

Characterizing Students' Behavioral Patterns in an Online Reading Test

Hercy N. H. CHENG*, Jian-Wen SUN, Zhi LIU, San-Ya LIU

National Engineering Research Center for E-Learning, Central China Normal University, China

*hercyheng.tw@gmail.com

Abstract: Understanding students' testing behaviors may help researchers design better computer-based assessment. For this reason, this study aims at characterizing students' behavioral patterns in online reading test by k-means clustering. The clustering algorithm adopts eight indicators: reading time, answering time, the number of choosing articles, the number of choosing questions, the number of selecting options, the number of marking questions, the number of revisiting a test and the final testing scores. The result identifies five clusters of student testers: slow readers, fast readers, question markers, fast responders, and re-readers.

Keywords: Testing behavioral patterns, k-mean clustering, online Chinese reading assessment

1. Introduction

Computer-based assessment not only helps teachers instantly evaluate students' abilities, but also allows students to reflect themselves because of immediate feedback. Besides the efficiency of evaluation, computer-based assessment has better objectivity and lower cost, compared with paper-based tests (Wirth, 2008). Given the new form of assessment, computers may be able to evaluate more competence that traditional tests cannot easily capture. For example, since last year, PISA started to adopt computers to evaluate students' ability to solve problems collaboratively. This is because computers are able to record and analyze students' behaviors of reading and answering questions. Previous research evaluated students' reading literacy skills by analyzing the number of their actions and reading time (Gil, Martinez, & Vidal-Abarca, 2015).

However, because of the complexity, students might not find an optimal solution easily in a computer-based testing environment (Sager et al. 2011). In a sense, students' behaviors are likely influenced. Fortunately, computers' ability to record students' actions allows us to analyze their testing behaviors, so that we may understand how computers influence students. For this reason, this study attempts to characterize students' behaviors in an online reading test. There are several methods for behavior analysis, such as clustering, frequent sequence mining, lag sequential analysis, and so forth. This study adopts clustering because clustering is an unsupervised method, which can classify data according to their natural properties. In this case, the study aims to classify student testers according to their online behaviors. If we can identify different clusters of online student testers, we may design more helpful online learning assessment in the future.

2. Method

2.1 Participants and setting

The participants were 120 fourth graders (73 male and 47 female students) and 120 fifth graders (64 male and 56 female students) from a primary school in Hubei province, China. The study was carried out in a computer lab in the primary school. Because of its capacity, there were 60 students each time. One researcher managed a test with four experts for solving possible technical problems. The participants were required to take at least one test. However, they were allowed to take several times, because the questions were randomly selected. Even so, this study only analyzed their first time of valid tests.

2.2 Online Chinese Reading Test System

An online Chinese reading test system was adopted in this study. The system was designed to evaluate students' Chinese reading abilities. In each test, there are three Chinese articles with 18 items in total. After starting a test, students have to choose one of the three articles. Figure 1 illustrates the interface of one test. Students are allowed to read the article and choose an item for answering. The items are multiple-choose questions with four options. They were also allowed to mark any items that they want to come back later. When they submit their answers, they are prompted to make sure they want to submit. If they want to revisit the articles and their answers again, they are allowed to go back to the test. Students' actions are automatically recorded in the database.

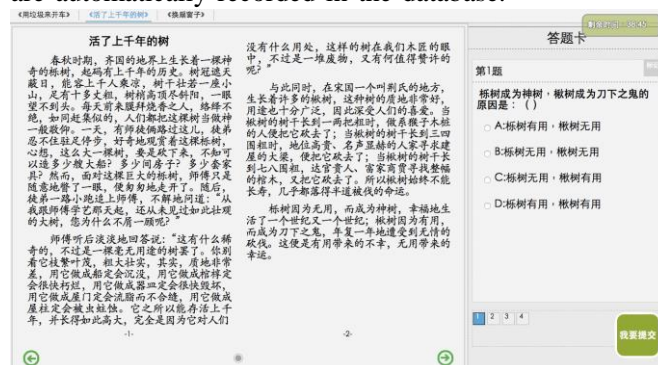


Figure 1. The online Chinese reading test system.

2.3 Data Analysis

This study adopted k-means clustering to characterize students' online testing behaviors. More specifically, the clustering algorithm used the following eight indicators: reading time, answering time, the number of choosing articles, the number of choosing questions, the number of selecting options, the number of marking questions, the number of revisiting a test and the final testing scores. The definitions of the eight indicators are described in Table 1. The value of k was determined by the formula of q value in the study of Frias-Martinez, Chen, Macredie, and Liu (2007). In this case, five clusters were an optimal result.

Table 1: The indicators of the clustering algorithm.

Indicators	Descriptions
Reading time	The time period between the first time that students choosing an article and choosing the first question.
Answering time	The time period between students choosing a question and choosing an option.
The number of choosing articles	The total times of students choosing articles in a test.
The number of choosing questions	The total times of students choosing questions in a test.
The number of selecting options	The total times of students selecting options in a test.
The number of marking questions	The total times of students marking questions in a test.
The number of revisiting a test	The total times of students deciding to revisit the test after submitting it.
Scores	The final testing scores that indicate students' reading abilities.

3. Results

As shown in Table 1, there were five clusters of behavioral patterns in online Chinese reading tests. ANOVAs show that the eight indicators can all significantly differentiate the clusters. Post hoc analysis further indicates the significant differences among clusters. Among the five clusters, the students in the first cluster spent the longest time on reading articles and questions. Therefore, the cluster was labeled as slow readers. The reason perhaps was that they frequently checked the articles and questions. They also tended to change the options of their answers.

Compared with the first cluster, the students in the second cluster spent the shortest time on reading. The cluster was labeled as fast readers. They did not frequently check the questions or change options. In the end, they had significantly lower scores than the other four clusters, implying that they had a worse testing behaviors or they might lack sufficient reading ability.

Although there were only 8 students in the third cluster, they demonstrated explicitly different behaviors. More specifically, they marked questions more than the other four clusters in an online test. Furthermore, they checked as many question as the first clusters, but significantly more than the other three clusters. They are thus labeled as question markers. However, they spent less reading and answering time, implying that they might not pay sufficient attentions on reading articles and questions.

The students in the fourth cluster spent similar time on reading articles to the other clusters, but they spent the shortest time on answering questions. Like the fast readers, they did not check the articles again. Although the fourth cluster shared similar behaviors with the fast readers, they had better scores than the fast readers, suggesting that the fourth clusters had sufficient reading abilities. For this reason, the fourth cluster is labeled as fast responders.

The students in the fifth cluster checked the articles the most frequently, while their reading time is not significantly different from those of other clusters. Meanwhile, they did not frequently check the questions and options, suggesting that they mainly read articles instead of examining answers. For this reason, they are labeled as re-readers.

Table 2: The result of clustering.

Cluster	N	Reading Time (s)	Answering Time (s)	# of Articles	# of Questions	# of Option	# of Mark	# of Revisit	Score
Slow Readers (1)	21	194.22	957.02	8.75	44.59	30.23	1.45	0.43	84.61
Fast Readers (2)	45	74.08	581.90	4.56	20.65	19.21	0.33	0.20	52.59
Question Markers (3)	8	113.45	734.45	6.75	40.97	21.13	19.49	0.75	68.89
Fast Responders (4)	113	142.73	501.87	4.74	21.17	20.27	0.68	0.41	84.99
Re-readers (5)	53	124.55	751.95	10.22	27.08	22.41	0.72	0.79	80.20
F		6.76**	30.10**	80.22**	58.55**	40.90**	192.26**	3.99**	41.46**
MSE		0.91	0.67	0.43	0.51	0.60	0.24	0.96	0.60
Post hoc Comparison		1>2,3	1>2,3,4,5 5>2,4 3>4	1,5>2,3,4 3>2,4	1,3>2,4,5 5>2	1>2,3,4,5 5>2	3>1,2,4,5	N/A	1,3,4,5>2

Acknowledgements

The authors would like to thank the National Social Science Fund Project of China (14BGL131) as well as the Minister of Education of China and Central China Normal University for financial support (CCNU16A02022, CCNU15A06073, CCNU16JYKX030, CCNU15A05010, CCNU15A06072).

References

- Gil, L., Martinez, T., & Vidal-Abarca, E. (2015). Online assessment of strategic reading literacy skills. *Computers & Education*, 82, 50-59.
- Sager, S, Barth, CM, Diedam, H, Engelhart, M, & Funke, J. (2011). Optimization as an analysis tool for human complex problem solving. *Journal of Optimization*, 21(3), 936-959.
- Wirth, J. (2008). Computer-based tests: Alternatives for test and item design. In J Hartig, E Klieme, & D Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 235-252). Cambridge/Göttingen: Hogrefe & Huber Publishers.
- Frias-Martinez, E., Chen, S., Macredie, R., & Liu, X (2007). The role of human factors in stereotyping behavior and perception of digital library users: A robust clustering approach. *User modeling and User-Adapted Interaction*, 17, 305-337