Chen, W. et al. (Eds.) (2017). Proceedings of the 25th International Conference on Computers in Education.
New Zealand: Asia-Pacific Society for Computers in Education

Using Learning Analytics to Support Computer-Assisted Language Learning

Huiyong LI[a*], Hiroaki OGATA[b], Tomoyuki TSUCHIYA[c], Yubun SUZUKI[c], Satoru UCHIDA[c], Hiroshi OHASHI[d] & Shin'ichi KONOMI[d]
[a]*Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan*
[b]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
[c]*Faculty of Languages and Cultures, Kyushu University, Japan*
[d]*Faculty of Arts and Science, Kyushu University, Japan*
*lihuiyong123@gmail.com

Abstract: Computer-assisted language learning (CALL) is often used as an approach to foreign language teaching and learning in higher education. The CALL course is offered at a national university in Japan to allow freshman students to perform self-regulated learning with e-learning materials for the purpose of developing language skills. However, as novice self-regulated learners, freshman students have low self-regulation skills and they are more likely to obtain lower achievement. In addition, it is difficult for instructors to grasp students' learning situation due to the large amount of evaluation work. Therefore, in this research, a total of 7,413,397 learning logs were analyzed, which were collected from 2,499 students' learning interactions in the CALL course. After that, a learning support system for freshman students is proposed. The system is provided for students and instructors through the learning dashboard. On the one hand, students can conduct self-monitoring and reflect their behaviors in a visual way. On the other hand, instructors can identify learning behavioral patterns and grasp individual learning situation to provide one-on-one instructions.

Keywords: Learning analytics, self-regulated learning, computer-assisted language learning

1. Introduction

Computer-assisted language learning (CALL) is often used as an approach to foreign language teaching and learning in higher education. The CALL course is widely offered at universities in Japan to allow freshman students to perform self-regulated learning (SRL) with e-learning materials to develop language skills in grammar, listening, and reading.

SRL is an active learning process used to regulate and monitor learning cognition, motivation, and behavior, thereby setting personal goals (Wolters, Pintrich & Karabenick, 2005). In the process of SRL, freshman students often do not recognize how they are learning and thus do not appreciate many beneficial learning strategies (Bjork, Dunlosky, & Kornell, 2013). The learning behavioral differences between novice and skillful self-regulated learners reveal that novice self-regulated learners avoid self-evaluation and have negative self-reactions (Schunk & Zimmerman, 1998). The students with low SRL skills are more likely to obtain lower achievement, and therefore it is critical to provide adaptive support to improve freshman students' SRL skills.

In addition, it is difficult for instructors to grasp students' learning situation due to the large amount of evaluation work. Since a large number of students and online learning materials need be paid attention to, it is limited for instructors to provide individual instruction to students in the CALL environment.

Learning analytics studies could be applied to improve education, create learning supports, establish learning models, and so on (Gray, 2014). Using learning analytics and the massive learning logs in current CALL course, it is able to support instructors and students efficiently. In this research, the learning logs were collected and extracted from the server of the CALL course, and time-based indicators from students' log data were analyzed to understand the learning activities. Utilizing time-based indicators, a learning support system on the CALL course is proposed. The aims of the system are to support students' self-monitoring and instructors' decision making easily.

2. Related Work

2.1 SRL in a Computer-based Learning Environment

SRL is defined as an active and constructive process through which learners can set goals, and monitor and control their cognition, motivation, and behavior (Pintrich, 2000). It is also characterized as a self-directive process as self-beliefs enable learners to transform their academic abilities (Zimmerman, 2008). Winne and Hadwin (1998) proposed that SRL included four phases: defining the task, setting goals and plans, enacting tactics, and adapting metacognition. Therefore, learners need to analyze the learning context and define tasks, set the appropriate learning goals and make plans, select the effective learning strategies to use, monitor the whole learning process, and evaluate their learning performance.

Previous studies indicated that SRL is the crucial skill for success in computer-based learning environments (Adeyinka & Mutula, 2010). However, learners cannot always regulate themselves successfully because of reasons such as lack of good strategy use, lack of metacognitive knowledge, failure to control of metacognitive processes, or lack of experience in learning environments with multiple representations. Thus, how to foster SRL ability has become a central issue in the field of education research and practice.

In order to support SRL in the CALL course, instruments that capture students' self-regulation are critical. Most studies on self-regulated learning have used self-report instruments, which not only are intrusive but also are limited to capturing actual self-regulated behaviors in learning contexts. However, this issue can be resolved via online trace data use, and such technologically mediated learning environments enable the collection of a comprehensive set of student learning behaviors that occur in learning environments (Pardo, 2014).

2.2 Learning Analytics

Learning analytics are driven by the collection and analysis of traces that learners leave behind (Greller & Drachsler, 2012). It can help to understand and optimise the learning process and the environments in which it occurs (Siemens & Long, 2011; Ogata & Mouri, 2015). Until now, learning analytics are mostly feedback to the users in web-based learning dashboards (Verbert et al., 2014). Those dashboards can support raising awareness and reflection of individual and peer performance, suggest additional learning activities or content and therefore can have an impact on the learning behavior. For instance, monitoring the state in a learning activity can motivate the learner towards the accomplishment of a learning goal. This cognitive process has been defined as “self monitoring”, and “understanding how to learn” (Candy, 1991).

However, there were few research conducted using massive learning data (2499 students’ learning data) in the field of learning analytics. Further, considering the fact that students participating in online learning exhibit a lack of time management regarding self-regulated learning, such as cramming and procrastinating, and it is therefore critical to provide individual feedback to help students take proactive actions. In this research, the massive learning data were analyzed in order to understand the actual behaviors of all students, and then a learning support system was proposed through the learning dashboard in order to provide prompt feedback to support students’ self-monitoring and instructors’ decision making.

3. Research Methods

3.1 Setting and Participants

A total of 2,631 students enrolled in 50 CALL classes at Kyushu University, our institution. The CALL classes were provided for freshman students from all departments with two credits from spring and fall semesters 2016. The students were supposed to perform self-paced language learning outside of the classroom. Table 1 shows the course schedule in the spring semester 2016. To increase the motivation of the students, four sub-deadlines were set in one semester. The students were required to complete the assigned materials from stage 1 to stage 3, with those for stage 4 meant as an option.

The 93 students (3.53%) who scored 520 or more on the semester-initial TOEFL-ITP applied for exemption from the CALL course. Additionally, there were 39 dropout students (1.48%), who did not access the learning materials during a whole semester. Thus, the exempted students and dropout students were removed, and the remaining 2,499 students (94.99%) participated in this research.

The e-learning materials of the CALL course contained grammar, listening, and reading sections and included 493 units with a total of 751 quiz items. The quiz items were uploaded for students to study at the beginning of each semester and all of the learning activities occurred online.

Table 1: Course schedule in the spring semester 2016

		Learning materials assigned		
Stage	Deadline	Reading	Listening	Grammar
1	Week 5	Reading1	Listening1	Grammar1
2	Week 10	Reading2	Listening2	Grammar2
3	Week 15	Reading3	Listening3	Grammar3
4(optional)	Week 21	Reading4	Listening4	Grammar4

3.2 Data Source

The data used in this research were collected from the CALL course server. The moment students practiced quiz items online, the learning behaviors were recorded in server logs concurrently. There were three types of learning logs in the CALL course server, including access to learning materials (access logs), completed quiz items (completion logs), and quiz answers (answer logs). A total of 7,413,397 learning logs were retrieved and analyzed from the server of the CALL course with 1,792,277 access logs, 1,117,375 completion logs, and 4,503,745 answer logs. The details of learning sessions were stored in logs, including user identifier, learning material id, quiz item id, access start time, access end time, completion flag, completion time, right answer id, and selected answer id.

3.3 Data Preparation and Analysis Procedures

Figure 1 illustrates the data preparation, analysis, and feedback procedures on the analysis server.

First, Search Query Language (SQL) queries were conducted to retrieve log data from the CALL course server and then log data were saved on SQL Server 2012-a database management system.

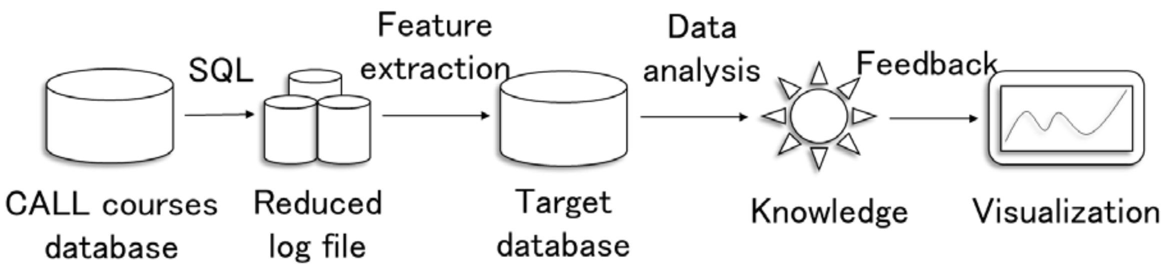

Figure 1. Data preparation, analysis, and feedback procedure

Second, the feature extraction phase was performed using reduced log files, which were cleaned by removing all useless, irregular, and incomplete data from the original CALL course logs. For example, sometimes the end time was not recorded when a student accidentally closed the web browser, or a student might do nothing for a long time with the website left open. In these cases, the related raw data were removed from the log files to reflect only normal learning activities of the students. Four primary variables were extracted through calculating or accumulating reduced log files: (1) Daily access items. (2) Daily access time. (3) Daily completion items. (4) Daily answer lists.

Third, the primary variables stored in the target database were analyzed and interpreted. The analysis phase included two sub-phases: (1) Descriptive analysis: an overview on the dataset was generated by summarizing and classification. The ratios of task completion at four stages were summarized from the primary variable of daily completion items. Further, the active days of the students on the course were cumulated from the primary variable of daily access items, then the students were classified by the active days. (2) Correlation analysis: The relationship between the total time spent on the course and quiz scores was investigated from the primary variables of daily access time and daily answer lists.

Finally, based on the analysis results of massive log data, learning dashboard was provided. The learning dashboard focused on the learning processes and tendencies in the CALL course.

4. Results

4.1 Overview of Data Analysis

Table 2 summarizes the ratio of task completion at four stages (N=2499). The ratios of tasks completed at stage 1, stage 2, and stage 3 were 98.08%, 93.20%, and 88.52%, respectively. In contrast, the number at stage 4 significantly declined to 42.70%, since the learning materials of stage 4 were not required but optional. The results reveal that most of the students completed tasks at the required stages but 11.48% (n=287) of the students were still not completed after the third required stage. Moreover, 57.30% (n=1432) of the students stopped learning the materials at the optional stage. The number of students who dropped out at each stage increased continuously, thus the students who dropped out might need more support.

Table 2: Course stages and the ratios of task completion

Stage	Total	Task completed		Not completed	
	N	N	%	N	%
1	2499	2451	98.08%	48	1.92%
2	2499	2329	93.20%	170	6.80%
3	2499	2212	88.52%	287	11.48%
4(optional)	2499	1067	42.70%	1432	57.30%

The active days of each student were also investigated. One active day was defined as a day on which one student accessed learning materials. About half of the students (50.82%) accessed learning materials with 11 to 20 active days, and most of the students (90.88%) accessed learning materials with 30 active days or less. Since the duration of the CALL course was 21 weeks, most of the students (90.88%) performed self-regulated learning with a low degree (fewer than 22% of available days) of the learning activeness. Thus, the most of students might need more support to participate in the course.

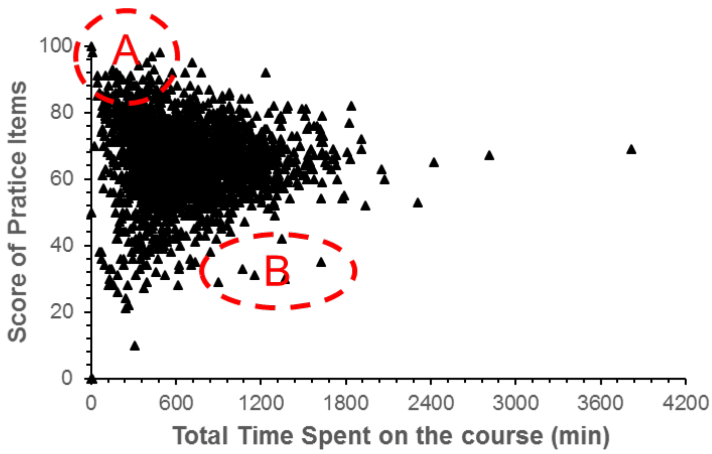

Figure 2. Scatters of total time spent on the course and quiz scores

To investigate the quality of the learning outcomes, a scatter plot is generated. Figure 2 illustrates the relationship between the total time spent on the course and quiz scores. The results reveal that the students in the area A spent less time than others on the course but still obtained high scores. In contrast, the students in the area B spent above-average time on the course but scored less than 40, thus they might need more attention and support.

4.2 Learning Dashboard

From the results of behavioral analyses, several learning patterns of the students were identified. It was clearly necessary to pay special attention to the students who need guidance. Therefore, the learning dashboard for students and instructors was developed.

The learning dashboard is shown in Figure 3, which contains two parts: learning progress and behavioral tendencies. (1) The left part is a line chart to show the learning progress. One curve represents one student and the green curve means the class average of learning progress. The horizontal axis shows the dates and the vertical axis displays the counts of completed quiz units. (2) The right part is a parallel coordinate to present the overall tendencies of learning behaviors. The vertical axes show the quiz scores, the number of completed units, the active days, and the total time spent, respectively. One polyline connecting the vertices on the vertical axes represents one student. Meanwhile, the quiz scores are mapped to a six-color ribbon. When the quiz score is higher, the color of the polyline will be mapped from green to dark red.

For students, personal learning progress and personal behavioral tendencies were shown on the dashboard. They could support self-monitoring in a visual way by showing learning process and result. In addition, comparing personal behaviors to that of class average, the student easily found the position of self, leading to increased motivation and participation.

For instructors, learning progress in one class and behavioral tendencies in one class were shown. The constructors can see the student id and name by moving the mouse pointer to one line, and therefore it is easy to identify the students' learning patterns and locate the student who need guidance.

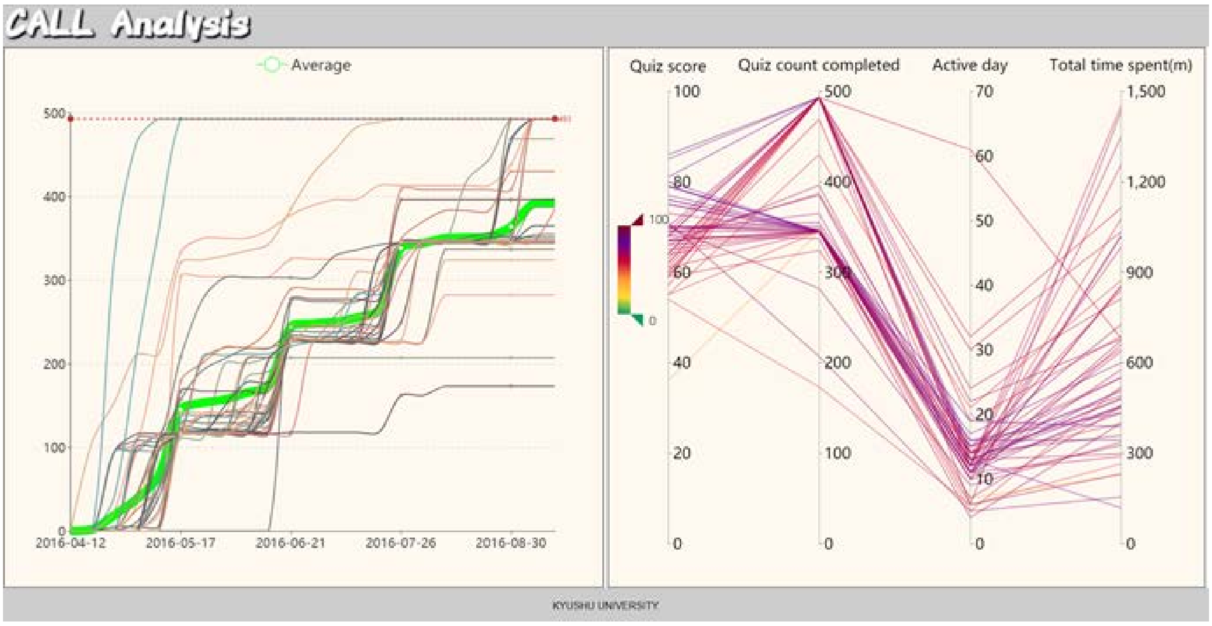

Figure 3. Learning dashboard

5. Conclusion and Further Work

In this research, a learning support system for the CALL course in higher education was proposed. A total of 7,413,397 learning logs collected from 2,499 students' behaviors were analyzed. The results revealed that 11.48% of the freshman students still drop out during the SRL processes even on the mandatory courses in CALL, and most of the freshman students (90.88%) performed SRL with a low degree (fewer than 22% of available days) of the learning activeness. The proposed system was based on actual learning logs, and it focused on the learning processes and tendencies. The system was provided for not only students but also instructors through the learning dashboard. On the one hand, students could conduct self-monitoring and reflect their behaviors in a visual way. On the other hand, instructors could identify learning behavioral patterns and grasp individual learning situation to provide one-on-one instructions. Moreover, the learning dashboards of the learning support system were easy to use on a variety of web browsers, since they were developed with JavaScript, which was a major browser scripting language.

For future work, the evaluation of usefulness and impact of the learning support system will be conducted. Further, in order to provide feedback more efficiently and identify at-risk students sooner, the learner models on learning progress and behavioral tendencies will be constructed.

Acknowledgements

The research has been partly supported by "Research and Development on Fundamental and Utilization Technologies for Social Big Data" (178A03), the Commissioned Research of the National Institute of Information and Communications Technology (NICT), Japan; and Grant-in-Aid for Scientific Research (S) No. 16H06304; Grant-in-Aid for Scientific Research (B) No. 25282059; Grant-in-Aid for Challenging Exploratory Research No. 26560122 from MEXT, Japan.

References

Adeyinka, T., & Mutula, S. (2010). A proposed model for evaluating the success of WebCT course content management system. *Computers in Human Behavior*, *26*(6), 1795-1805.

Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist*, *40*(4), 199-209.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual review of psychology*, *64*, 417-444.

Candy, P. C. (1991). *Self-Direction for Lifelong Learning. A Comprehensive Guide to Theory and Practice*. Jossey-Bass, 350 Sansome Street, San Francisco, CA 94104-1310.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society, 15*(3), 42.

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education*, *104*, 18-33.

Ogata, H., & Mouri, K. (2015, July). Connecting dots for ubiquitous learning analytics. In *International Conference on Hybrid Learning and Continuing Education* (pp. 46-56). Springer International Publishing.

Ogata, H., Yin, C., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based Learning Analytics in University Education. In *International Conference on Computer in Education (ICCE 2015)* (pp. 401-406).

Pardo, A. (2014). Designing learning analytics experiences. In *Learning Analytics* (pp. 15-38). Springer New York.

Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of educational psychology*, *92*(3), 544.

Schunk, D. H., & Zimmerman, B. J. (Eds.). (1998). *Self-regulated learning: From teaching to self-reflective practice*. Guilford Press.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review, 46*(5), 30.

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., & Klerkx, J. (2014). Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing, 18*(6), 1499-1514.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. *Metacognition in educational theory and practice*, *93*, 27-30.

Wolters, C. A., Pintrich, P. R., & Karabenick, S. A. (2005). Assessing academic self-regulated learning. In *What Do Children Need to Flourish?* (pp. 251-270). Springer US.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American educational research journal*, *45*(1), 166-183.