

The Role of Essay Tests Assessment in e-Learning: A Japanese Case Study

Minoru Nakayama¹, Hiroh Yamamoto¹ and Rowena Santiago²

¹Tokyo Institute of Technology, Japan

²California State University, San Bernardino, USA

nakayama@cradle.titech.ac.jp

Abstract: e-Learning has some restrictions on how learning performance is assessed. Online testing is usually in the form of multiple-choice questions, without any essay type of learning assessment. Major reasons for employing multiple-choice tasks in e-learning include ease of implementation and ease of managing learner's responses. To address this limitation in online assessment of learning, this study investigated an automatic assessment system as a natural language processing tool for conducting essay-type tests in online learning. The study also examined the relationship between learner characteristics and learner performance in essay-testing. Furthermore, the use of evaluation software for scoring Japanese essays was compared with experts' assessment and scoring of essay tests. Students were enrolled in two-unit courses which were taught by the same professor as follows: hybrid learning course at bachelor's level, fully online course at bachelor's level, and hybrid learning course at masters level. All students took part in the final test which included two essay-tests at the end of course, and received the appropriate credit units. Learner characteristics were measured using five constructs: motivation, personality, thinking styles, information literacy and self-assessment of online learning experience. The essay-tests were assessed by two outside experts. They found the two essay-tests to be sufficient for course completion. Another score, which was generated using assessment software, consisted of three factors: rhetoric, logical structure and content fitness. Results show that experts' assessment significantly correlates with the factor of logical structure on the essay for all courses. This suggests that expert evaluation of the essay is focused on logical structure rather than other factors. When comparing the score of experts' assessment between hybrid learning and fully online courses at the bachelor's level, no significant differences were found. This indicates that in fully online learning, as well as in hybrid learning, learning performance can be measured using essay tests without the need for a face-to-face session to conduct this type of assessment.

Keywords: online learning, essay-testing, learner characteristics, learning performance

1. Introduction

One of the learning goals of university instruction is to develop students' logical thinking and writing (Biggs, 1999). This is true even with online courses which are gaining popularity in higher education and are taught as hybrid or fully online courses. E-learning, however, has its restrictions on how learning performance is assessed. Online testing is usually conducted through multiple-choice questions, without using any essay type of learning assessment. Major reasons for employing multiple-choice tasks in e-learning include ease of implementation and ease of managing learner responses.

On the other hand, conventional face-to-face classes often employ essay-type examinations for the purpose of assessing the learners' meta-cognitive understanding and ability to build logical structures beyond the understanding of basic knowledge (Biggs, 1999; Brown and Knight, 1994).

To address this limitation in online assessment of learning, this study investigated an automatic assessment system as a natural language processing tool for conducting essay-type tests in online learning. The study also examined the relationship between learner characteristics and learner performance in essay-testing. In addition, the use of evaluation software for scoring Japanese essays was compared with experts' assessment and scoring of essay tests.

2. Method

2.1 Experimental procedure

Three credit courses (Nakayama et al., 2008), which were offered in the Spring and Autumn terms of 2006-2007 were selected for this survey project. The course title of the first two courses was "Information Society and Careers", a 2-unit bachelor-level class for university freshmen, with one course offered as a fully online course and the other as hybrid course. Students could choose to attend either course, in accordance to their preference.

The third course was "Advanced Information Industries", a 2-unit master's class for students in their first year of graduate work. Most master's students have had some experience with hybrid courses during their bachelor years. Most of the students who took this course were to major in Engineering.

The three courses were taught by the same professor at a Japanese national university. The hybrid courses consisted of regular 15-week face-to-face sessions, supplemented with e-learning components in the form of online modules and tests. Students attended the face-to-face class and were also able to access the online content outside of class.

The e-learning components were originally designed for a fully online course. The modules include video clips of the instructor and the lecture for that session, plus the presentation slides which were used in the face-to-face lecture. Most tests were conducted in the multiple-choice format. Learners can assess their responses and view their individual scores after completing the test. They are given as many opportunities as needed to retry and answer each question until they are satisfied with their own scores. This in turn motivated them to learn the course content well, using the accompanying video clips and presentation slides. To encourage maximum participation in e-learning, students in the hybrid courses were given the opportunity to earn extra points.

Student enrolment in these courses is as follows: hybrid learning course at bachelor's level had 47 participants, fully online course at bachelor's level had 39 participants, and hybrid learning course at master's level had 78 participants. All students took part in the final test which included two essay-tests at the end of the course, and received the appropriate credit units. Learner characteristics were measured using five constructs: motivation, personality, thinking styles, information literacy and self-assessment of online learning experience.

2.2 Survey instruments

To extract learner characteristics among Japanese students, five constructs were surveyed, using the same constructs and questionnaires in previous studies conducted in 2006 and 2007 (Nakayama et al., 2006, 2007a, 2007b). These constructs were: motivation (Kaufman and Agars, 2005), personality (Goldberg, 1999; IPIP, 2004), thinking styles (Sternberg, 1997), information literacy (Fujii, 2007) and self-assessment of online learning experience. In this paper, the relationship between essay tests and two of these constructs (information literacy and learning experience) were investigated. Further descriptions of these two metrics are given in the following sections.

Information literacy

Fujii (2007) defined and developed inventories for measuring information literacy. For this construct, the survey consisted of 32 question items, and 8 factors were extracted: interest and motivation, fundamental operation ability, information collecting ability, mathematical thinking ability, information control ability, applied operation ability, attitude, and knowledge and understanding.

Secondary factor analysis was conducted on the above ten-factor scores for information literacy, and as a result, two secondary factors were extracted (Nakayama et al. 2008). The first secondary factor (IL-SF1) consists of "operational confidence and knowledge understanding"; the second one (IL-SF2) consists of "attitude issues".

Learning experience

Students' online learning experiences were assessed using a 10-item Likert-type questionnaire. This questionnaire was administered twice: during the second week of the term and at the end of the course. As in previous studies, three factors were extracted from this instrument: Factor 1 (F1): overall evaluation of e-learning experience, Factor 2 (F2): learning habits, and, Factor 3 (F3): learning strategies (Nakayama et al, 2006, 2007a, 2007b).

Learning performance

The students' final grade for the course was based on various learning activities. Here, three indices were identified and used as indicators of learning performance: the number of days attended (NDA), the number of completed modules (NCM), and the online test scores (OTS) (Nakayama et al, 2006, 2007a, 2007b). They were analyzed for their relationship with essay-test scores.

2.3 Final test for the courses

For bachelor students, the final test was conducted by a proctor during the scheduled finals week at the university. All students gathered in a lecture room, and answered four questions – two questions included some multiple-choice tasks and the other two questions were essay-tests.

For masters students, the final test was a written report based on their research work on two self-selected questions out of a given set of five themes.

3. Results

3.1 Essay-test assessment

Although the style of essay tests is a little bit different between bachelors and masters, this type of assessment was conducted for both levels.

The essay-tests were reviewed by two outside experts and were found to be sufficient for course completion. Before the assessment, the two experts independently evaluated the essays using a 3-points scale (0-2) which was applied to each of the five aspects of the essay test: certainty, fitness for learning content, argument, various aspects and figuring. For this study, all usable data were used for this analysis.

Table 1: Mean of expert's assessment and their correlation coefficients

Mean(SD)	Expert 1	Expert 2	r
Essay test 1 (N=398)	7.7(1.4)	6.6(1.6)	0.56
Essay test 2 (N=398)	7.5(1.5)	6.1(1.8)	0.63
Total	15.3(2.4)	13.1(2.7)	0.67

Assessment scores from the two experts who evaluated the two essay questions are summarized in Table 1. Here, scores for the two essays (essay 1 and essay 2) at the master's level were used based on the two essay reports that the students wrote. The ratings that the experts gave for these two sets of essay tasks were very close and almost similar. Correlation coefficients are also summarized in Table 1. Overall, assessment scores from each of the two essays strongly correlated with each other ($r=0.67$), therefore they could be merged to form a single score.

For the automated Japanese essay assessment, an automated scoring system (Ishioka and Kameda, 2003) was used. It is possible to use this system via web site. As a result, another set of scores was generated using the assessment software and these scores measured three factors: "rhetoric", "logical structure" and "content fitness".

The relationship between experts' assessment score and the automated assessment score were examined. Correlation coefficients (r) are summarized in Table 2.

Table 2 shows that experts' assessment significantly correlates with the factor of "logical structure" on the essay for all courses ($r=0.30$). There are no significant relationship between experts' assessment and "rhetoric" or "content fitness" of the automated essay-scores.

This suggests that expert evaluation of the essay is focused mainly on "logical structure" rather than other factors.

Table 2: Correlation between expert's assessment scores and automated assessment scores

N=209	Expert 1	Expert 2	Total
rhetoric	(-.02)	(-.12)	(-.07)
logical structure	0.16	0.39	0.30
content fitness	(-.05)	(-.07)	-0.01
Total	0.16	0.46	0.35

3.2 Performance of essay-test between hybrid course and fully online course

The university credit courses used in this study have served as examples of courses where essay-tests were conducted in both hybrid and fully online settings. Therefore, it would be very interesting to find out whether the essay-test performances of students in these two online learning settings are equivalent or not.

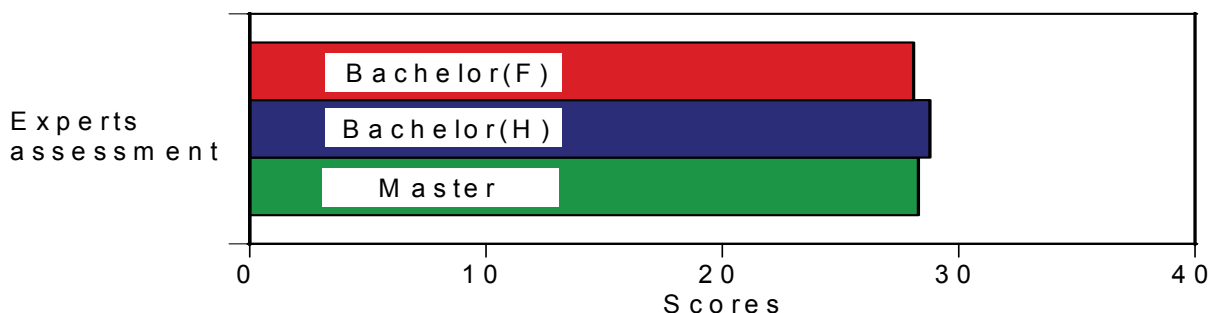


Figure 1: Expert's assessment scores

Experts' assessments for the three groups, namely bachelor-hybrid, bachelor-fully online, and masters-level, are compared in Figure 1. As the figure illustrates, assessment scores are almost of the same level across all three groups. Experts evaluated these essay-tests using common criteria, so that it is possible to compare the scores among them. The task style was different between bachelors and masters, however, these results did not mean that bachelor students and masters students had the same level of performance when it comes to essay-writing.

When comparing the experts' assessment scores between hybrid learning and fully online courses at the bachelors level, no significant differences were found ($t(73)=0.47, p=.64$). This indicates that in fully online learning, as well as in hybrid learning, learning performance can be measured using essay tests without the need for a face-to-face session to conduct this type of assessment.

3.3 Automated essay assessment

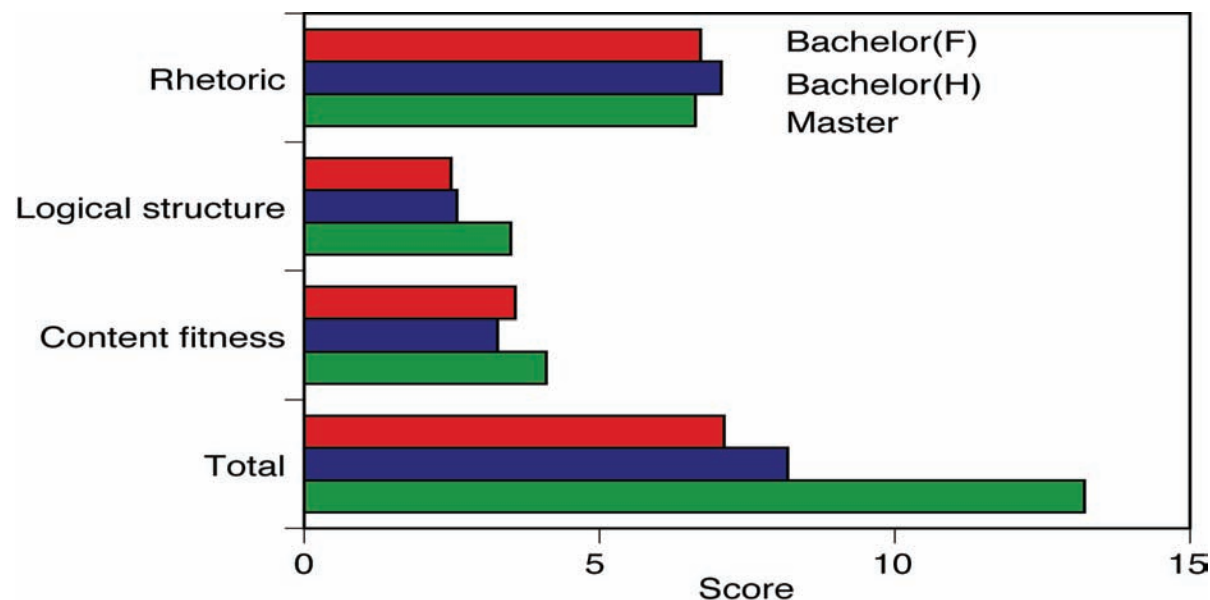


Figure 2: Assessment scores for the essay-test using the automated assessment system

Assessment scores for the essay-test using the automated assessment system are illustrated in Figure 2. In this figure, performance scores in the three groups are compared. It was not possible to analyze the difference between all bachelors and masters scores because testing styles were totally different, but scores for the master group are higher than scores for bachelor groups on two factors -- "logical structure" and "content fitness". This system automatically adjusts some scores in relation to other factors in the given situation using a minus-points system (Ishioka and Kameda, 2003), with the result that the total score does not simply reflect the sum of three factor scores. Among all the total scores, the

score for masters is the highest. This suggests that the point deducted for masters is smaller than ones for bachelors, so that the result of the total scores seems reasonable.

3.4 Multiple choice task and essay-test

As explained in the introduction, most online test consists of multiple-choice tasks. For the bachelor students' final paper-and-pencil test, both multiple-choice tasks and essay tests were included. According to the comparison of scores for multiple-choice tasks between the two groups, the scores in the hybrid courses are significantly higher than scores in fully online courses ($t_{(73)}=5.1, p<0.01$). This could mean that bachelor students in hybrid courses have a better understanding and knowledge of the learning content in weekly face-to-face sessions. On the other hand, students in fully online course who may possess some previous knowledge of information technology which was the main topic of the course (Nakayama et al., 2008) seem to have the ability to summarize this previous knowledge for an essay test. Thus, it may be possible that there is no significant difference in the scores of between essay-tests in hybrid courses and essay-tests in fully online courses.

To determine the effect of performance in multiple-choice tasks on scores for essay-tests, correlation coefficients in scores between multiple-choice tasks and essay-test were extracted and summarized in Table 3.

Table 3: Correlation coefficients between multiple-choice test scores and essay test scores

	B(Hybrid)	B(Fully)
	N=45	N=30
Reviewers	(-.06)	(-.09)
rhetoric	0.37	(0.28)
logical structure	(-.17)	(-.12)
content fitness	(-.15)	0.54
Total	(-.01)	-0.36

The correlation coefficients between scores for the multiple-choice tests and experts or automated assessments for hybrid and fully online courses are summarized in Table 3. Almost all coefficients are not significant, and this result confirms that the essay-test measures the different aspects of learner performance in multiple-choice tasks.

3.5 Relationship with learners' metrics

The conventional metrics which relate to learner characteristics may affect the essay-test scores as measures of learning performance, thus, a correlation analysis was conducted. The correlation coefficients are summarized in Table 4.

Table 4: Correlation coefficients between expert's assessment and survey metrics

	Masters	Bachelors	
		Hybrid	Fully online
	N=76	N=45	N=30
Information literacy			
Operational confidence and knowledge	(0.08)	-.47	(0.03)
Attitude	(-.13)	0.00	(0.01)
Learning experience			
E-learning evaluation	(-.04)	-.41	(-.21)
Learning habit	(-.01)	(-.16)	(0.26)
Learning strategy	(-.11)	(-.01)	(0.12)
Behavioral data			
Number of days attended	0.40	(0.02)	-
Number of completed modules	(0.22)	(0.05)	(0.00)
Online test scores	(0.22)	(-.03)	(0.25)

As the results indicate, additional interesting correlations were observed: essay-test scores negatively correlated with "e-learning evaluation" scores ($r=-.41$); and, the information literacy factor scores of "operational confidence and knowledge" ($r=-.47$) in hybrid learning and essay-test scores correlated with "the number of days attended" ($r=0.40$) in hybrid learning for masters students.

As discussed in earlier sections above, the possible reason may depend on the learning content and learning environment. Therefore, these findings could be related with the content of the courses. Other learner characteristics did not affect students' essay-test scores. Further detailed analysis of these factors will be the topic of our next study.

4. Conclusion

To determine the students' learning performance in online learning activities, essay-tests were introduced and examined as assessment tools for hybrid and fully online courses in the bachelor's level and for hybrid courses in the master's level. Also, to study the results of using an automated assessment system for essay-tests, a proto type system was introduced and results were discussed. As the results indicated, a common factor, "logical structure", was assessed in the essay-test by both experts and automated system.

There is little difference in scores resulting from the use of essay test as an assessment tool in hybrid and fully online courses at the bachelor's level. However, there is a significant difference in scores for multiple-choice tasks. Though most of these results may depend on learning content and environment, it is possible to measure learning performance through the use of essay-tests in either hybrid or fully online courses.

Acknowledgements

This research was partially supported by the Japan Society for the Promotion of Science (JSPS), Grant-in-Aid for Scientific Research (B-19300274: 2007-2009).

This manuscript appears in "Online Learning Management and Learner's Behavior - A Case Study Of Online Learning in Japan" as a book chapter in "Developing And Utilizing E-Learning Applications", edited by Fotis Lazarinis, Steve Green, and Elaine Pearson, which is published by IGI Global. The authors would like to thank that IGI Global granted to us the permission of this publication.

References

- Biggs, J., "Teaching for Quality Learning at University", Buckingham, UK:SRHE and Open University Press (1999)
- Brown, S., & Knight, P. "Assessing Learners in Higher Education", Kogan Page (1994)
- Fujii, Y., "Development of a Scale to Evaluate the Information Literacy Level of Young People -Comparison of Junior High School Students in Japan and Northern Europe", Japan Journal of Educational Technology, Vol. 30(4) (2007), pp.387-395
- Goldberg, L. R., "A Broad-Bandwidth, Public Domain, Personality Inventory Measuring the Lower-Level Facets of Several Five-Factor Models". In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), Personality Psychology in Europe, Vol. 7, pp. 7-28, Tilburg, The Netherlands: (Tilburg University Press, 1999).
- International Personality Item Pool, "A Scientific Collaboratory for the Development of Advanced Measures of Personality Traits and Other Individual Differences" (2001) Retrieved 27 October, 2004 from <http://ipip.ori.org/>.
- Ishioka, T., Kameda, M., JESS: An Automated Japanese Essay Scoring System, Bulletin of the Computational Statistics of Japan, Vol.16, No.1, pp.3-19 (2003)
- Kaufman, J. C. & Agars, M. D., "Kaufman-Agars Motivation Scale", Unpublished instrument (2005).
- Nakayama, M., Yamamoto, H. & Santiago, R., "Investigating the Impact of Learner Characteristics on Blended Learning among Japanese Students", Proceedings of ICEL 2006, pp.361-370 (Academic Conferences, 2006).
- Nakayama, M., Yamamoto, H. & Santiago, R., "Relationship Between Learner Characteristics and Learning Performance in Hybrid Courses Among Japanese Students", Proceedings of ICEL 2007, pp.341-349 (Academic Conferences, 2007a).
- Nakayama, M., Yamamoto, H. & Santiago, R., "The Impact of Learner Characteristics on Learning Performance in Hybrid Courses among Japanese Students", The Electronic Journal of e-Learning, Vol.5, Issue 3, pp.195-206 (2007b).
- Nakayama, M., Yamamoto, H. & Santiago, R., "Impact of Information Literacy and Learner Characteristics on Learning Behavior of Japanese Students in Online Courses", International Journal of Case Method Research & Application, Vol.XX, Issue No.4, pp.403-415 (2008).
- Sternberg, R. J., Thinking Styles, Cambridge University Press, Cambridge, UK (1997).