

The London Scholarship of Teaching and Learning Conference, 2006

Diverse approaches to assessment: an audit of current practice

ABSTRACT

This paper will present an analysis of an audit of over 250 assessments on five undergraduate courses in three universities. This related to the conference theme of practising the scholarship of teaching and learning. The audit was undertaken as part of the 'Towards Learning Creatively' (TLC) project. TLC is a three-year Fund for the Development of Teaching and Learning (FDTL5) project, based at three universities in southern England. It aims to encourage tutors to develop more diverse assessments, using non, or partly-written methods.

TLC focuses on the creation of more inclusive assessment methods, with particular attention on the needs of students with dyslexia. The project does not aim to advocate non-written assessment instead of written assessment, but to encourage tutors to review what and how they assess, and then to question whether written assessments are the most valid ways to assess intended learning outcomes.

It is generally agreed that assessment matters. It is fundamental to the way students learn (Brown, G., 2001; Brown, S. and Glasner, 1999). Moreover it may be argued that the traditional methods of assessment were applied uncritically in a Higher Education (HE) system that was elitist. Furthermore that the HE system has changed. HE has been part of a global shift to new ways of creating and using knowledge, focusing on problem-solving and the needs of the customer (Ramsden, 2003). HE is now for the many not the few; and learner centred rather than teacher centred (Jarvis et al, 2005).

The assessment audit revealed a traditional pattern of essays, reports, unseen examinations and seminar presentations, with significant differences between the use of written and non or partly written assessments, and between the universities. These findings will be compared and contrasted with those of the Higher Education Academy's assessment audit carried out in 2005. The primary research is then discussed in the context of changes in HE teaching and assessment perspectives, especially the apparent discrepancies between theory and practice (Elton, 2005; Ramsden, 2003; Elton and Johnson, 2002).

Diverse approaches to assessment: an audit of current practice

Liam Higgins, Southampton Solent University

Jenny Anderson, Southampton Solent University

Lyn Bibbings, Oxford Brookes University

Claire Griffiths, Southampton Solent University

Caroline Jackson, Bournemouth University

Introduction

The purpose of this paper is to introduce a Fund for the Development of Teaching and Learning (FDTL) 5 project 'Towards Learning Creatively' (TLC). Second, to present an analysis of audit findings about the types of assessment used in five hospitality, leisure, sport, and tourism subject area undergraduate courses at three universities. Lastly, to explore some of the reasons university tutors assess the way they do.

The TLC project's aim was to enhance the student learning experience by developing creative and inclusive modes of assessment which do not rely solely on written evidence. This aim does not advocate non or partly written instead of written assessment. Rather it wants to encourage tutors to review what and how they assess, and to ask themselves whether written methods are the most appropriate ones to assess the learning outcomes.

The project is part of a greater FDTL initiative, not to just provide techniques for teaching, but also to encourage an understanding of teaching, "... that it is what the student's do, not what the teacher does, is what really matters" (Ramsden, 2003: 10). The project was HEFCE funded 2004-7, involving three universities situated in southern England. There were two main elements: developing and trialing different types of assessment; and enabling lecturers to implement them.

More varied and inclusive approaches to assessment

The project team agrees with the argument that mass higher education and diversification has created a need for more varied approaches to teaching, learning and assessment. HE has changed: from the few to the many; from being teacher centred to learner centred. With this growing diversification of the student population there is a need for educational renewal and, in particular, more varied and inclusive approaches to learning, teaching and assessment (Jarvis et al, 2005; Elton and Johnston, 2002).

The project has focused on inclusive assessment, particularly for students with dyslexia. The reason for this is that about four per cent of all students in HE have dyslexia. and approximately half of those dyslexic students are diagnosed whilst in HE (CADISE, 2000; NWPDHE, 1996). Furthermore some subjects attract more students with dyslexia, for example outdoor recreation, and tourism. Dyslexia is not related to intellectual ability; rather dyslexic

students have different difficulties while studying in HE. Common difficulties are: reading, which is slow and difficult to understand; writing and spelling, especially with the sequencing of words; short-term memory; attention span; and organisational skills. These difficulties result in additional work and additional stress. However, dyslexic students may have strengths too, such as strong visual-spatial 'right hemisphere' strengths: holistic perception; lateral thinking; problem solving ability; creativity; and interpersonal and intuitive skills.

Assessment as measurement or as a judgment

Some HE tutors would argue that the use of written assessments is the most valid way of assessing undergraduates. All assessment judgments must be valid, and so any non or partly written assessment should be as valid as a solely written one. However, it could also be argued that because teaching in HE has been scrutinised from a standards point of view: by the QAA; the media; and by external examiners, the focus has been on assessment as 'measurement'. On the other hand, we argue that assessment ought to be a 'judgement' by interested and knowledgeable scholars as to what is the current 'best answer' and this may not be in the written form. In addition there are other external drivers for using more diverse assessments, such as the Special Educational Needs Discrimination Act, (2001) and Pt IV (2006); HEFCE Guidance on Base-level provision; and the QAA Code of practice for students with disabilities (Waterfield and West, 2002).

Multiple intelligences, learning styles and preferences

Traditionally, we argue, academic subjects have been taught in ways that largely involve two intelligences: linguistic and logical-mathematical (Gardner, 1993). Now consider what an IQ test basically measures: ability with words and numbers. Students who are naturally strong in linguistic and mathematical intelligences do well on the standard Stanford-Binet IQ test. Therefore it is a fairly good predictor of success at school, because the way we teach (lectures) and the material with which we deal (logically constructed books) depend heavily on these two intelligences. Since teachers are drawn from people who do well at school, it is a self-perpetuating system.

There are links with Learning Styles theory too. The Honey and Mumford, (1986) Activator, Reflector, Pragmatist, Theorist typology is very well known. However it is 'VARK': visual; aural; read/write; and kinaesthetic sensory learning preferences theory (Fleming, et al, 2005) that is most pertinent to this project team. VARK is a variation on 'VAK', a classic learning styles theory from the 1920's, used by psychologists and teachers (Keller, Orton, Gillingham, Stillman, Montessori) in relation to dyslexic students. The VARK learning preferences model does not overlay Gardner's 'multiple intelligences'; rather the VARK model provides a different perspective for understanding and explaining a person's preferred or dominant thinking and learning style. Gardner's theory is one way of looking at learning styles; VARK is another. But both approaches argue that for effective learning to occur, the messages

sent by the teacher must be in a form that is easily understood by the learner, and this may not always be effectively done in the written format.

Assessment as a cause for concern

In short, we argue that assessment in HE may be a cause for concern. Knight and Yorke, (2003) cited a 2001 survey of 23 (what was) LTSN subject centres, informed by the QAA subject reviews. There was concern for the assessment component of the Teaching, Learning and Assessment review. One particular problem was with the use of assessment practices that all staff would follow. There appeared to be ‘... many problems with assessment, including those of matching assessment methods to the range of intended learning outcomes,...’ (Knight and Yorke, (2003: viii). The authors went onto report criticism of the assessment practices and the fact that only one in three universities received a top grade of 4 for Teaching, Learning and Assessment. They argued that factors such as increasing student numbers, the need to promote and describe ‘... complex, fuzzy achievements that can be associated with graduate employability...’ (ibid: ix), and the increasing expectations of stakeholders in HE, have called into question the effectiveness of traditional assessment practices. They added that if at the primary and secondary school levels there is uncertainty about the reliability and validity of assessment, and there it has been done carefully, how much faith should we have in HE, where, “... understanding of assessment - especially what assessment *cannot* do – is likely to be less developed? (ibid: 16).

The Assessment Audit

When planning the project, the team decided to carry out an assessment audit, as it seemed that there was a general lack of data on it. For example, Race (1999: in Elton and Johnston, 2002:5) suggested that:

“... something like 90% of a typical university degree depends on unseen time constrained written examinations and tutor-marked essays and /or reports”

but Elton and Johnston added that 90% is presumably a rational guess, “... since quantitative research does not exist”.

The audit questionnaire was based on the work of Brown (2001) and the Bioscience LTSN (2005). However, there were other sources that could have been used. For example, the QAA refer to 47 different assessment modes, Knight and Yorke (2003) listed 51, and one university has over 150. The project partners piloted the audit in April and May, 2005, and the audit was carried out between June 2005 and May 2006. Purposeful sampling was used: partners selected five undergraduate courses at their universities that were thought to have high numbers of dyslexic students. All of the core units were audited plus as many of the option units as could be arranged. The main sources of information were the unit specification and annual unit reports.

Results and analysis

The method of analysis was carried out using both descriptive and inferential statistics. The assessments and whether they were wholly, partly or non written, were subjected to measures of difference such as ANOVA, so that any changes in assessment could be measured over the course of the three year project.

Fig i. The frequency of mode of assessment

127 units and 276 assessments were sampled. There were 30 types of assessment (Brown, 2001) listed in the survey. The most common types of assessment were: an essay; followed by an unseen exam; a student led seminar presentation; a report; and in-class test. These results support those of the Higher Education Academy's (HEA) National Survey of Assessment Practice in Hospitality, Sport and Tourism (HEA, 2005), which sampled 49 HE and FE colleges. It found that the same types of assessment were identified as most commonly used, but in a slightly different order to the TLC audit: reports; essays; exams; and presentations. It should be noted that the methodology was different: the HEA survey was conducted by telephone with mostly Course Leaders. In both surveys, the most common assessments types were in writing, with just one not solely written: student presentations.

The analysis of the audit data found no significant effects for: the units' CATS value; whether a core or option unit; the unit level; or the number of students or dyslexic students on the unit. However, there were significant differences in whether the assessment was written, non or partly written.

Fig ii University and use of written or non and partly written assessment

There was a significant effect ($p = 0.030$) of the university on the number of written assessments between A and B, as well as between A and C. In this sample, C gave significantly ($p = 0.000$) more non - written assessments than either A or B.

Fig iii ANOVA of university and individual assessment

An ANOVA found that for individual assessments there were significantly ($p = 0.000$) more in writing. There were no significant differences between the universities except for B, in partly written only.

Fig iv ANOVA of university and group assessment

For group assessments there were significantly ($p = 0.000$) more partly written group work than written and non-written, except for University C, which had significantly ($p = 0.001$) more non written group work than the other universities.

Discussion

To briefly recap on the assessment audit. Three universities were sampled, consisting of five undergraduate courses, made up of 127 units and 276 assessments. The most common types of assessment were: an essay; followed by an unseen exam; a student led seminar presentation; a report; and an in-class test. There were significant effects of the university and whether written, non or partly written assessment was used, as well on whether the assessment was group or individual work. This paper now moves onto discuss some of the possible reasons why tutors assess the way they do and why there may be differences between the universities.

Traditional approaches to assessment

It may be that tutors in HE assess the way they do because it is traditional: it is a habit (Biggs, 2003; Ramsden, 2003). Many tutors were taught and assessed using read/write approaches and therefore used them in turn on their students, and as a result the unseen exam has always been a mainstay in HE. Yet:

“... it would appear that traditional assessment, whichever form this may take in different disciplines, is often accepted uncritically, which is much less the case for innovative assessment (Elton and Johnston, 2002:4).

We would argue that a tutor's aim ought to be to get the learner's engagement with the learning task. One way to do this would be to move away from the more traditional approaches to assessment where, "... you have to assess everything that moves in order to capture student's time and energy" (Gibbs and Simpson, 2004:8).

Tutor's theory of teaching and approach to assessment

It may be that the tutor's choice of assessment is affected by the tutor's theory of teaching. For instance, Ramsden (2003) modelled three hierarchical theories of teaching: 'telling'; 'organising student activity'; and 'facilitating learning'. Teaching as 'telling' or transmission tends to be positivist in its approach, and often uses exams and essays for summative assessment. This tends to result in surface learning by students, which in turn affects the teaching: the so-called 'backwash' effect. Biggs (2003) asserted that with this approach, there may be a genuine belief in the 'measurement' model ('my job is to sort the sheep from the goats') rather than the 'judgement' model, which represents an acceptance of valid answers by a process of 'connoisseurship' (Knight and Yorke, 2003). Teaching as 'organising student activity' is a mix of positivist and constructivist approaches, and often uses group work. Teaching as 'facilitating learning' is constructivist in nature, and it often uses formative assessment. As a result it tends to encourage a student-focused approach to teaching, resulting in deeper learning.

The need for validity and reliability

On the other hand, tutors may argue that their choice of assessment method is the most valid and reliable one. Indeed, the assessment literature insists that both 'measurement' and 'judgment' should be valid. The main issue here is that in education it is seldom possible to measure exactly what learning has taken place. Assessors are forced to measure a 'surrogate' for what they are really interested in (Jarvis, et al. 2005). This has led to lots of concerns about how valid assessments are. For example, according to Knight and Yorke (2003), assessments should engage with the programme or module's learning intentions and not with pallid, easy to handle simplifications.

In addition to being valid, the assessment literature stresses the need for assessment instruments to be reliable. The reason why is that generally, assessments are applied to large groups of people, not just an individual, and so needs to be fair. However, research has consistently suggested that marker reliability is low, except when pure knowledge recall is tested. This is probably because the markers lack training in assessment (Elton and Johnson 2002). Reliability, then, really matters with summative assessment, and this is not always easy to do, as:

“... reliability is greatest when assessments are simplest. There is a problem, then, if we aim to engender complex learning, because assessment processes that tend to simplify will undermine it” (Knight and Yorke, 2003:18).

Due to the prevalence of a positivistic approach to assessment, where there is, for example, a reliance on statistical models of distribution, dominance has been given to reliability over validity: 'fairness' over educational relevance, of 'doing the wrong things righter' (Elton, 2005). A constructivist approach to learning and assessment challenges the assumptions made by a positivist approach. For example, what is the 'right' answer? A questioning of the arbitrariness of divisions in degree classes, where over 70% is a First, and 69% is a 2:1; and the assumption that marks should be normally distributed, since good teaching ought to legitimately skew the distribution (Elton and Johnson, 2002).

To add to the complexity of dealing with assessment, it should be noted that reliability and validity are not mutually independent from each other. In high stakes assessment, as in degrees, there are stakeholders, for example the QAA, and employers, who demand reliability yet:

“Unfortunately, these are the very objectives which are of comparatively little importance in degree programmes, where it is usual to expect higher abilities and skills to be developed, and where increasingly learning programmes differ for different students” (Elton and Johnson, 2002: 28).

Thus there has to be a trade off between reliability and validity. For example, exams are often used to deter plagiarism in higher education, resulting in a testing of recall rather than higher order thinking: description rather than critical analysis.

Tutors' attitude to change

So how may the differences between the universities and their use of non, partly and solely written forms of assessment be explained? It could be argued that the explanation is rooted in their tutors' attitude to changing their assessment practices. An example would be where a change in assessment would conflict with a university's assessment regulatory framework (Elton, 2005; Elton and Johnston, 2002:5). They refer to the Dept of Education and Skills, HEFCE and the QAA as the most positivist of assessors. However the QAA, at least, may be changing its approach. The latest Assessment Guidelines (2006) includes more advice about using diverse and formative assessment. In addition, professionals often "... resist changes that are perceived to threaten their core values and practices" (Pennington, 2003:7). In academia the espoused core values tend to be a commitment to collegiality, and may be rooted in the subject discipline's own approaches to assessment (Serow, 2004; Silver, 2003).

Barriers to changing assessment

There may be barriers to change in assessment practices. For instance, there may be internal barriers such as: a perception of how a HE tutor should assess; a lack of understanding of teaching and assessment theory; and a lack of intrinsic motivation to change, because of the fear of failure or the fear of judgement by one's peers. Or there may be external barriers, such as the

university's and professional body's assessment policies; the available resources (time, and rooms) and having to teach large groups of students.

The context

Many educationalists agree that when assessing students it is the context that is central (Jarvis, et al, 2005; Brown, 2004; Biggs, 2003; Ramsden, 2003; Elton and Johnson, 2002). The cultural background of the university and of the subject discipline particularly influences this context. The interaction between the university's culture and the subject discipline's approach to assessment may result in tutors being what Pennington (2003:9) termed: 'Rational adopters', individuals who respond to new ideas by analysis, discussion and evaluation; or 'Pragmatic sceptics' who need proof that the change works before adopting it; or 'Resisters/Defenders' who are unconvinced about and actively work against a change being adopted and embedded. We can only speculate as to the size and nature of these groupings in the respective universities sampled, but the audit results and analysis do seem to suggest that there are differences in tutors' approaches to the use of non, partly or solely written assessments, as well as the mix of individual and group work.

Conclusion

The TLC project is concerned with encouraging the use of non or partly written assessment, that is inclusive. However, an audit of assessment at three universities suggested that assessment was mainly written, and that the use of non and partly written assessment differed between the universities. These results support the view that whilst there is a rhetoric about using diverse assessment (Elton, 2005; Brown, 2004; Biggs, 2003; Ramsden, 2003), the practice seems to be to mainly use traditional written methods.

One reason for this may be that the learning preferences of the tutors do not necessarily match that of the learners. Tutors, who tend to prefer read/write assessment methods, may not be understood well by learners who often prefer visual or kinaesthetic communications (Fleming et al, 2005). Another reason for the different approaches to using more diverse assessment may be rooted in the context that it takes place in: the interaction between the culture of the university department and the subject discipline may affect the choice of assessment (Jarvis et al, 2005; Serow, 2004; Silver, 2003; Elton and Johnson, 2002).

As a result of this audit, the project may go on to research tutors' attitudes towards using non or partly written assessment. To research what, if any, are the qualitative differences between tutors' attitudes to assessment at the different universities.

References

- BIGGS, J., (2003) *Teaching for Quality Learning at University* 2nd ed. Buckingham: Open University Press.
- BROWN, G., (2001). *Assessment: a guide for lecturers*. LTSN Nov 2001. http://www.heacademy.ac.uk/resources.asp?process=full_record§ion=generic&id=3. Accessed 7/01/06
- BROWN, S., (2004). Assessment for Learning. *Learning and Teaching in Higher Education*. Issue 1. Pp 81-89.
- BROWN, S., and GLASNER, A (Ed), (1999). *Assessment matters in Higher Education: choosing and using diverse approaches*. Buckingham: SRHE.
- CADISE (Consortium of Arts and Design Institutions in Southern England), (2000). *Dyslexia in HE Art and Design: HEFCE Dyslexia Project Report*. www.cadise.ac.uk/interest_groups/dyslexia_index.htm
- Elton, L., (2005). *Changing assessment – danger of doing the wrong thing righter*. Student Staff Partnership for Assessment Change and Evaluation (SPACE) conference, 9 November, 2005, Plymouth University.
- ELTON, L. and B. JOHNSTON, (2002). *Assessment in universities: a critical review of research*. York: LTSN Generic Centre.
- GIBBS, G. and SIMPSON, S., (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*. Issue1. 3-31.
- GARDNER, H., (1993). *Frames of mind: the theory of multiple intelligences*. 2nd ed. London: Fontana
- HIGHER EDUCATION ACADEMY, (2005). National Survey of Assessment Practice in Hospitality, leisure, Sport and Tourism. <http://www.hlst.heacademy.ac.uk/projects/benchmarking.html> Accessed 7/01/06
- JARVIS, P., HOLFORD, J., and C. GRIFFIN, (2005). *The theory and practice of learning*. 2nd ed. Oxford: RoutledgeFalmer
- KNIGHT, P., and M., YORKE, (2003). *Assessment, Learning and Employability*. Maidenhead: Open University Press.
- LTSN Centre for Bioscience, (2005). Assessment Audit Tool. <http://www.bioscience.heacademy.ac.uk/resources/Audit.htm> Accessed 7/01/06

- NWPDHE (National Working Party for Dyslexia in Higher Education), (1999). *Dyslexia in HE ; policy, provision and practice*. Hull: University of Hull
- PENNINGTON, G., (2003). *Guidelines for promoting & facilitating change*. York: LTSN Generic centre
- RAMSDEN, P., (2003). *Learning to Teach in Higher Education*. London: RoutledgeFalmer
- SEROW, R., (2004). "The culture of undergraduate teaching at research universities." *Innovation in Higher Education*. 27 (1), 25-37.
- SILVER, H., (2003). "Does a University have a culture?" *Studies in Higher Education*. 28 (2), 157-169.
- SPACE (Student Staff Partnership for Assessment Change and Evaluation project). <http://www.space.ac.uk/assess.php>
- WATERFIELD, J., WEST, B. (2002). *SENDA compliance in Higher Education. An audit and guidance tool for accessible practice within the framework of teaching and learning*. University of Plymouth: South West Academic Network for Disability Support (SWANDS)

Fig i The frequency of mode of assessment

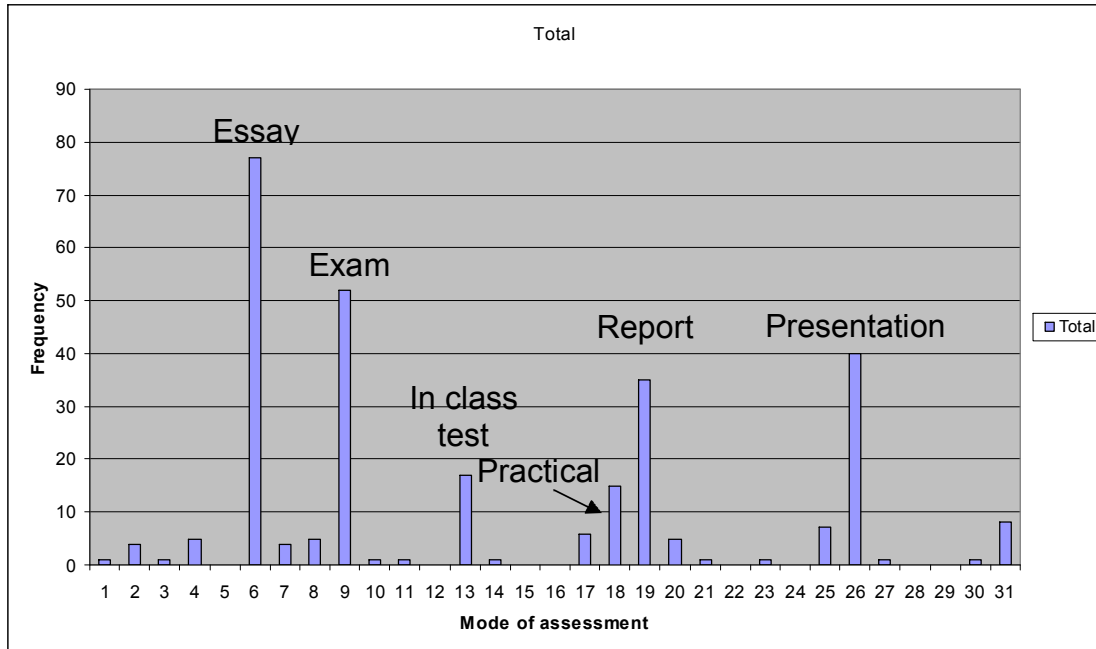


Fig ii University and use of written or non and partly written assessment

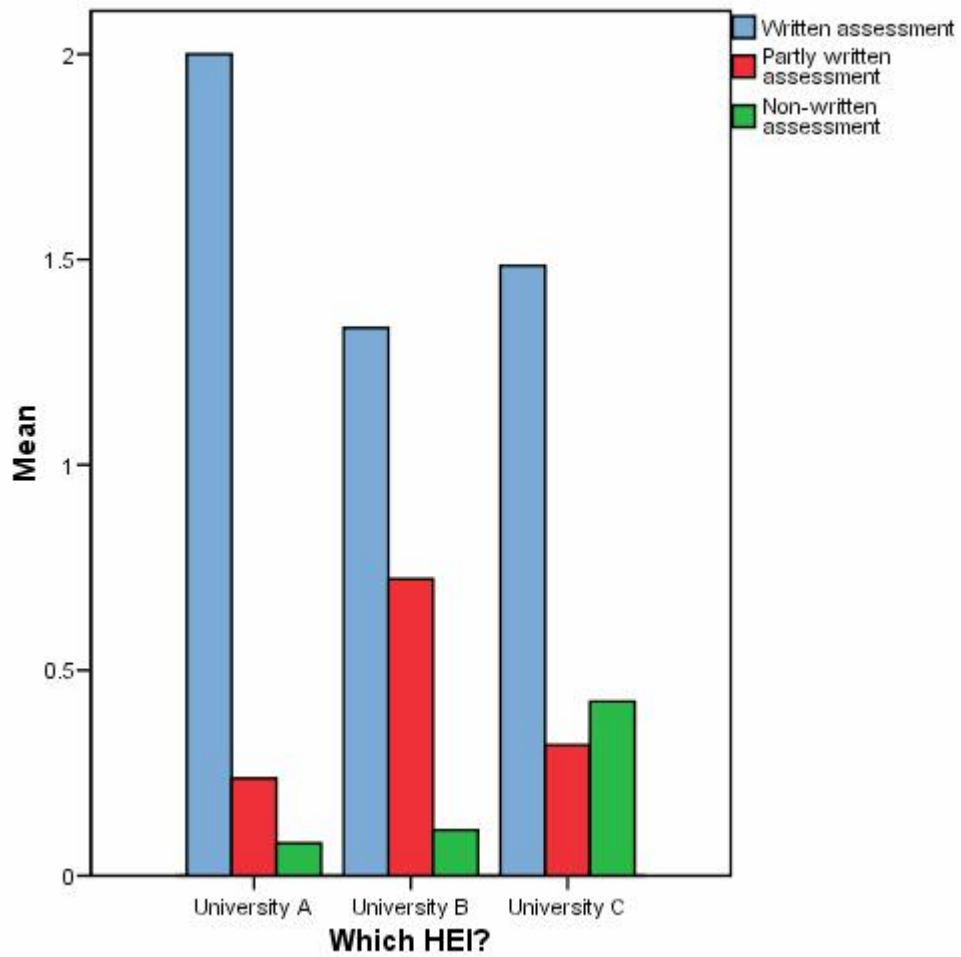


Fig iii ANOVA of university and individual assessment

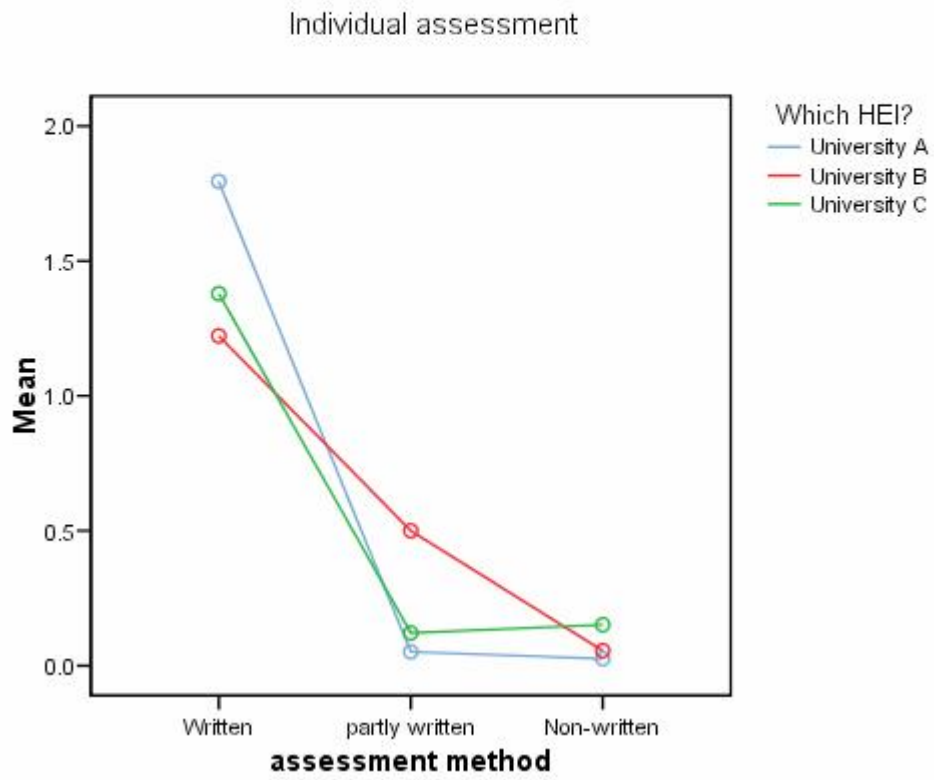


Fig iv ANOVA of university and group assessment

