

Essays are not the only way: A case report on the benefits of authentic assessment

SIOBHAN B. G. MACANDREW¹ AND KATHERINE EDWARDS
University of Abertay Dundee

Authentic assessments resemble the tasks that graduates will be required to perform in the workplace. This report compares grades obtained for essays with those for two authentic assessments, namely, the presentation of a poster and the publication of an information pack. Implementation of authentic assessments requires carefully devised assessment criteria, cooperative staff, and students who have been coached on the requirements of style and content. The authentic assessments reported here are valid, reliable, fulfilling and capable of motivating students to achieve. The results are set in the context of marking workload, didactic benefits and student and lecturer job satisfaction. The data illustrate some of the difficulties encountered in 'action research' in the classroom.

INTRODUCTION

You are likely to spend at least 240 hours grading papers every academic year. That represents five years of your career. Most of what you read will take the form of essays. But how worthwhile is this activity?

There is growing unease about the practice of setting and marking traditional essays. Essays are prone to plagiarism and labour intensive to mark. Moreover, variety in assessments would show potential employers that our graduates have a range of talents. In particular, 'authentic' assessments resemble the tasks that graduates will be required to perform in the workplace. Such assessments are aimed at improving the goodness of fit between a student's degree outcome and their likely performance in a job.

We provide an evaluation of two authentic assessment methods in comparison with traditional essays. The authentic assessments were devised by the first author and have been in use for more than eight years. We compared students' mean grades for two assessed essays on cognitive and on social psychology with grades for two authentic assessments. These were the preparation and presentation of a poster on psychology of language and the preparation and publication of an 'information pack' for sufferers of a neuropsychological condition. We embarked on our study hoping to provide clear empirical evidence on the efficacy of authentic assessments. The data provide intriguing hints, but there are severe limitations since we had little control over some important variables. This report is offered as a record of our experience. We hope it will facilitate others' experimentation with authentic forms of assessment.

Characteristics of posters

In this type of assessment the student prepares and presents a poster of similar specification to those employed at academic conferences. Students display their work at a meeting convened for the purpose in

front of an audience that usually comprises their peers and at least one lecturer. Students are allocated approximately five minutes to describe their work and answer questions posed by the lecturer. The presentation is usually recorded on videotape. This provides a record of the work, facilitates external examiners' and second markers' inspection, and allows more useful feedback to the student. Students choose the topic for the poster based on their own interests. The material required to produce the poster is constrained only in terms of surface area and style. Our posters are usually approximately 125 cm x 125 cm. The word limit for our third-level posters is approximately 3,000 words. The time allowed to complete the assignment is approximately eight weeks.

Characteristics of 'information packs'

An information pack consists of an organised body of information written with a specific reader in mind. The pack must explain academic literature in a way that is comprehensible to the chosen reader. The writer of the pack must strive to create a product that is highly informative yet engaging. The pack is intended to educate the reader in an effective and interesting way. Examples of this would be a neuropsychology student producing a information pack about Parkinson's disease to be used by sufferers and their carers, a psycholinguistics student producing a pack about aphasia aimed at student nurses, or an honours project student publishing a pack for the University library about how to complete a psychology project. As with posters, students choose the topic for the pack based on their own interests. The material required to produce the pack is completely unconstrained in terms of type or format. On rare occasions, students have chosen to produce a web site, 'television documentary' or 'radio programme'. Students with visual impairments or dyslexia are attracted to the option of producing work in audio format. The word limit for third-level information packs is approximately 3,000

¹ Correspondence concerning this article should be addressed to Siobhan MacAndrew at: Division of Psychology, University of Abertay Dundee, Marketgait House, 158 Marketgait, Dundee DD1 1HG. Email: s.macandrew@tay.ac.uk

words. The time allowed to complete the assignment is usually eight weeks.

Advantages of authentic assessments

The most powerful arguments for such assessments are didactic. Authentic assessments are student-focussed. The student must decide what they are interested in, what the issues are, and how to present the information. Both packs and posters require a different way of *researching, thinking about, and presenting* academic material. Students must have 'ownership'¹ of their knowledge to be successful. In essence, they are discovering what it is like to participate in the research community.

There is less opportunity for or advantage in plagiarism. There is no point in copying another student, as everyone is working on different things. Nothing is gained from copying directly from a text or article since one of the assessment criteria is that academic information must be 'understandable to the general public'. The only way to achieve this is by understanding the original source and rendering the arguments into non-technical language. Of course, we are not arguing that these assessments are cheat-proof, only that cheating may turn out to be more effort and less likely to be rewarded than doing the work.

Students report a feeling of 'discovery' in completing posters and information packs. They contrast this with the experience of 'second guessing the lecturer' that they associate with answering an essay question. Traditional essay questions often promulgate a notion of authority and the expectation of a 'right answer'. As a consequence, students may become disempowered as learners. Our students report being 'interested', 'fulfilled' and 'excited' by the work they produce in posters and information packs.

Potential disadvantages of authentic assessments

The main arguments against authentic assessments are administrative. Many lecturers prefer to have control of the questions tackled by students. They argue it is 'fairer' and also easier to establish the level at which a student is performing. This is particularly true of those who use model answers in grading. A further concern for markers is that every student submits a qualitatively and quantitatively different piece of work. These points are underscored by the tension between *norm-referenced* and *criterion-referenced* grading.

¹ *Ownership* of knowledge implies that the student is sufficiently well informed and confident about academic material that they can manipulate this information appropriately in response to a variety of circumstances. A student who does not own information for an assessment relies excessively on quotes from textbooks, has great difficulty paraphrasing, and slavishly presents information in the same order, style and format as they have read in a textbook or article. A student who owns information for an assessment does not rely on quotes, paraphrases with ease, and can confidently turn any information to the relevant purpose.

In norm-referenced grading, students' scores are related to the norm group of everyone taking the assessment. As raw scores are plotted this is sometimes called 'grading on a curve' (Ormrod, 1995). A normal distribution is assumed. The majority of candidates are given an average grade and there are few high and low grades. If the whole class performs well, a good score could mean a poor grade (the opposite would be true in a class of low achievers). It is also difficult to assign grade boundaries fairly. As the grade is related to others' performance, it gives little information on a student's progress, or how they could improve. This is to be contrasted with criterion-referenced grading, which assesses a student's mastery of the assessment criteria unrelated to others' performance. In theory, anyone could get an 'A' so long as they fulfil the assessment criteria. Since the student is aware of the criteria for each grade in advance, their goals in completing the work should be clear. Criterion-referenced grading relates the eventual score to the learning outcomes. This makes the purpose of the assignment concrete and facilitates feedback to the student. In contrast to norm-referenced grading, it is meaningful for the student to monitor their own progress by contrasting their marks over time.

Interim summary

Authentic forms of assessment facilitate students' ownership of knowledge, encourage them to take responsibility for their own learning, and are consistently rated as enjoyable and fulfilling. However, their use requires care in devising assessment criteria and in ensuring students and markers share the same vision of what is required. Traditional essays have a long history and allow the lecturer to have complete control over the material the student must know to fulfil the assessment. A corollary of this is that students describe writing essays as 'playing a sort of game' in which you second-guess the lecturer to get a good grade.

We are not arguing that it is impossible for traditional essays to provide the benefits observed with authentic assessment methods, only that it is less likely. An easy way of summarising the contrast between traditional and non-traditional assessment is that traditional methods focus on the *product* of the assessment whereas non-traditional methods focus on the *process* of completing the assignment.

A CASE STUDY FROM ABERTAY

Background information on University of Abertay degree structures

Our institution is a small Scottish University that gained its charter in 1994. We teach in two semesters and have modular degrees. Students take ten modules per year for four years. Most modules are assessed by at least one essay and at least one examination, so the average Abertay student has more than forty instances of assessment in their degree. There is a wide choice of modules, spread of student ability and

variety of student past experience, and also a large number of local and mature students.

Background to the observed assessments

Students attend workshops to explain the assignments and the assessment criteria. It is made clear that this involves the same amount of work and academic content as an essay. They are given a handout and a copy of the *Powerpoint®* lecture presentation explaining the assessment (available from first author).

In the poster workshop, they are told this is a standard form of presentation at psychology conferences. They are given examples of submission criteria (for example, those of the Cognitive Science Society). The students present their posters to their peers at a

conference and a video recording is made for the external examiner.

In the information pack workshop, they are told that the pack is a source of reliable facts from academic literature made understandable to the sufferer of a neuropsychological condition or their carers.

What are the assessment criteria?

The criteria were developed during discussions with lecturers and students. The process of deciding the key components of the assignments had a huge didactic benefit and later assisted the students in preparing their work. Students write essays from childhood and are familiar with the territory. The discussion allowed them to see that the authentic

PS346 Psychology of language poster mark sheet	
Student name:	Noam Chomsky
Title of poster:	Do baby chimps learn language like baby humans?
Markers:	Siobhan MacAndrew, A. N. Other
Research	
Has student read sufficiently widely?	No - Only one textbook chapter
Has student chosen appropriate literature to present their case?	No - Relies on tabloid newspaper reports
Does the research material include sufficient academic material?	No
Does student understand the literature mentioned?	No relationship between studies mentioned and structure of argument
Is there over-reliance on quotes and little input in student's own words?	Yes
Is there over-emphasis on non-academic web sites?	Yes
Presentation	
a) Comprehensibility	
Does the poster have an appropriate title?	Yes
Are the aims of the poster clear?	Yes
Are the aims of the poster met?	No, rather rambling and doesn't get to the point
Does the student explain concepts well?	No, see above and doesn't explain terms
Is the structure coherent?	No
Is the content superficial?	No
b) Appearance	
Is the poster easy to read?	Yes
Is the structure clear?	No - argument meanders through the text
Is the structure reflected in visual organisation of the poster? (i.e. not visually confusing)	No - hard to work out what order to read the panels
Is there appropriate use of definitions, tables, graphs, illustrations?	No
Are there references to academic material in text and a reference list in APA format at the end?	Yes, but not enough of them
Comments on particular strengths and weaknesses	
Promising idea, but rather rushed - did you really work on this for 7 weeks? Arguments are a bit rambling. Need to read more and not just introductory textbook. Even one journal article would have helped the grade. Have more confidence to put things in your own words!	
Agreed grade: 8 (Third class)	

Figure 1
An example of a marking schedule for a poster

PS347 Neuropsychology Mark sheet	
Student name:	<i>Alois Alzheimer</i>
Title of poster:	<i>There is help, there is hope: A guide for carers of people with dementia</i>
Markers:	<i>Siobhan MacAndrew, A. N. Other</i>
Objectives specific to this assessment: Does the pack contain	
What is the cause of the disorder?	<i>Excellent and clear</i>
What are the consequences for daily living?	<i>Carefully thought out</i>
Explanation of academic research clearly in terms the person will understand?	<i>Good</i>
Interpretation of what experimental findings may mean for “everyday life”?	<i>Good</i>
Ideas for rehabilitation based on psychological knowledge?	<i>Very creative!</i>
Is the pack appealing and interesting?	<i>Beautifully presented and entertaining</i>
General skills:	
Research	
Has the student read sufficiently widely?	<i>Excellent</i>
Has the student chosen literature appropriate to their case?	<i>Yes</i>
Does the pack contain sufficient academic material?	<i>Yes, (almost too much!)</i>
Does the student understand the literature mentioned?	<i>Yes</i>
Is there over-reliance on quotes and little in student’s own words?	<i>No</i>
Is there over-emphasis on non-academic web sites?	<i>No</i>
Presentation	
a) Comprehensibility	
Are the aims of the pack clear?	<i>Yes</i>
Are the aims of the pack met?	<i>Yes</i>
Does the student explain concepts well?	<i>Yes, writes very well</i>
Is the structure coherent?	<i>Yes</i>
Is the content superficial?	<i>No</i>
b) Appearance	
Is the pack easy to read?	<i>Yes</i>
Is the structure clear?	<i>Yes</i>
Is there appropriate use of definitions, tables, graphs, illustrations?	<i>Yes, excellent glossary, good graphs and even cartoons!</i>
Are there references to academic material in the text and a reference list in APA format at the end?	<i>Yes</i>
Comments on particular strengths and weaknesses	
<i>An excellent blend of academic findings and practical suggestions to help. Very well written. Researched widely, even beyond psychology. Very creative. Almost too much reading done – take care to pace yourself! Has fulfilled all the assessment criteria – well done!</i>	
Agreed grade:	<i>2 (Mid First)</i>

Figure 2

An example of a marking schedule for an information pack

assessments could be decomposed into recognisable requirements. Both types of assessment focus on quality of research, writing style and presentation. In addition, markers of posters evaluated whether the presentation effectively showed the link between data and theory. Markers of the information packs evaluated whether the pack explained causes of the

condition, the consequences for cognition, and the effect on everyday life of the person. Examples of completed marking schedules are provided in Figures 1 and 2 (note that in the poster conference students were not graded on their oral presentation or ability to answer questions).

Unfortunately, we could not use an identical marking schedule to grade the essays. The university did not permit this, as they have a standardised essay marking scheme for all disciplines. However, our markers were familiar with marking both authentic assessments and essays at Abertay and strove to be consistent. This represents one of the variables that was out of our control and reminds us of the difficulties of carrying out 'action research' in the classroom.

The research questions

We addressed three main research questions. Do students obtain higher grades for non-traditional assessments than for traditional ones? Are posters and information packs functionally equivalent with respect to what they assess? Lastly, are any students disadvantaged by their use?

Data collection and analysis

The essay questions chosen for the study are in key areas of the psychology curriculum and represent assessments typical of most UK psychology departments. Our external examiner confirmed this judgement.

For academic years 1999-2000 and 2000-2001 we collated the grades of all students taking any or all of the four observed modules (third-level Social Psychology, Cognition, Psychology of Language and Neuropsychology). This amounted to 177 students. We edited the set to include only those students who had done all four modules. This left us with 50 students¹. We took the mean grade of a student's cognitive and social psychology essays. This was to get a score of 'general essay ability' unhampered by preference for social or cognitive psychology. We also noted the Grade Point Average (GPA) of each individual. This is the aggregate of ten module grades from second-level studies. We assumed that the GPA is a measure of 'general academic ability'. It is not an ideal measure, but is the only one available. Finally, we noted the degree programme of each student.

We carried out a multiple linear regression of GPA versus three independent variables of mean grade for essays, grade for poster, and grade for information pack. We also performed a 3 x 3 ANOVA with three levels of the between-subjects variable of degree programme (these were behavioural science, psychology with biology, and psychology with computing) and three levels of the within-subjects variable of assessment type (these were essay, poster, and information pack).

¹ There is a wide choice of modules in our degrees, hence the difficulty of obtaining high numbers of students taking identical modules. In addition, some parts of assessment were incomplete for some students at time of writing which further eroded numbers.

Table 1

Correlations between measures	
Comparison	Correlation
Grade Point Average – Essay	.37*
Grade Point Average – Poster	.51*
Grade Point Average – Pack	.31*
Essay – Poster	.30*
Essay – Pack	.41*
Poster – Pack	.32*

Results

GPA correlated significantly with all three independent variables of essay, poster and information pack grade. These variables also correlated with each other. This suggests that the authentic assessments are similar to essays as a measure, and that all three types of assessment have a relationship with GPA (see Table 1).

Overall the regression model was significant ($F(1,48) = 7.40$, $p < .005$). Interestingly, the only significant predictor of GPA is the information pack ($t = 3.1$, $p < .005$, $R = +.341$). See Table 2 for details.

Table 2

Multiple linear regression: GPA versus essay, poster and information pack grades

	Coefficient	t value
Essay	0.148	1.25
Poster	0.076	0.99
Pack	0.341	3.08**

In the ANOVA there was no main effect of degree programme, but there was a significant main effect of assessment type ($F(2,47) = 5.97$, $p < 0.005$). A Neuman-Keuls *post hoc* test showed the significant

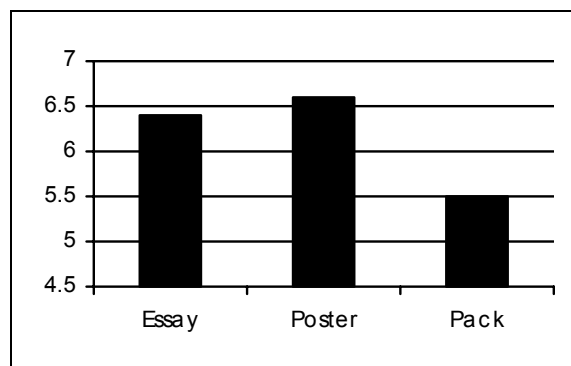


Figure 3

Mean grades for essays, posters and information packs.

Note that our University uses a grading system that ranges from 1 (top first class) to 18 (no work submitted). Grades of 1, 2, & 3 are first class, 4 & 5 are upper second, 6 & 7 are lower second, 8, 9 & 10 are third class.

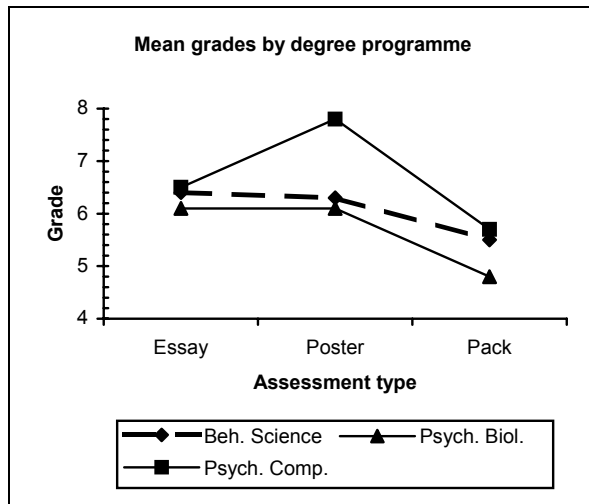


Figure 4

Mean grades by degree programme.

Note that our University uses a grading system that ranges from 1 (top first class) to 18 (no work submitted). Grades of 1, 2, & 3 are first class, 4 & 5 are upper second, 6 & 7 are lower second, 8, 9 & 10 are third class.

differences are between the essay and information pack grades and between the poster and information pack grades. There was no interaction of degree programme and assessment type. Figure 3 shows the mean grades for each assessment type, and figure 4 shows the grades broken down by degree programme.

DISCUSSION

Let us return to the research questions. First, do students obtain higher grades for innovative assessments than for traditional ones? In this demonstration they do achieve higher grades for information packs, but their performance on posters is more erratic. Generally, information packs are significantly the best 'retrodictors' of GPA, so we could say that this is a good way of identifying able students. However, this is premature until the project is complete and we can compare information pack grades with exit awards. Our next question was whether posters and information packs are functionally equivalent with respect to what they assess. Grades for posters were lower than for the other two types of assessment. Identifying the cause of this leads us to the issue of whether students were disadvantaged by the use of authentic assessments. Surprisingly, psychology with computing students performed poorly on posters (see figure 4). The ANOVA showed this difference was not statistically significant, but nevertheless we find this cause for concern. Since posters are more 'authentic' to the workplace vocational students might be expected to excel. The psychology of computing students have 'learnt the game' of essays but were bewildered by the poster assignment. They remained poor communicators and inflexible in their approach to learning, but this could be addressed by specific training.

This also illustrates a feature of our data that make them less generalisable. Our degrees are modular; therefore every class contains students from at least three degree constituencies. It is intriguing to speculate what might have happened had we carried out the same study with a class of psychology students.

The design of this study was retrospective so we did not have full control over all the variables affecting the outcome. Clearer results would be generated by a longitudinal design following students from only one non-modular degree programme, with all grading done using the same criteria and markers and with reliability checks made on all grades.

What are the take-home messages?

The regression analysis shows that the information pack is a reliable test of 'general ability' and the ANOVA indicates that this assessment produced the highest grades overall. Posters and information packs are discovery-oriented, reliable, less prone to plagiarism, and facilitate students in achieving their best performance. We hope these findings inspire more people to use authentic assessment and give heart to those who are already doing so.

ACKNOWLEDGEMENTS

Siobhan MacAndrew is an LTSN Psychology specialist coordinator for innovative assessment. This paper reports interim results of a miniproject funded by LTSN Psychology on which Katherine Edwards worked as research assistant. We would like to thank students and staff at the University of Abertay for their cooperation in this research.

REFERENCE

Ormrod, J. E. (1995). *Educational Psychology: Principles and applications*. Englewood Cliffs, New Jersey: Merrill.

FURTHER READING

For the sake of space, this paper does not introduce you to the literature in this area. The following list gives useful information that should be readily available in your institution's centre for learning and teaching.

Brown, G. (2001). *Assessment: A guide for lecturers*. LTSN Generic Centre Assessment Series, York, UK: LTSN. (Available from www.ltsn.ac.uk/genericcentre)

Brown, S. and Glaser, A. (1999). *Assessment matters in higher education: Choosing and using diverse approaches*. Buckingham: The Society for Research into Higher Education & Open University Press.

Biggs, J. (1999). *Teaching for quality learning at University*. Buckingham: Open University Press.

Heywood, J. (2000). *Assessment in higher education: student learning, teaching, programme and institutions*. London: Jessica Kingsley.

Manuscript received on 25 April 2002

Revision accepted for publication on 11 October 2002