A Pilot Study on the Impact of the Web-based Essay Critiquing System on Writing at the Tertiary Level


aDepartment of Computer Science, Hong Kong Baptist University, Hong Kong. 
 bLanguage Centre, Hong Kong Baptist University, Hong Kong. 
 cInterMedia, University of Oslo, Norway. 
 *Corresponding Author: kckwong@comp.hkbu.edu.hk

ABSTRACT

As English is widely used worldwide, it is the preferred second language in Hong Kong. Many students find essay writing stressful because they do not have sufficient ideas to fully cover the topic of the essay. To alleviate the aforementioned learning barriers, a web-based essay critiquing system was developed using Latent Semantic Analysis (LSA), an automatic text analysis technique, for providing just-in-time feedback to students. The feedback takes two forms: new sub-themes suggested to be included, and the visualization of the existing sub-themes’ organization. In this paper, we present our findings on students’ performance and their perception of the usefulness of this system.

Keyword: Essay Writing, Latent Semantic Analysis, Critiquing System

1. Introduction

Acquiring good essay writing skills is important, and yet challenging for students in Hong Kong. Many of them find the essay writing task stressful because they are short of ideas to fully discuss the essay topic and complete the essay. If it is an in-class writing exercise, it will be difficult for the teacher to give immediate content-related hints to each individual student. If it is a take-home writing exercise, getting immediate feedback from the teacher is simply impossible. Students usually make multiple drafts before finalizing their essays. Even if the teacher can afford to provide feedback to each draft, the turn-around time will be in days at least. This often makes the learning process quite ineffective.

To alleviate the aforementioned learning barriers, a computer-supported critiquing system was developed using Latent Semantic Analysis (LSA), an automatic text analysis technique, for providing just-in-time feedback to students. The feedback takes two forms:

• new sub-themes suggested to be included, and
• the visualization of the existing sub-themes’ organization.
With the help of the feedback, students can further revise their essays accordingly. As
the feedback is immediate, students do not need to wait for teachers’ hints/comments
to further revise their essays. Apart from it, the system is a web-based one, making the
learning to take place anywhere, as long as there is Internet connection. It is believed
that the immediate feedback and easy access will provide assistance and encourage
students to better improve their writings before submitting the final version for
grading.

The rest of this paper is organized as follows. It is first started by briefly explaining
how LSA is incorporated as part of the corresponding system architecture. Next, the
evaluation methodology and the findings are described. Finally, some concluding
remarks and suggestions for further development are made.

2. A Critiquing System for Essay Writing

The students write their essays using the Essay Critiquing System. When they require
assistance, they submit their essays to the system which provides feedback (critique)
by suggesting the new sub-themes which can be considered to be included in the
essays. The texts in the student essays that match against one of the sub-themes of the
sample essays are highlighted in colour along with a sentence describing the sub-
theme. This can help students take a fresh look at the organization of their essays.

Latent Semantic Analysis (LSA) is a mathematical technique for computing the
semantic similarity between pieces of textual information (for example, sentences,
paragraphs or essays) with the help of a large corpus (Landauer and Laham, 1998;
Landauer and Psotka, 2000). It is incorporated as part of our essay critiquing system
to compare student and model essays, and provide relevant critic feedback to the
students when using the system.

The components of the proposed architectur e are as follows: teacher input, student
input, database that stores student answers and reference materials from external
sources, text segmentation and preprocessing engine, LSA engine, semantic matcher
and critic feedback to students (Figure 1).

The teacher decides on an essay topic for students, and collects some relevant
background materials of the topic such as articles from the Internet resources or
textbooks to build a corpus. Then, a list of possible sub-themes, which may be
included in the essay, is extracted from a set of sample essays, which can be some
good student essays of past years or model essays from some books. Before entering
the data into the LSA engine, all articles are first broken down into sentences and
preprocessed, e.g. stop-word removal and stemming. Then the LSA engine computes
the word-segment association matrix and the semantic similarity between all possible
pairs, one from the student essay and the other from the sub-theme list extracted from
good samples. When it is completed, the sub-themes on the list that are found missing
in the students’ essays can be identified. The final step is to determine the missing
sub-themes to be reported to students for their consideration when revising their
essays.
3. Evaluation Methodology

In order to evaluate the feasibility of this system, a pilot study was conducted. The students in this study were Year 1 and Year 2 undergraduates. They were randomly assigned to the treatment group and the control group. Both groups consisted of 14 students and were asked to write a 300-word essay on the topic “Should there be rules against fast food in schools?” within 80 minutes. The treatment group was allowed to use the essay critiquing system for feedback in the forms of new sub-themes and the visualization of the existing sub-themes’ organization. There was no limit on the number of submissions for system feedback. The control group, however, only used the computers to type their essays. Both groups were allowed to use paper dictionaries and/or online dictionaries. At the end of the study, a questionnaire was administered and an individual interview was conducted to all students in the treatment group. The questionnaire consisted of six questions on student perception of the usefulness of the system and some questions on demographics. The final version of the student essays were marked by two English teachers of a university, based on the criteria of grammar, content and organization and the full mark is 20.

4. Results and Discussion

4.1 Student Performance

The length of essay can be one of the indicators to tell if students have sufficient ideas to fulfill the number of word requirement set by the teacher. Table 1 shows that the

Figure 1. Essay Critiquing System Architecture
essay average length, as measured by the number of words, of the treatment group is 26.3 words more than that of the control group. However, this difference is not statistical significant at 0.05 level.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>14</td>
<td>387.86</td>
<td>85.06</td>
</tr>
<tr>
<td>Control</td>
<td>14</td>
<td>361.57</td>
<td>67.30</td>
</tr>
<tr>
<td>Overall</td>
<td>28</td>
<td>374.71</td>
<td>76.44</td>
</tr>
</tbody>
</table>

Table 1. Descriptive statistics on essay length

One of the goals of developing this essay critiquing system was to help students improve the performance of essay writing. Table 2 shows that the essay average score of the treatment group is slightly higher than that of the control group, but the difference is not statistical significant at 0.05 level.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>14</td>
<td>13.38</td>
<td>1.74</td>
</tr>
<tr>
<td>Control</td>
<td>14</td>
<td>12.66</td>
<td>1.23</td>
</tr>
<tr>
<td>Overall</td>
<td>28</td>
<td>13.02</td>
<td>1.53</td>
</tr>
</tbody>
</table>

Table 2. Descriptive statistics on essay score

Although both the essay average score and the essay average length of the treatment group are slightly higher than those of the control group, the differences are not statistical significant. This may be due to the fact that the treatment period is insufficient to significantly influence the students’ writing quality. In this study, the students only used the system to write one essay within 80 minutes. It is expected that a longer treatment period may yield a significant difference.

4.2 Student Perceptions

When this system was first made available for students to use, we were eager to seek feedback from them. At the end of the study, an anonymous questionnaire was given to each student in the treatment group in order to ask about their perceptions of the system. All students returned the questionnaires. Tables 3 to 6 show the student responses to the questionnaire. The responses in percentage are shown in brackets.

<table>
<thead>
<tr>
<th>This system is:</th>
<th>very easy to use</th>
<th>easy to use</th>
<th>difficult to use</th>
<th>very difficult to use</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8 (57%)</td>
<td>6 (43%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

Table 3. Number and percentage of student responses to question 1.

<table>
<thead>
<tr>
<th>Question</th>
<th>very useful</th>
<th>useful</th>
<th>useless</th>
<th>very useless</th>
</tr>
</thead>
<tbody>
<tr>
<td>The suggestions on the missing sub-themes in</td>
<td>0 (0%)</td>
<td>14 (100%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

- 4 -
your essay provided by this system are:

<table>
<thead>
<tr>
<th>The covered sub-themes in your essay indicated by the system are:</th>
<th>1 (7%)</th>
<th>12 (86%)</th>
<th>1 (7%)</th>
<th>0 (0%)</th>
</tr>
</thead>
</table>

Table 4. Number and percentage of student responses to question 2 and 4.

<table>
<thead>
<tr>
<th>Question</th>
<th>0 (0%)</th>
<th>1-2 (43%)</th>
<th>3-4 (21%)</th>
<th>More than 4 (36%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>How many sub-themes suggested by this system did you use in your essay?</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 5. Number and percentage of student responses to question 3.

<table>
<thead>
<tr>
<th>Question</th>
<th>No.</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you think it is more appropriate to set a maximum for the number of times that one can get suggestions from the system?</td>
<td>9 (64%)</td>
<td>5 (36%)</td>
</tr>
<tr>
<td>Do you think the teacher should continue to adopt this system for your essay writing in the future?</td>
<td>2 (14%)</td>
<td>12 (86%)</td>
</tr>
</tbody>
</table>

Table 6. Number and percentage of student responses to question 5 and 6.

All students agreed that the system suggestions on the missing sub-themes for their consideration to be included in their essays were useful. Fifty-seven percent of them included three or more sub-themes provided by the system into their essay. The rest included at least one. All except one student indicated that the visualization of the sub-themes’ organization in their essays was useful.

Although most of the students are non-computer science students, all of them found the system easy to use. Sixty-four percent of the students preferred no limitation was set on the number of essay submissions for getting feedback from the system while the other preferred to set the maximum number of submissions in the range of 3 to 5. Eighty-six percent of the respondents agreed that the teacher should continue to adopt this system for their essay writing in the future. Those holding opposite view also agreed that the system was useful, but they thought that students should not rely on the system. They have to create the ideas and write the essays on their own.

Interview findings are consistent with the results of questionnaires. Most interviewees expressed that the system was easy to use and the suggested sub-themes were useful. Some further added that the new sub-themes suggested by the system could broaden their views in writing the essays. However, a few of them indicated that the system detected some sub-themes in their essays, but no texts in the essays were highlighted. This was mainly because the system only highlighted those texts whose semantic
similarity coefficient was higher than a preset threshold. This problem can be overcome by adjusting the threshold.

In order to improve our system, the interviewees were also asked to suggest some functions they expected the system would provide in the next version. Many of them suggested adding the grammar checking function. Some further commented that the description of the suggested sub-themes should be in greater detail, not just some keywords, and some examples should also be given. In fact, the descriptions of the suggested sub-themes provided by our earlier prototype were quite detailed (Lee et al., 2006), but we found that some students just copied the suggested sub-theme descriptions into their essays. With the intention of avoiding this ill-practice, the system under this study only provides keywords of the suggested sub-themes.

5. Conclusion

Since the system was well-received by undergraduates, it is worth adapting it for use in other educational sectors, such as secondary schools. The Quality Education Fund Steering Committee of the Hong Kong SAR Government has approved a grant for us to further develop this system and promote it in local secondary schools. Hence, a 6-month evaluation study will be conducted in a secondary school in Hong Kong in August, 2007. It is also hoped that the system can be enhanced by developing its corpus which includes and offer more essay topics for students, and generate some analysis reports for teachers to understand the performance of their students.

REFERENCES

