

LECTURERS' EXPECTATIONS OF A COMPUTED-BASED ESSAY MARKING SYSTEM

Saadiyah Darus and Siti Hamin Stapa
Universiti Kebangsaan Malaysia

ABSTRACT

Marking essays is a difficult and tedious task for most lecturers especially when they have to mark long essays in vast numbers. The situation is more problematic at higher learning institutions where essays constitute a major component of a course. In order to alleviate some of these problems, computer-based essay marking systems have been developed. Although some of these systems are already available in the market, Malaysian lecturers do not utilize them. The aim of this paper is to investigate the reasons to explain this phenomenon. This article introduces computer-based essay marking systems and some of the available software. The second part of the article describes the methodology used in this research. The results highlight the extent to which these lecturers are exposed to computer-based essay marking systems as well as their expectations of such systems. Hence, the results of this research will give possible directions for future research in this area in the Malaysian context.

Introduction

Writing essays plays an important role in education especially at higher learning institutions because they promote higher order learning skills (Hounsell, 1984; Biggs, 1988). At the tertiary level of education, essays are generally given to students as written assignments that need to be submitted by a certain deadline. In marking essays, lecturers need to be consistent and ideally, students should get prompt feedback to essays submitted. As a consequence, marking essays manually is an overwhelming task for lecturers, especially when the number of students is large or if they teach more than one course and several essays are given as written assignments for a particular course.

In order to alleviate some of these problems, computer-based essay marking (CBEM) systems have been developed. CBEM systems utilize computer technology to mark essays. They have been developed by various researchers from different fields and can be categorized into two groups: automated and semi-automated computer-based marking systems (Darus, 1999). Automated CBEM systems are Project Essay Grade (PEG), Intelligent Essay Assessor (IEA) and electronic essay rater (e-rater).

Page et al (1968) developed PEG. PEG-1 was able to predict the scores that human graders give to the students' essays, while PEG-2 showed that the computer was more accurate in scoring creativity and style than in scoring mechanics when compared to human judges. PEG-3 used specific content words and their synonyms as additional features in the system. Landauer et al. (1998) developed the Intelligent Essay Assessor (IEA) which uses Latent Semantic Analysis for extracting and inferring relations of expected contextual usage of words in a passage of discourse and is designed to measure the knowledge content of essays.

E-rater was developed by Burstein et al. (1998a). It was based on the writing characteristics specified in the holistic scoring of Graduate Management Admission Test (GMAT) essays. It is able to analyze essay features such as discourse, syntax and topic content (Burstein et al., 1998b). It is claimed that the system can be used to evaluate non-native speakers' writing (Burstein & Chodorow, 1999).

On the other hand, semi-automated computer-based essay marking systems are Methodical Assessment of Reports by Computer (MARC) and Markin 32. MARC is a report-marking programme for the engineering course at Papua New Guinea University of Technology. Marshall and Barron (1987) developed the programme to enable instructors to provide three pages of individualized feedback on every student's report, based on certain criteria for report writing.

Markin 32 developed by Holmes (1996), is a correction tool that allows instructors to mark written work submitted by students in the form of electronic documents. Markin 32 version 1.2 provides five marking facilities; annotation button, add feedback, add comment, add a grade, and compile error statistics. It allows instructors to return marked written work to students in three different formats: web page, word-processor file and text file.

A detailed review of computer-based essay marking systems is documented in Darus (2000).

The automated marking systems, namely e-rater, IEA and PEG are available on the Web. However, to our knowledge Malaysian lecturers at higher learning institutions do not utilize them.

Aim of the Study

The aim of this study is to investigate the reasons why Malaysian lecturers at institutions of higher learning do not utilize currently available CBEM systems.

In particular, the study will address the following research questions:

1. Are the lecturers aware of the availability of computer-based essay marking systems? To answer research question 1, this study examined the extent to which lecturers have heard of and used a CBEM system.
2. What are the lecturers' opinions about a CBEM system? To answer research question 2, this study examined a) whether the lecturers believe it is possible for computers to mark essays effectively and b) whether it will be beneficial to them if computers can be used to mark essays.
3. What are the expected features that lecturers look for in a CBEM system? To answer research question 3, the study examined the desirable functions that a CBEM system should provide.

Methodology

A survey in the form of a questionnaire was developed and distributed to lecturers in the form of an e-mail. Some of these lecturers are currently teaching at UKM and were selected from the university's handbook and homepage. These lecturers give written assignments to students and can be contacted through e-mail. A total number of 80 respondents participated in this study and the data collated were then analyzed.

The questionnaire consists of 21 questions that are divided into three sections. The first section is about the respondents' backgrounds. The second section illuminates their approach in teaching the courses and the final section focuses on their opinions about a CBEM system.

Results

i) Background of the respondents

Out of 80 lecturers who participated in this study, 52 were female (65 %) and 28 were male (35 %). In terms of area of specialization, 22 lecturers specialized in economics and business (27.5 %), 26 lecturers in arts and social science (32.5 %) while 40 lecturers (50 %) specialized in language and education.

When analyzing the lecturers' approach in teaching the courses, it is concluded that they usually pay attention to mark the following essay traits: creativity, style, organization of ideas, knowledge or topic content and syntax. The number of lecturers who mark these traits is shown in Figure 1. From this Figure, it is clear that 74 lecturers (92.5 %) mark organization of students' ideas in essays. 70 lecturers (87.5%) mark knowledge or topic

content of an essay. 52 lecturers (65.0%) mark creativity and 42 lecturers (52.5 %) mark style. Only 38 lectures (47.5 %) mark syntax.

Fig. 1: Essay traits that lecturers mark

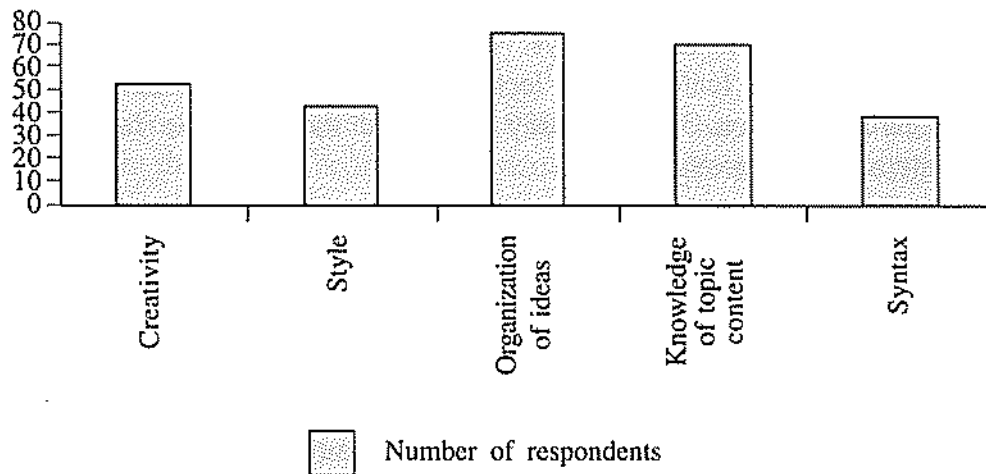
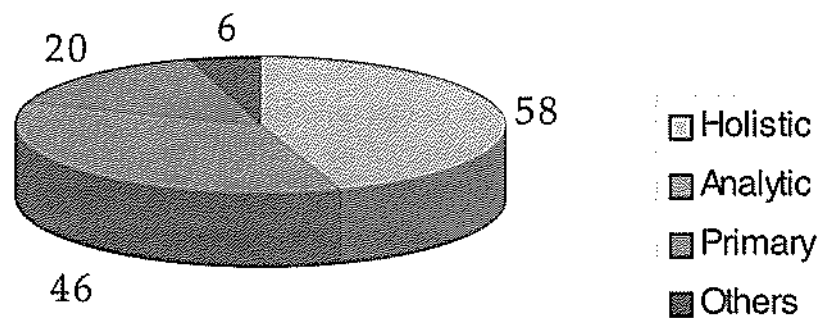


Figure 2 shows the type of scoring procedures that lecturers used in marking essays. The most common scoring procedure used by the lecturers is holistic scoring, used by 58 lecturers (72.5%). The next popular scoring procedure is analytic scoring used by 46 lecturers (57.5%). 20 lecturers (25.0%) used primary trait scoring while 6 lecturers (7.5%) do not use any of these three scoring procedures.

Fig. 2: Types of scoring procedure used



Lecturers' awareness of the availability of computer-based essay marking systems

Fig. 3: Lecturers' awareness of CBEM systems

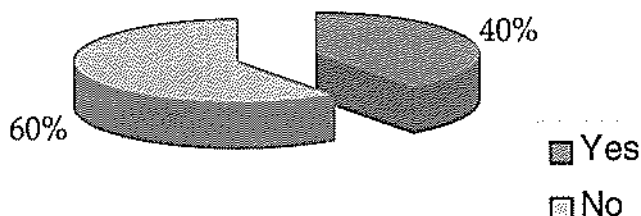
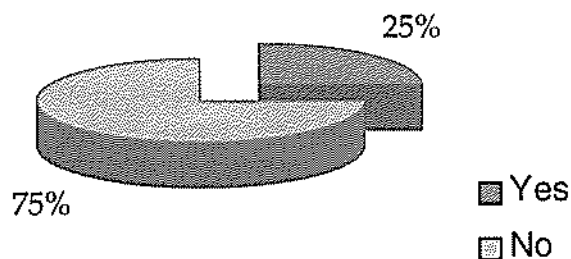


Figure 3 shows that most of the lecturers (60%) have not heard about the CBEM systems. However, it is encouraging to note that at least 40% of them have heard about the CBEM systems. The results clearly show that most of the lecturers are not aware of the availability of CBEM systems.

Lecturers' opinions about CBEM systems

Lecturers have mixed opinions as to whether computers are able to mark essays effectively. It is interesting to note from Figure 4, that although most of the lecturers (75 %) do not believe that computers can mark essays effectively, 67 % of them believe that computers are beneficial to them if computers can mark essays as shown in Figure 5. This implies that to a certain extent, the lecturers believe that computers can mark essays while some other aspects should be left to human raters.

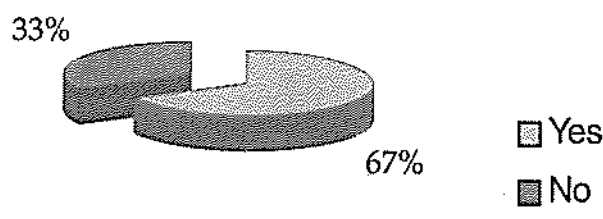
Fig. 4: Lecturers' opinions on whether computer can mark essays effectively



Expected features that lecturers look for in a CBEM system

The results from Table 1 show that most of the lecturers (57.7 %) have the opinion that a CBEM system should be able to **indicate errors** in essays. The next three important functions that a CBEM system should be able to perform are as follows: **mark syntax and provide error statistic** (47.5%); **mark non-native speaker's writing and produce a letter grade** (42.5%); and **mark organization of ideas** (40.0%). A plausible reason for these expectations is that the lecturers believe that a computer can easily and effectively perform these functions while the rest of the functions should be performed by lecturers since these can best be carried out by humans.

Fig. 5: Lecturers' opinions on whether computers are beneficial if they can be used to mark essays



Other desirable functions (below 40 %) that CBEM systems should provide in decreasing importance are mark surface features (37.5 %), rhetorical structure (37.5 %), topic content (35.0 %), give individual feedback (35.0 %) mark holistically (35.0 %), mark knowledge content (32.5 %) and mark analytically (32.5 %). It is not surprising that to mark according to discipline (30.0 %) is the least expected function for a computer to carry out because presumably humans can perform this more efficiently than a computer can.

Table 1: The desirable functions that CBEM systems should provide

Functions that should be provided by CBEM systems	No. of respondents (percentage)
Indicate errors	46 (57.5%)
Mark syntax	38 (47.5%)
Provide error statistics	38 (47.5%)
Mark non-native speaker's writing	34 (42.5%)
Produce letter grades	34 (42.5%)
Mark organization of ideas	32 (40.0%)
Mark surface features	30 (37.5%)
Mark rhetorical structure	30 (37.5%)
Mark topic content e.g. look at vocabulary	28 (35.0%)
Give individual feedback	28 (35.0%)
Mark holistically	28 (35.0%)
Mark knowledge content e.g. look at semantics	26 (32.5%)
Mark analytically	26 (32.5%)
Mark according to disciplines	24 (30.0%)

Suitability of CBEM systems

Table 2 lists the functions that are supported by some of the CBEM systems. In the table, "X" indicates that a particular function can be performed by the systems. The right-most column lists the desirable functions that lecturers expect from a CBEM system, together with an indicator to mark its importance. For example, **indicate errors** is the most desirable feature and thus is denoted as **1**. The second most desirable feature, that is **mark syntax** is ranked as **2**.

From Table 2, it seems that none of the available CBEM systems satisfy all fourteen desirable functions. The system that comes closest to the lecturers' expectations is e-rater which can perform nine out of the fourteen functions. PEG comes next, as it is able to perform six functions. IEA and Markin 32, both can perform four functions while MARC can only perform two functions. This implies that the currently available CBEM systems are not suitable for the Malaysian environment since they do not address the needs and expectations of Malaysian lecturers.

Table 2: Functions supported by currently available CBEM systems and the desirable functions expected by lecturers

Functions	MARC	Markin 32	PEG	IEA	E-Rater	Lecturers' Expectations
Letter grade/score	X	X	X	X	X	4
Indicate errors		X				1
Provide error statistics		X				3
Mark surface features e.g. essay length			X		X	6
Mark topic content e.g. by looking at vocabulary			X		X	7
Mark knowledge content e.g. by looking at semantics (measure concepts)				X		8
Mark rhetorical structure e.g. Issues and argument					X	6
Mark syntax			X		X	2
Mark non-native speakers writing (i.e. respondents writing in English)					X	4
Mark reports (engineering)	X					*
Mark style			X	X		*
Mark mechanics e.g. misspelled words			X	X		*
Detect plagiarism				X		*
Mark expository essay (biology, psychology, history)				X		*
Mark subject-matter essays e.g. science, social studies, English, foreign language			X			*
Holistic marking			X		X	7
Coherence of text				X	X	*
Organization of ideas					X	5
Creativity			X			*
Give individual feedback	X	X	X	X	X	7
Mark analytically						8
Mark according to disciplines				X		9

* Indicates that the respondents were not questioned on this particular item.

Conclusion

This study was carried out to investigate the reasons why Malaysian lecturers at institutions of higher learning do not utilize currently available CBEM systems. The study focused on three research issues. The first issue is to find out whether the lecturers are aware of the availability of computer-based essay marking systems. The results of the study show that most of the lecturers are not aware that CBEM systems are available.

The second issue is to investigate the lecturers' opinions about a CBEM system. The results of the study seem to suggest that while a minority of the lecturers think that it is possible for computers to mark essays effectively, the majority of them believe that computers will be beneficial to them if they can be used to mark essays.

We have compared the functions that are provided by available CBEM systems against lecturers' expectations. While some of these functions can be provided by these systems, none of the systems can provide all fourteen functions desired by lecturers.

There is one significant contribution of this study. The study indicates that there is a need to develop a new CBEM system for the Malaysian environment. This conclusion is in agreement with the results obtained from our previous study (Darus, Hussin and Stapa, 2001). The developer of this new system can make use of the results of the study to develop a system that satisfies Malaysian lecturers' expectations for marking essays.

References

- Biggs, J.B. 1998. "Approaches to Learning and Essay Writing", R.R. Schmeck (ed.) *Learning Strategies and Learning Styles*. New York: Plenum Press.
- Burstein, J., & Chodorow, M. 1999. Automated Scoring for Non-native English Speakers. Presented at the Joint Symposium of the Association of Computational Linguistics and the International Association of Language Learning Technologies, Workshop on Computer-Mediated Language Assessment and Evaluation of Natural Language Processing, College Park, Maryland.
- Burstein, J., Kukich, K., Wolff, S., Lu, S. & Chodorow, M. 1998a. Computer Analysis of Essays. Presented at the NCME Symposium on Automated Scoring, April 1998.
- Burstein, J., Kukich, K., Wolff, S., Lu, S. & Chodorow, M. 1998b. Enriching Automated Scoring Using Discourse Marking. In the Proceedings of the Workshop on Discourse

- Relations and Discourse Marking. Annual Meeting of the Association of Computational Linguistics, August 1998. Montreal, Canada.
- Darus, S. 1999. A Prospect of Automatic Essay Marking. Paper presented at SEAMEO Regional English Language Center Seminar, 19-21 April 1999, Singapore.
- Darus, S., Stapa, S.H, Hussin, S. & Koo, Y.L. 2000. A Survey of Computer-Based Essay Marking (CBEM) Systems. In the Proceedings of the International Conference on Education and ICT n the New Millennium. Universiti Putra Malaysia, Kuala Lumpur. pp. 519-529.
- Darus, S., Hussin, S. & Stapa, S. H. 2001. Students' Expectations of a Computer-Based Essay Marking System. Paper presented at the International Education Conference (IEC 2001), 27-28 August 2001, Hilton Petaling Jaya, Malaysia.
- GMAT. 1997. <http://www.gmat.org/>
- Holmes, M. 1996-1998. Markin 32 Version 1.2 available at: <http://www.arts.monash.edu.au/others/calico/review/markin.htm>
- Hounsell, D.J. 1984. "Learning and Essay Writing". In F. Marton, D. Hounsell & N. Entwistle (eds.) *The Experience of Learning*. Edinburgh: Scottish Academic Press.
- Landauer, T.K., Foltz, P.W. & Laham, D. 1998. Introduction to Latent Semantic Analysis. *Discourse Processes*, 25: 259-284.
- Marshall, S. & Barron, C. 1987. MARC-Methodical Assessment of Reports by Computer. *Systems*, 15 (2):161-67.
- Page, E.B., Fisher, G.A. & Fisher, M.A. 1968. Project Essay Grade: A FORTRAN Programme for Statistical Analysis of Prose. *British Journal of Mathematical and Statistical Psychology*, 21: 139. (Abstract)