

# A Preliminary Usability Evaluation for Online Annotation and Student Clustering System

Miao-Han CHANG<sup>a</sup>, Rita KUO<sup>b\*</sup>, Fathi ESSALMI<sup>c</sup>, Maiga CHANG<sup>a</sup>, Vive KUMAR<sup>a</sup>, & Hsu-Yang KUNG<sup>d</sup>

<sup>a</sup>*School of Computing and Information System, Athabasca University, Canada*

<sup>b</sup>*Department of Computer Science and Engineering, New Mexico Institute of Mining and Technology, USA*

<sup>c</sup>*Research Unit of Technologies of Information and Communication, University of Kairouan, Tunisia*

<sup>d</sup>*Department of Management Information System, National Pingtung University of Science and Technology, Taiwan*

\*rita.mcsl@gmail.com

**Abstract:** This research evaluated an annotation recommendation system – GRACE (General Rapid Annotation Clustering Enhancement) – which gives suggestions of what annotations might be missing when students were reading the teaching materials. A pilot is done in two computer science classes in a college in Tunisia. The results show that system’s usability is acceptable for students doing online reading activities on GRACE.

**Keywords:** Annotation, bio-inspired approach, clustering, usability, diffusion of innovation

## 1. Introduction

When students read learning materials like textbook, they usually take notes and highlight important words/sentences on it (Hoff, Wehling, & Rothkugel, 2009; Chen & Chen, 2014). Students have different annotation preferences while reading; for instance, they may annotate words in different ways (e.g., underlining, highlighting, or double underlining). However, they may intend to not annotate some important keywords or simply overlook those words while reading an article. When they review the annotated texts for preparing exams and doing homework, they might skip those un-annotated content on the texts as they believe that all important content have been well annotated by them earlier.

## 2. Online Annotation Systems

The research team has developed GRACE (General Rapid Annotation Clustering Enhancement) platform that can group students based on their annotations and provide annotation suggestions for them on the missing content. After teachers created reading activities, students can find these activities on the GRACE. Figure 1 shows the interface for reading activity on GRACE. At the top panel, students can choose the annotation styles they prefer, including highlight, underline, bold, italics, increase font size and change font color. They can also use mixed annotations, such as the use of bold face and underline at the same time for the words “England that this” (see Figure 1a).

Students can decide whether or not they would like to receive the recommendation(s) by checking and unchecking the “Hidding suggestions” check-box on the top. If the recommendation function is enabled, students will get the annotation suggestions for the paragraph when they finish annotating it. Take Figure 1a for example, the student is annotating “wheelbarrow”, which is also the last word in the first paragraph. GRACE will retrieve all students’ annotations and group the student with others who have similar annotation behaviours. It then finds out “chemicals and wastes” were annotated by others in the

same group which this student belongs to; therefore, it reminds the student that he or she might be missed to annotate those words (see Figure 1b).

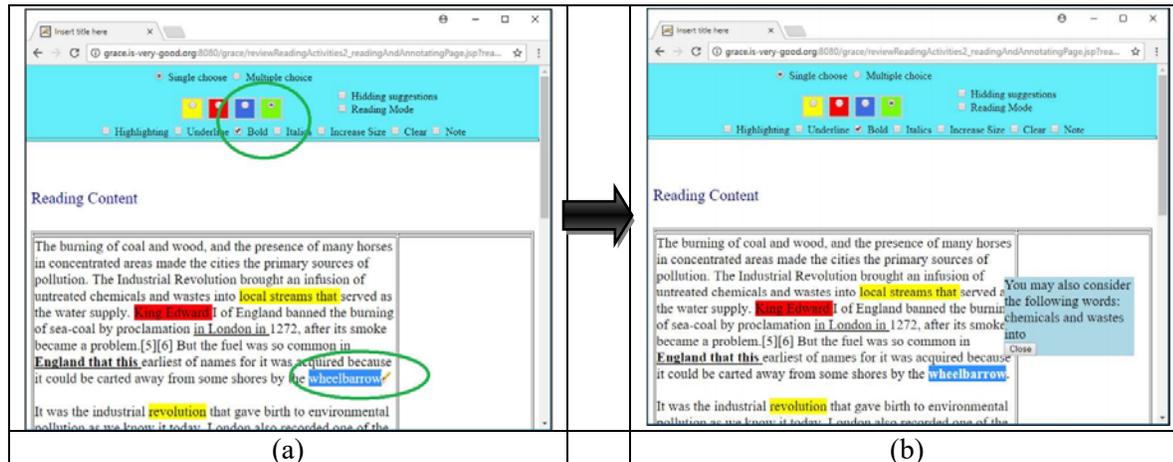


Figure 1. The interface for reading activity on GRACE.

### 3. Research Methods

#### 3.1 Evaluation Process

To verify the system's usability and the influence of students' diffusion of innovation in using GRACE, this research proposed a research model (see Figure 2). The research adopts the System Usability Scale (SUS) to evaluate GRACE's usability (Brooke, 1996) and also adds a few questions from (AlQudah, 2014), (Lewis, 1995), and (Lund, 2001) as well as several self-developed items to the questionnaire. In addition, the Perceived Enjoyment (PE) factor is designed to understand how satisfy students were using GRACE. The research also investigates whether or not students' acceptance of using new technology like e-reader for reading activities will influence their overall satisfaction toward GRACE. The research uses two of the factors in Diffusion of Innovation proposed by Rogers (Rogers, 2003) and adopts the questions designed by (Park & Chen, 2007) and (Quadir, Yang, Chen, & Shih, 2017).

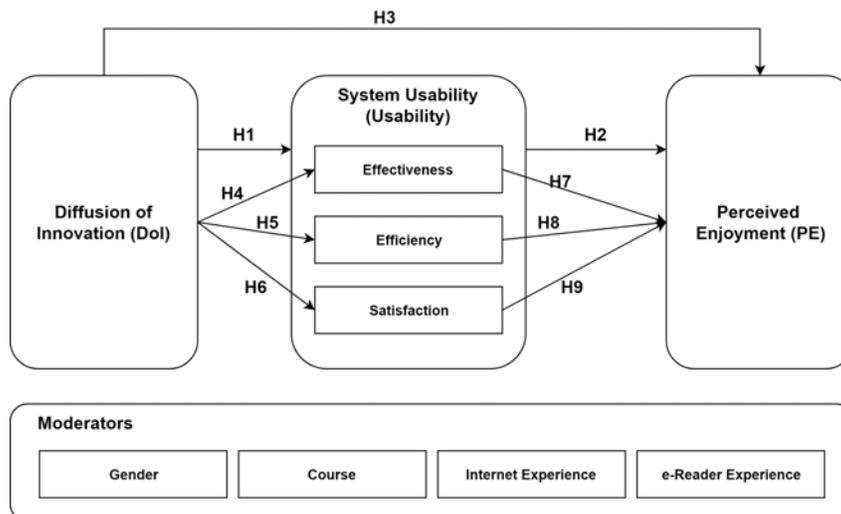


Figure 2. The research model.

To find out students' perceived usefulness toward GRACE as well as what are the influential factors, the research team has conducted a pilot in two computer science classes in a college in Tunisia with nine valid samples collected. The validity and reliability for the two major factors (Diffusion of Innovation and Perceived Enjoyment) and the three sub-factors (Effectiveness, Efficiency, Satisfaction) in the System Usability factor are in excellent ( $\alpha \geq 0.9$ ) or good ( $0.8 \leq \alpha < 0.9$ ) and shows the questionnaire is reliable (George & Mallery, 2010).

### 3.2 Findings

To understand whether or not the system is useful to students, we calculated the SUS value based on Brooke's study in 1996. The average value of SUS is 55.556 which shows that the system's usability is *acceptable* for students doing online reading activities on GRACE (Bangor, Kortum, & Miller, 2009).

The research team also discovered some unexpected findings. First of all, students' perceived complexity of using any e-reader system has no significant relation with the usability score that they gave for GRACE. The finding tells us that even a student who believes an e-reader is a complex system, he or she might still believe GRACE is a useful system. Another finding shows that students' perceived ease of use toward GRACE is not affecting their intention of using the system, but students' perceived usefulness is. The finding suggests that if a system is useful for its users, even it is at the development stage and gives only fair user experience, users might still want to use the system to improve the efficiency or effectiveness of their works.

Another interesting finding is that students who like to try a new technology before officially adopt it (i.e., the triability factor in the diffusion of innovation) have higher satisfaction toward GRACE and have more intention to use it in the future. However, there is no significant relation found between the triability and perceived ease of use nor perceived good user interface design toward GRACE. This finding indicates that users treat triability is important for them to adopt new technologies and a system with triability can make users have higher intention of using it, even when it has only fair enough user interface design.

## 4. Conclusion

The developed bio-inspiration clustering algorithm is more stable than the previous algorithm designed by the research team. However, the precision and recall rate is still not good enough. Because the efficiency and effectiveness of GRACE highly affect students' intention of using it, giving more relevant annotation recommendations for students is extremely important. The research team will review the clustering algorithm in order to give more appropriate annotation recommendations for students.

The sample size of the study is small but is still sufficient for the Human Computer Interaction studies, which is  $10 \pm 2$  samples defined in Hwang and Salvendy's research in 2010. However, if the research team wants to find out more potential issues to improve the system, an evaluation with a larger sample size is required (Nielsen, 2012). On the other hand, this study only discuss students' perceived efficiency toward the system with questionnaire. To find out whether or not GRACE is really useful for students studying, pre-test and post-test should be integrated in the new evaluation plan.

## References

- AlQudah, A. A. (2014). Accepting Moodle by academic staff at the University of Jordan: Applying and extending TAM in technical support factors. *European scientific journal*, 10(18), 183-200.
- Bangor, A., Kortum, P., and Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3), 114-123.
- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7.
- Chen, C. M., & Chen, F. Y. (2014). Enhancing digital reading performance with a collaborative reading annotation system. *Computers & Education*, 77, 67-81.
- George, D., & Mallery, P. (2010). *SPSS for windows step by step: A simple guide and reference 18.0 update* (11th ed.). Boston, MA: Allyn & Bacon.

- Hoff, C., Wehling, U., & Rothkugel, S. (2009). From paper-and-pen annotations to artefact-based mobile learning. *Journal of Computer Assisted Learning*, 25(3), 219-237.
- Hwang, W., & Salvendy, G. (2010). Number of people required for usability evaluation: the 10±2 rule. *Communications of the ACM*, 53(5), 130-133.
- Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1), 57-78.
- Lund, A. M. (2001). Measuring usability with the USE questionnaire. *Usability Interface*, 8(2), 3-6.
- Nielsen, J. (2012). How many test users in a usability study? Retrieved 21 May 2019, from <https://www.nngroup.com/articles/how-many-test-users/>
- Park, Y., & Chen, J. V. (2007). Acceptance and adoption of the innovative use of smartphone. *Industrial Management & Data Systems*, 107(9), 1349-1365.
- Quadir, B., Yang, J. C., Chen, N. S., & Shih, M. J. A. (2017). The Effects of Perceived Innovation Game Attributes by Learners on Learning Performance in a Game-Based Achievement Learning System. In *Innovations in Smart Learning* (pp. 155-164). Singapore: Springer.
- Rogers, E. M. (2003). *Diffusion of innovations* (5<sup>th</sup> ed.). New York, NY: Free Press.