to their own past success with particular assessment forms.

It is also worth noting that we deliberately chose a boundary case since our overarching aim was to 'sense check' the results of the general literature to a different academic subject. That is, our sample consisted of students at a research-intensive higher education institution which demands the highest entry grades in mathematics for its students and whose employment criteria for academics has a strong focus on research. The development of the notions of value and of what distinguishes a good mathematics student from a poor one will be bound up with students' enculturation in mathematics. Skovsmose and Nielsen (1996) point out that even when values are not explicitly discussed by lecturers, students can infer them from their actions. Part of that enculturation is the messages students receive from mathematicians in the form of assessment tasks, and the extent to which they represent authentic activity in the subject mirrors the extent to which they represent mathematicians' values. Those implicit messages may be different (or differently received) in less research-intensive institutions or in different higher education cultures. However, the key issue is that (in this context at least) the generalist findings do not appear to be applicable and so the calls for change, based on that literature, may be inappropriate here.

It might be argued that the more traditional the subject and the university, the more likely that students are to hold very traditional views. However, as the case of oral examinations shows, our findings may not simply be a reflection of inherent conservatism or the students' own successful experiences with assessment methods but may be genuinely held beliefs about what forms of assessment focus on what they deem important and what forms embody the values of their subject. Most importantly, however, we believe the results call into question the extent to which results from generalist literature can be used to guide choices of assessment methods in specialist subjects and suggests that the views of professional bodies (such as the LMS, 2010) about the particular nature of a subject need to be heeded by policy makers when promoting change.

**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: 359-374.

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

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

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

Birenbaum, M. 1997. Assessment preferences and their relationship to learning strategies and orientations. *Higher Education* 33: 71-84.

Borsboom, D., G. Mellenbergh and J. van Heerden. 2004. The concept of validity. *Psychological Review* 111: 1061-1071.

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: 273-294.

Challis, N., K. Houston and D. Stirling. 2004. *Supporting Good Practice in Assessment in Mathematics, Statistics and Operational Research: Briefings and Guides*. Birmingham: MSOR.

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 .

Falchikov, N. and K. Thompson. 2008. Assessment: What drives learning? *Journal of University Teaching and Learning Practice* 5: 49-60.

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

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.

Iannone, P. and A. Simpson. 2011. The Summative Assessment Diet: How we assess in mathematics degrees. *Teaching Mathematics and its Applications*.

Jankvist, U.T. 2009. A categorization of the 'whys' and 'hows' of using history in mathematics education. *Educational studies in mathematics* 71: 235-261.

Kehm, B.M. 2001. Oral examinations at German universities. *Assessment in Education: Principles* 8: 25-31.

LMS. 2010. *Mathematics degrees, their teaching and assessment*, London Mathematical Society.

Messick, S. 1989. Validity. In *Educational Measurement,* ed. R.L. Linn, 13–104. New York: Macmillan.

Meyer, L. H., S. Davidson, L. McKenzie, M. Rees, H. Anderson, R. Fletcher, and P.M. Johnston. 2010. An Investigation of Tertiary Assessment Policy and Practice: Alignment and Contradictions. *Higher Education Quarterly* 64: 331–350.

Ramsden, P. 1988. Studying learning, improving teaching. In *Improving Learning: New Perspectives,* ed P. Ramsden, 13-31. London: Kogan Page.

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.

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

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

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, 1257–1288. 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: 331-347.

Tzanakis, C., and A. Arcavi. 2000. Integrating history of mathematics in the classroom: An analytic survey. In *History in mathematics education,* ed. J. Fauvel and J. van Maanen, 201-240. Dordrecht: Kluwer.

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: 645-658.

## Appendix

### A. Validity

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**<br>(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**<br>(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**<br>(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**<br>(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**<br>(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**<br>(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**<br>(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**<br>(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 | | | | | |

**B. Value**

Over the course of a year, imagine that students undertake a variety of assessments of their mathematical ability as listed below.
The department has to weight the outcomes of all of these assessments to give as good a measure of *mathematical ability* as it can, 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.
Please rank the different assessment from 1 (meaning that the department should put *most* weight on this assessment because it is most likely to distinguish good mathematicians from poor ones) to 8 (meaning the department should put the *least* weight on this assessment, because it is least likely to distinguish good mathematicians from poor ones)

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

Table 1. Academic subjects and year groups of studies in Struyven at al. (2005), by numbers of studies and total sample size.

| Subject | Number of studies | Total Sample Size |
|---|---|---|
| Psychology | 9 | 1389 |
| Medicine and biology | 7 | 690 |
| Accountancy, economics and finance | 6 | 587 |
| Education | 6 | 320 |
| Engineering | 5 | 627 |
| History | 4 | 55 |
| Social Sciences (unspecified) | 3 | 344 |
| Arts Humanities (unspecified) | 3 | 120 |
| Environment | 2 | 77 |
| Science (unspecified) | 1 | 128 |
| IT/computing | 1 | 12 |
| Languages | 1 | 12 |
| Law | 1 | 10 |
| Physics | 1 | 10 |
| Literature | 1 | 5 |

| Year group | Number of studies | Total Sample size |
|---|---|---|
| 1 | 13 | 2567 |
| 2 | 11 | 805 |
| 3 | 8 | 444 |
| 4 | 5 | 253 |
| Postgraduate | 4 | 62 |

Table 2. Taxonomy of assessment modes

| **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) |

Figure 1. Partial ordering of assessment methods – student views of valid assessments of memory

| Group 1 | Group 2 | Group 3 | Group 4 | Mean Rank |
|---|---|---|---|---|
| Closed Book | | | | 7.40 |
| | Oral exams | | | 5.89 |
| | Multiple choice | | | 5.57 |
| | | Project presentations | | 4.11 |
| | | Weekly example sheets | Weekly example sheets | 3.41 |
| | | | Projects | 3.27 |
| | | | Open book | 3.23 |
| | | | Dissertations | 3.13 |

Figure 2. Partial ordering of assessment methods – student views of valid assessments of understanding

| Group 1 | Group 2 | Group 3 | Group 4 | Mean rank |
|---|---|---|---|---|
| Oral exams | | | | 5.44 |
| Closed Book | Closed Book | | | 5.17 |
| Dissertations | Dissertations | | | 5.16 |
| | Weekly example sheets | Weekly example sheets | | 4.63 |
| | Projects | Projects | | 4.61 |
| | | Presentations | | 4.25 |
| | | Open Book | | 3.90 |
| | | | Multiple choice | 2.85 |

Figure 3. Student views of dominance of understanding over memory

| Direction | Assessment methods | | | | | |
|---|---|---|---|---|---|---|
| + | Dissertations $z=-6.023$, $p < 0.001$ | Weekly example sheets $z=-5.898$, $p < 0.001$ | Project $z=-4.952$, $p<0.001$ | Open-book exams $z=-5.394$, $p<0.001$ | Presentations $z=-4.952$, $p<0.001$ | Oral exams $z=-4.108$, $p<0.001$ |
| 0 | Multiple choice $z=-0.352$, $p=0.725$ | | | | | |
| - | Closed book exams $z=-2.149$, $p<0.05$ | | | | | |

Figure 4. Partial ordering of assessment methods – student views of value

| Group 1 | Group 2 | Group 3 | Group 4 | Group 5 | Group 6 | Mean ranks |
|---|---|---|---|---|---|---|
| Closed Book | | | | | | 5.70 |
| | Open Book | | | | | 4.56 |
| | Projects | Projects | | | | 4.19 |
| | | Weekly example | | | | 3.72 |
| | | Dissertations | Dissertations | | | 3.42 |
| | | | Oral exams | Oral exams | | 2.66 |
| | | | | Multiple choice | Multiple choice | 2.24 |
| | | | | | Presentations | 1.64 |