Abstract

The use of timed, essay examinations is a well-established means of evaluating student learning in higher education. The reliability of essay scoring is highly problematic and it appears that essay examination grades are highly dependent on language and organisational components of writing. Computer-assisted scoring of essays makes use of language features and has demonstrated strong similarity to human ratings. Studies of examiner behaviour show that attention to content and language features contributes to grading decisions. However, given the time constraints on essay examinations, an overemphasis on language aspects may weaken the validity of essay examination grades. This article suggests alternative approaches to the standard essay prompt which should raise the validity of essay tasks and scoring in higher education. Suggested options include redesigning tasks so that organisational and language features are less influential in scoring and the use of content maps.

Understanding essay examinations

The paper begins with a discussion of essay examinations with reference to key literature:

- Essay examinations demand that students demonstrate knowledge of academic content taught during a course and synthesise an understanding of that content in a relatively brief period of time (Gentile, 2000; Bereiter, 2003).
- As a first draft, the exam essay is not fully representative of a student's written language abilities.
- Heavy weighting on criteria related to written communication may reduce the validity of exam essay marking.
- Essay requirements that encourage or allow students to memorise pre-set questions or “data-dump” everything they know on a topic undermine the intention of assessing student cognitive complexity or depth of understanding.
- Appropriate organisation is material is fundamental to demonstrate of depth and complexity.

Issues in scoring examination essays

This section includes a comprehensive review of the literature on essay marking - reliability and validity. It concludes that the organisation of material and language unduly influence the marking decisions of assessors and that the weighting given to these language factors by individual assessors varies considerably regardless of the grading criteria provided.
Automatic essay scoring and its validity for examinations
This is a critical discussion of computer-based essay scoring tools.

Improving the validity of essay examinations
Brown argues that, given the lack of validity of over-emphasising written language ability over content and cognitive complexity for essays produced in examination conditions, “it may be useful to consider whether there are means of reducing the effect of language or surface features on essay examinations scores when the goal is coherent, knowledge-transforming explication of deep reasoning” (p. 283). He proposes some practical strategies:

1. Require all students to use the same organisational pattern for their essay. An example is given to illustrate how an essay test item can be written with prompts that provide students with detailed guidelines so that all essays follow the same organisational pattern.

2. A second example provides students with sequencing or transition phrases that they are expected to use. For example “The evidence on this topic generally says….”. Four contrasting findings from the literature on this topic are….”. (285).

“These two alternative approaches to essay examination prompt design have the potential to raise both the reliability of scoring and the validity of essay scores as measures of content knowledge and understanding” (285).

3. The third suggestion requires “students to create concept maps and permitting them to be used in examinations to help their ability to compose written essays under pressure that more accurately reflect their real learning” (287).

Conclusion
Brown concludes that “while it is clear that these suggestions are not novel in so far as their application in writing courses or in coursework written assignments is concerned, their use in examinations has not been well established. Higher education needs further studies into validating examination essay grades, one of our primary means of evaluating student learning” (288).

This series of briefs on assessment topics has been prepared by the UQ Teaching and Educational Development Institute (TEDI) for UQ teaching academics. “UQ ASSESSMENT BRIEFS” of journal articles, book chapters, reviews, websites, reports etc are distributed to Faculty and School Teaching and Learning Chairs in a form designed to encourage wider distribution.

Please send feedback, requests or suggestions for future “UQ ASSESSMENT BRIEFS” to:

Dr Clair Hughes (clair.hughes@uq.edu.au)
Senior Lecturer in Higher Education
Teaching and Educational Development Institute (TEDI)
The University of Queensland, St Lucia Qld 4072
Telephone +61 7 336 52456
Facsimile +61 7 3365 1966