The International Workshop on Learning Analytics and Educational Data Mining (LAEDM 2016) In conjunction with CRIWG/CollabTech 2016

Kanazawa, Japan

September 14, 2016

Preface

The aim of the workshop is to bring together international scholars and researchers working on various aspects of interest regarding recent developments of big data analytics in educational fields. In this workshop, a keynote speech is given and a total of 11 papers are accepted for presentation. The topics of the papers presented in this workshop include learning technologies and tool developments for personalized and adaptive learning, deployment and implementation of learning analytics, theoretical and pedagogical issues, and privacy aspects of learning analytics.

Workshop Organizer

Professor K. Robert Lai, Center for Big Data and Digital Convergence, Yuan Ze University, Taiwan
Professor Weiqin Chen, College of Applied Science, Oslo and Akershus University, Norway

Program Committee

Gwo-dong Chen, National Central University, Taiwan
Yong-Sang Cho, Korean Education and Research Information Service, South Korea
Kenji Hirata, University of Tokyo, Japan
Tore Hoel, Oslo and Akershus University College of Applied Sciences, Norway
Ulrich Hoppe, University of Duisburg-Essen, Germany
Jon Mason, Charles Darwin University, Australia
Hiroaki Ogata, Kyushu University, Japan
Shu-Fen Tseng, Yuan Ze University, Taiwan
Barbara Wasson, University of Bergen, Norway

Program

Date: September 14, 2016

Venue: Seminar Room A of Shiinoki Cultural Complex, Ishikawa Prefecture

10:00-12:30 morning session (Section chairs: Professor K. Robert Lai)

10:00-10:30 Keynote speech (Professor Dai Griffiths, University of Bolton, United Kingdom)

Title: **Learning analytics state-of-affairs – summing up three years of experience building a community of stakeholders in schools, universities and the workplace**

10:30-12:30 Paper presentation session

ID	Authors and paper title	Page
1	Ren-Hao Pan, Van Lam Ho, Phan Dinh Van, Hsiu-Chen Hsu, Robert K Lai and Chien-Lung Chan. *Development of daily activity and lifestyle data visualization tool for the college student learning analytics*	1
8	Sébastien ANDRÉ , Atsushi SHIMADA and Hiroaki OGATA. *Automatic Summarization System of Lecture Slides*	6
11	Ryo ABO, Takayuki KOGA, Izumi HORIKOSHI, Kimiaki YAMAZAKI and Yasuhisa TAMURA. *Data Visualization Framework for Learning Analytics*	8
5	Cheng-Wei Lee, Huan-Yi Pan, Liang-Chih Yu, Chien-Lung Chan, Chih-Yueh Chou and Shu-Fen Tseng. *Academic Performance Estimation by Visualization of Learning Patterns in Campus Environments*	10
9	Shu-Fen Tseng, Liang-Chih Yu, Ho-Shing Wu, and Po-Yao Chao. *Measuring Engagement: Student Profiling and the Effects of Remedial Learning Counseling*	14

12:30-14:00 lunch

14:00-16:30 afternoon session (Section chairs: Professor Tore Hoel)

ID	Authors and paper title	Page
2	Kazuharu Toyokawa. *Monitoring Learning Attitude and Engagement from Students' Free-style Comments via Lexicon-based Analysis*	18
4	Yugo Hayashi. *Collaborative learning in a knowledge integration task using PCAs: Investigation based on dyad eye movements*	22
12	Chen-Yu Lee, Gwo-Haur Hwang, Long-Siang Huang and Hsu-Yu Lee. *The Analysis of Learning Effectiveness of Chinese Proficiency Test Tutorial System Including Capability Indicators*	26
7	Chyi-In Wu. *Computer Assisted Learning to Improve Educational Performance: A Study of Aboriginal Taiwanese Kids*	30
13	Leo A. Siiman and Margus Pedaste. *Response time analytics for evaluating the quality of survey data from school-age students*	33
3	Tore Hoel and Weiqin Chen. *Implications of the European Data Protection Regulations for Learning Analytics Design*	37

Development of daily activity and lifestyle data visualization tool for the college student learning analytics

Ren-Hao Pan
Innovation Center for Big data and Digital Convergence, Yuan Ze University, Taoyuan 330, Taiwan, (R.O.C)
E-mail: pan@51donate.com

Van Lam Ho
Department of Computer Science and Engineering, Yuan Ze University, Taoyuan 330, Taiwan, (R.O.C)
E-mail: s999114@mail.yzu.edu.tw

Phan Dinh Van
Department of Information Management, Yuan Ze University, Taoyuan 330, Taiwan, (R.O.C)
E-mail: dvan2707@due.edu.vn

Hsiu-Chen Hsu
Department of Information Management, Yuan Ze University, Taoyuan 330, Taiwan, (R.O.C)
E-mail: s941402@gmail.com

Robert K. Lai
Department of Computer Science and Engineering, Yuan Ze University, Taoyuan 330, Taiwan, (R.O.C)
E-mail: krlai@saturn.yzu.edu.tw

Chien-Lung Chan
Department of Information Management and Innovation Center for Big data and Digital Convergence, Yuan Ze University, Taoyuan 330, Taiwan, (R.O.C)
E-mail: clchan@saturn.yzu.edu.tw
(Corresponding Author)

Abstract

In this study, 30 college students used a wearable tracking device for 24-hour measurement of daily physical activity and lifestyle factors over a period of 56 days. The collected data included heart rate, number of steps, activity level, sleep-related information, and calories burned. These data were stored on our Hadoop-based in-memory data warehousing platform. A Behavior Visualization System (BVS) was developed to illustrate the students' lifestyle and behaviors using dynamic visualization tools such as line charts, bar charts, etc. A descriptive statistics module, data filtering and exporting tool were also included for advanced activities analysis for specific individuals or groups. An association analysis of physical activity and learning performance was conducted using the BVS, with the criterion that the device was worn for over 20 hours per day. The results revealed that examination performance was associated with sleep duration during the one day prior to examination (Pearson correlation coefficient 0.365, $p<0.05$). This finding suggested that using the wearable device for daily physical activity and sleep tracking may have a beneficial effect on students' learning.

1. Introduction

Previous studies have indicated that day-to-day physical activity (PA) influences concentration level and is highly correlated with disease occurrence [1,8,9]. For example, attention deficit hyperactivity disorder (ADHD) is a syndrome observed in children and is identified by investigation of physical activity, including sleep duration and daily activity level [2]. Jeremy reported a positive association between physical activity and sleep [10,11], and a longer duration spent performing moderate or moderate-to-vigorous physical activity has also been shown to be related to sleep quality [3]. Moreover, a low number of steps taken daily was observed to contribute to the high prevalence of adult obesity in the United States [4]. Sleep efficiency affects the neurocognitive function, attention level and academic performance of students at various levels of education. A lack of sleep related to the prefrontal cortex (PFC) can cause weaknesses in learning, attention, decision-making, and normal neurocognitive functioning [5]. Many instruments have been developed to measure physical activity (PA) and sleep, but a new method for measuring regular PA and sleep and analyzing the associations with cognitive function is needed [6]. The aim of this study was to develop a new approach for investigation of PA and attention capacity. We observed the daily activities of students at Yuan Ze University by utilizing a wearable tracking device and assessed attention capacity in order to examine the relationships of PA and lifestyle with memory and attention capacity.

2. Methods

2.1. Factors

The wearable tracking device used in this research was a Fitbit Charge HR, which is produced by Fitbit, Inc. The tracker can record 24 hours of daily information regarding physical activity and lifestyle, including heart rate, calories burned, steps taken, distance moved, floors climbed, elevation, sedentary duration, duration of light activity, duration of moderate activity, duration of high activity, sleep duration, waking duration, number of awakenings, duration spent in bed, sleep efficiency, sleep start time and depth of sleep. All measured data were synchronized to the Fitbit cloud, then transcribed from the cloud to a Hadoop-based in-memory data warehousing platform. Table1 shows the 19 factors measured, the sampling frequency and the total amount of recorded data derived from 30 students.

Table 1: Factors measured by the tracking device.

#	Factor	Frequency	Total rec-

			ords
1	Heart rate	5 second	10,705,550
2	Calories	minute	2,419,200
3	Steps	minute	2,419,200
4	Distance	minute	2,419,200
5	Floors	minute	2,419,200
6	Elevation	minute	2,419,200
7	Sedentary duration	minute	2,419,200
8	Duration of light activity	minute	2,419,200
9	Duration of moderate activity	minute	2,419,200
10	Duration of high activity	minute	2,419,200
11	Sleep duration	Day	1,680
12	Waking duration	Day	1,680
13	Number of awakenings	Day	1,680
14	Duration spent in bed	Day	1,680
15	Sleep efficiency	Day	1,680
16	Sleep start time	Day	1,028
17	Sleep depth - asleep	Day	470,060
18	Sleep depth - awake	Day	29,420
19	Sleep depth - very awake	Day	3,142

2.2. Tracking Data Collection

Java programming language was used to design programs and access data through the wearable device cloud API (Application Programming Interface) for secure data access. First, data from the wearable device were updated by mobile phone using Bluetooth ver. 4.0. The client app next redirected to the tracker authorization page for data collection authentication. A personal token was obtained from a callback URL and was applied to synchronize the cloud data. When the above steps were completed, a data retrieval program began to collect data via the JSON exchange format and stored it in the data warehousing platform.

2.3. Data Storage & Processing

In this research, a Behavior Visualization System (BVS) was developed for data analysis, which had the ability to manage a group of tracking devices and synchronize multiple tracker measurement data to the data warehouse. The collected data had different sampling frequencies: heart rate was measured every 5 seconds, and calories, steps, distance, floors, elevation, sedentary duration, duration of light activity, duration of moderate activity and duration of high activity were measured every minute. Data obtained daily included sleep duration, waking duration, number of awakenings, duration spent in bed, sleep efficiency, sleep start time, sleep depth – asleep, sleep depth – awake, and sleep depth – very awake.

All data collected were stored on the Hadoop-based in-memory data warehousing platform, Impala. Impala is based on the Hadoop distributed file system (HDFS) and has the ability to access terabyte-level data rapidly using the in-memory data processing technique. Impala SQL can be used to obtain a data summary, implement data queries, and analyze data in the Impala data warehousing platform. Impala can also be employed to explore information in an interactive manner, and to implement reusable batch processing with the advantages of high-performance and low-latency SQL queries.

Compared with the traditional Client-Server architecture, the In-Memory access framework significantly enhanced the ability to manage fast and large data distributed write and read tasks. Efficiency is improved and the amount of time required for data access is reduced; hence, the development of a real-time terabyte-level analytic application was possible.

2.4. Data Processing

In this study, data regarding the daily activities of students were displayed using the BVS as per the data processing workflow shown in Figure 1.

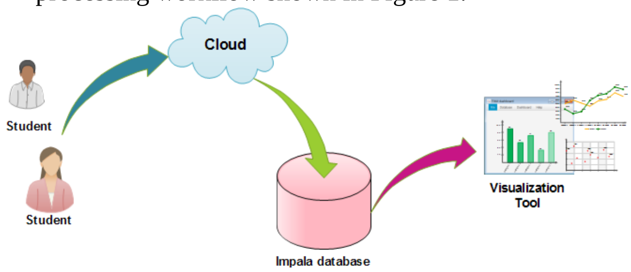


Figure 1: Data processing workflow.

The data transfer procedure from the tracker to the BVS was as follows: Step 1 – data regarding daily activities of students were collected by the tracker and synchronized to the cloud [7]. Step 2: The BVS provides authorization and collects data from the cloud for transcription to the data warehouse. Step 3: Visualization tools implemented using the JavaFX technique were applied in the BVS to illustrate activity and lifestyle data in charts (line, bar, scatter) according to time

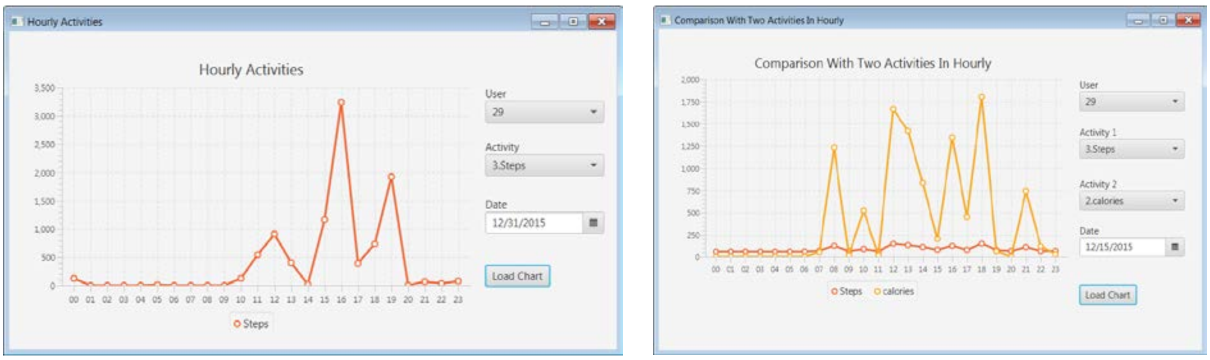


Figure 2: Display of hourly activities.

series and user series. The BVS also provided descriptive statistics for individual or group analysis. For advanced analytics, the BVS enabled data filtering by criterion-setting followed by export to a text file for further analysis.

3. Results

3.1. Materials

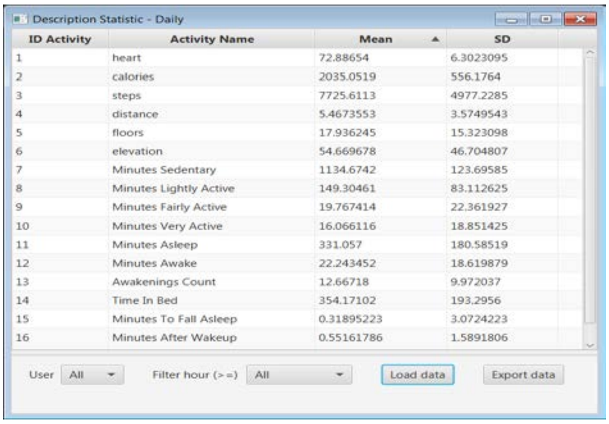
The device used in this study collected various data at different frequencies, e.g., heart rate data were collected every 5 seconds, and other data were collected every minute or every day. Therefore, management of the large amount of data collected (over 30,000 records generated from a single participant per day) posed a challenge, and data representation for all 30 subjects during the study period was also difficult. In addition, the data collected were sometimes incomplete and non-continuous, as day-to-day activities of the students may render the sensor unable to collect data; for example, when the device was removed during bathing, or when the tracker battery was depleted. Thus, if all the data collected were used, the results obtained through analysis would not be close to a representation of the real activities of the students. Therefore, conditions were established to filter missing daily physical activities data. In this study, heart rate data were employed as a condition to filter the final data for analysis. During periods in which the tracker was removed, heart rate data were not obtained, and hence the system must omit the missing data for that period. The system calculated the duration in which heart rate data were obtained in a day, and set the duration for which the tracker was worn as a criterion to provide a filter and enable targeted subject selection. The following analyses were conducted following filtering of the data using the criterion that the daily duration of wear must be greater than 20 hours.

3.2. Activity Profiling

The BVS provides various visualization and profiling options, which include hourly data presentation, daily activities visualization, and weekly statistics. Figure 2 shows a line chart representing hourly activities from 0 to 23 hours for each user and for all users on one day. The system also allows selection of user, activity and date; a line chart is then produced, and comparison between two activities is possible. Figure 3 shows how daily activities can be visualized as activity data obtained during a measurement period in which the students were attending a class; the BVS can illustrate the activity histogram of a selected user or the mean of all users. This system also supports data filtering based on the number of hours, adjusted by user heart rate data per hour and wear length per day. Users of the BVS are able to set a threshold for refreshing of the visualized results in order to observe the changing profile of a histogram, and the filtered data can be exported as a CSV file for further analysis. Figure 4 shows the behavior of the subjects each day for a week. This tool can be used to present data for each user or data for all users for each measured activity. However, the device batteries require recharging after a few days, and other reasons may also result in missing data; hence, a filter for the length of wear per day was required.

Figure 3: Display of daily activities

Figure 4: Display of weekly activities.



ID Activity	Activity Name	Mean	SD
1	heart	72.88654	6.3023095
2	calories	2035.0519	556.1764
3	steps	7725.6113	4977.2285
4	distance	5.4673553	3.5749543
5	floors	17.936245	15.323098
6	elevation	54.669678	46.704807
7	Minutes Sedentary	1134.6742	123.69585
8	Minutes Lightly Active	149.30461	83.112625
9	Minutes Fairly Active	19.767414	22.361927
10	Minutes Very Active	16.066116	18.851425
11	Minutes Asleep	331.057	180.58519
12	Minutes Awake	22.243452	18.619879
13	Awakenings Count	12.66718	9.972037
14	Time In Bed	354.17102	193.2956
15	Minutes To Fall Asleep	0.31895223	3.0724223
16	Minutes After Wakeup	0.55161786	1.5891806

Figure 5: Descriptive statistics for daily activity.

3.3. Descriptive Statistics

Using the statistics module, the developed tool can be used to calculate various descriptive statistical indicators, such as the means and standard deviations of all activities of each user or of all users during the research period. Again, the tool is capable of filtering based on the duration for which heart rate data were obtained per day, and Figure 5 shows the data can be exported for further study.

4.4. Correlation Analysis between Daily PA and Learning Performance

Using the data collected, we were able to analyze the correlations between the learning outcome of the 36 subjects and various PA-related factors. PA data were filtered according to the criterion that the device was worn for over 20 hours per day. Learning outcome was assessed based on five examination results from every subject during the investigation period. Association analysis was conducted using SPSS after combining examination results and PA data from the day prior to the exam. Table 2 shows the analysis results, which indicated a positive correlation between sleep duration on the day prior to the examination and examination outcome, with a correlation coefficient of 0.365 (p<0.05).

Table 2: Correlations between PA factors and examination outcome.

Factor (One day Before test)	*N*	*Correlation*	*P-value*
Calories	36	-.103	.549
Steps	36	.091	.598
Distance	36	.086	.619
Floors	36	-.055	.751
Elevation	36	-.055	.751
Minutes Lightly Active	36	-.144	.402
Minutes Fairly Active	36	.120	.485
Minutes Very Active	36	.168	.327
Minutes Asleep at Night	36	$.365^{*}$	.029
Minutes Asleep in Whole Day	36	.056	.746
Awakenings Count	36	.261	.124
Time In Bed	36	$.365^{*}$	.029
Sleep Efficiency	36	.008	.962
Start Time	36	-.220	.198

5. Conclusion

Data including 19 factors were collected from 30 students and stored in the Impala data warehouse in this study. A BVS was designed to represent data using charts and calculate descriptive statistics. Moreover, the BVS also enabled filtering of data based on the number of hours during which the device collected heart rate data, and allowed export of data as a CSV file for advanced analysis such as data mining. Correlation analysis was performed between the data collected during the one day prior to examination in one subject and the examination outcome. Despite the small sample size (n=36), the results of the correlation analysis showed that sleep duration on the one day prior to the examination had a positive effect on examination performance (as measured by the examination score). In this study, a brand new method for behavior analysis was introduced, which differed from questionnaire-based analysis or short-term experiments. A wearable tracking device was used for 24-hour daily physical activity measurement in students, and analysis of the association between various activity-related factors and examination performance was conducted. Moreover, a Behavior Visualization System was designed to collect physical activity and lifestyle data such as sleep-related information from the cloud and enable statistical analysis and illustration via charts, in order that researchers are able to analyze the physical activity, behavior and lifestyle of subjects for related research.

Acknowledgments

This study was supported by the Ministry of Science and Technology, MOST 102-2221-E-155-027 and MOST104-3115-E-155-002.

6. References

[1] J. Vanhelst, et al., Physical Activity Is Associated with Attention Capacity in Adolescents. *J Pediatr,* 2016. **168**: p. 126-31.e2.

[2] F. De Crescenzo, et al., The use of actigraphy in the monitoring of sleep and activity in ADHD: A meta-analysis. *Sleep Med Rev*, 2016. **26**: p. 9-20.

[3] M. Kishida and S. Elavsky, An intensive longitudinal examination of daily physical activity and sleep in midlife women. *Sleep Health*, 2016. **2**(1): p. 42-48.

[4] D. R. Bassett, Jr., et al., Pedometer-measured physical activity and health behaviors in U.S. adults. *Med Sci Sports Exerc*, 2010. **42**(10): p. 1819-25.

[5] G. Curcio, M. Ferrara, and L. De Gennaro, Sleep loss, learning capacity and academic performance. *Sleep Med Rev*, 2006. **10**(5): p. 323-37.

[6] C. Lang, et al., The relationship between physical activity and sleep from mid adolescence to early adulthood. A systematic review of methodological approaches and meta-analysis. *Sleep Med Rev,* 2015. **28**: p. 28-41.

[7] Suzana Loshkovska, Saso Koceski. ICT Innovations 2015: Emerging Technologies for Better Living. *Springer*, 2015. p.5 - 8.

[8] Y.-J. Tsai, C.-L. Hung, C.-L. Tsai, Y.-K. Chang, C.-J. Huang, and T.-M. Hung, "The relationship between physical fitness and inhibitory ability in children with attention deficit hyperactivity disorder: An event-related potential study," *Psychology of Sport and Exercise*, 2016.

[9] E. Iuliano, A. di Cagno, G. Aquino, G. Fiorilli, P. Mignogna, G. Calcagno, *et al.*, "Effects of different types of physical activity on the cognitive functions and attention in older people: A randomized controlled study," *Exp Gerontol,* vol. 70, pp. 105-10, Oct 2015.

[10] M. Quante, E. R. Kaplan, M. Rueschman, M. Cailler, O. M. Buxton, and S. Redline, "Practical considerations in using accelerometers to assess physical activity, sedentary behavior, and sleep," *Sleep Health,* vol. 1, pp. 275-284, 2015.

[11] N. Cellini, E. A. McDevitt, S. C. Mednick, and M. P. Buman, "Free-living cross-comparison of two wearable

monitors for sleep and physical activity in healthy young adults," Physiol Behav, vol. 157, pp. 79-86, Apr 1 2016.

Automatic Summarization System of Lecture Slides

Sébastien ANDRÉ, Atsushi SHIMADA, Hiroaki OGATA

Kyushu University

Abstract In this paper, we propose a new tool to enhance people learning. This tool is a Web Application that permits to summarize lecture slides. The proposed system provides a shorter version of an original lecture slide with keeping important contents as much as possible.

Keyword Lecture slide, material, web application, summarization.

1. Introduction

An automatic summarization methodology, which creates a summary of a given set of lecture materials[1], was proposed and has been test on more than 300 students by doing quiz tests at the beginning of each course. Results show that students who follow summarized lecture slides have better score than these who preview the full material.

In this paper, we are motivated by our previous work to develop a web application which implement this method.

2. Technical System Architecture

Our web application is composed of two main module, frontend and backend. Hereafter we will use the term “frontend” to talk about our website and the term “backend” that refers server(*figure 1*).

Frontend responsibility is to offer a user friendly website that allow uploading lecture material (ppt, pptx or pdf) and downloading shorter lecture slide into a single pdf file.

Backend will operate every processing treatment and manage our database. Both sides will communicate using http protocol in JSON format. The backend was developed in Java programming language that offer a powerful and versatile environment to develop our tool. We use Spring Boot Framework developed by Pivotal Company.

Regarding the frontend in order to make our website dynamic we will use Angular2 language from Google. Angular2 is a framework provided for Typescript language that is a superset of JavaScript. Beyond we can see the overview of our system.

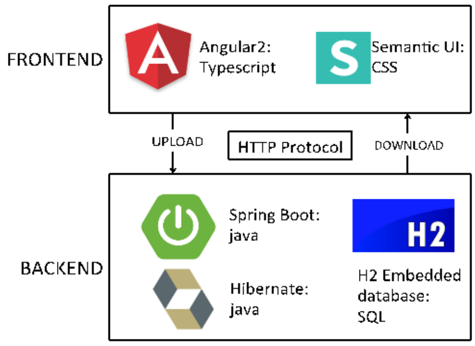


Figure 1. Technical Architecture

3. Backend

Backend responsibility is doing all business logic of our application. It has to operate various processing algorithm. For the details of the algorithm, refer to our previous work[1]. Our system exploits slides content such as text, image and time (*figure 2*). We will combine this information with mathematical formulas and provide a score for each slide.

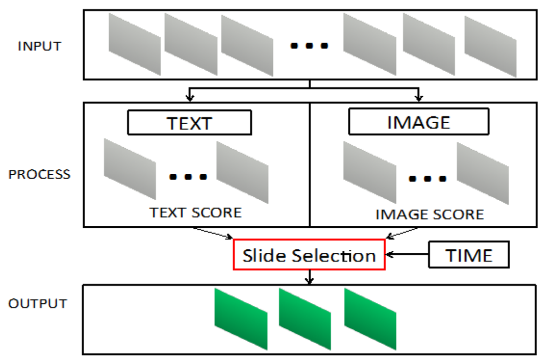


Figure 2. Functional Architecture

The functional architecture above gives an overview of what our system realizes. As shown above, our system is based around three main modules: text processing, image processing, time and slide selection.

Our backend API should behave different according to lecture material upload. For example: file uploaded can be pdf, ppt or pptx; lecture material could be in Japanese or English. To avoid getting algorithm structure like a tree with many branches we will use some factory design pattern. Thus our system has a unique algorithm process for every kind of material.

4. Frontend

The frontend is developed with Angular2 framework(*figure 3*); this allow us to build our website around component. Every component is independent and handle a small part the website, component can be seem as feature. Let's see which component are composing our interface:

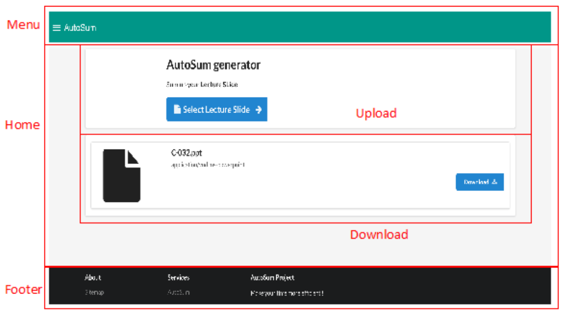


Figure 3: Website

5. Example

In this section we will give an example of our application with a real Japanese lecture slides. Originally this course was 30 slides but we will summarize only 16 slides. For this example, we will choose a maximum time length of 8 minutes. In this example each slide has 1-minute browsing time set up, so our system will select 8 slides. Let's see the input and output below in figure 4 and 5.

Figure 4: Lecture slides uploaded

Figure 5: Result summarization

6. Conclusion

In future work, we will optimize the process. Image subtraction is not fully accurate in our case, so we will try other image subtraction techniques. The process could last 2 minutes for 30 slides in Japanese, we can reduce the time by using multiple thread. Each part of processing (text, image and time) could be allocated to their own thread.

7. Reference

[1] Atsushi Shimada, Fumiya Okubo, Chengjiu Yin, Hiroaki Ogata, *"Automatic Summarization of Lecture Slides for Enhanced Student Preview"*, The 23rd International Conference on Computers in Education (ICCE2015), pp.218-227, 2015.

Data Visualization Framework for Learning Analytics

Ryo ABO[*1], Takayuki KOGA[*1], Izumi HORIKOSHI[*2],
Kimiaki YAMAZAKI[*3], Yasuhisa TAMURA[*1]
[*1] Graduate School of Science and Technology, Sophia University
[*2] Graduate School of Informatics, University of Educational Systems
[*3] TEK Software Co. Ltd.
ytamura@sophia.ac.jp

Abstract

The authors propose a top-down data visualization framework for learning analytic activities on K-12 schools. In this framework, 4 types of entities are set: (1) data generator(s), (2) learning contents, (3) time, and (4) viewers of visualization. Among them, (1), (2) and (3) have some scales; e.g. (1) data generator might mean a learner himself, another learner, a group, a class, same grade learners in a school, and all learners in a school. Since the authors set 6 kinds in (1), also 7 in (2), 5 in (3) and 4 in (4), total 840 (6 x 7 x 5 x 4) patterns of visualization are possible. Among them, some are unnecessary for viewers, others might be harmful for viewers. The authors studied the usefulness (uselessness) and harmfulness on each combination, in order to arrange possible and useful combination of (1), (2), (3) and (4).

1. Introduction

In recent years, learning analytics (LA) researches have been focused in educational technology field [1]. LA includes learning activity history data collection, storage, analysis, evaluation and feedback for learners, teachers and other stakeholders.

Because data volume collected by the LA is huge, it is difficult to recognize the tendency from raw data. Therefore, it is useful and convenient to visualize the data in the form of such a graph. There are many preceding proposals for this visualization example. Among them, there are some preceding researches to show overall framework of how and what to visualize. Ferguson [2] showed 5 categories of LA visualization for collaborative learning; (1) Social Learning network analytics, (2) Social learning discourse analytics, (3) Social learning content analytics, (4) Social learning disposition analytics, and (5) Social learning context analytics. It also proposes a tool called SocialLearn Dashboard to visualize LA data in Open University UK. Dyckhoff [3] proposed a LA support tool kit called eLat for teachers. Duval [4] took an example of visualization for jogging history and said: “At the first time they see a dashboard, they are impressed. However, in order to overcome this stage and continue their activity, the dashboard should clarify the goal and visualize how far they achieved toward the goal”. Also he showed 3 types of dashboards named Wakoopa, CAM, and Student Activity Monitor (SAM). Verbert [5] proposed a visualization tool named StepUp!, which consisted 4 stages of Awareness, Reflection, Sense making, and Impact in visualization.

In this paper, the authors propose a LA data visualization framework in K-12 learning activities. In this framework, 4 types entities are set: (1) data generator(s), (2) learning contents, (3) time, and (4) viewer of visualization. Among them, (1), (2) and (3) have some scales; e.g. (1) data generator might mean a learner himself, another learner, a group, a class, same grade learners in a school, and all learners in a school. Figure 1 is a 3-dimentional graph to show scales of (1), (2) and (3). Because these entities are independent, it can be represented in a 3-dimentional graph. One graph to visualize certain LA result is located in a cross point.

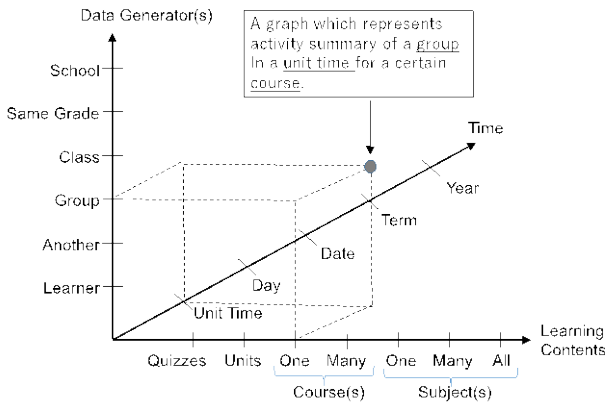


Figure 1: 3-dimentional data granularity

On the other hand, (4) viewer of visualization are classified into 4 groups: learners, guardians, teachers, and managers. Principals and board members of education are included in the “managers” group. Requirements from these stakeholders are totally different. For example, a learner wants to know the position in a class, on the other hand managers like principals want to know overall tendency of his school or teacher dependent parameters to use School Assessment.

2. Proposing visualizing framework

In order to arrange these requirements, the authors assumed the main needs of the four types of stakeholders.

A learner, the first type of stakeholders, wants to understand his own situation in his group or the class. Also, it is useful to inform long-term situation (Term or Year) to encourage his learning activity. However, it is harmful to inform him another learner’s situation, because he will start to compare his achievement with others. So this type of information should be hidden. Also parents, want to know that their children learn continuously. They focus on course-grained views for learning contents, e.g. they tend course and subject level rather than quiz and unit level.

Teachers want to grasp situations of his class to more detailed level, but not whole classes and a school. However, managers focus on a school and classes for relative evaluation and school assessment. Viewers of visualization have such a different focus. Based on these needs and focuses, the authors assumed possible, unnecessary, and prohibited visualization as shown in Table 1. Sub-table (a)-(d) shows different viewers of visualization. For each table, X axis indicates options of learning contents, Y axis time options, and cell contents show options of data generators. Cell contents of “-“ means that this visualization is unnecessary, while “X” means to be prohibited. Main reasons of “X” are privacy protection. It also prevents overheated competitions between learners.

3. Conclusion, future works

In order to show top-down mannered LA visualization framework in K-12 schools, the authors introduced 4 axes to arrange possible combination of data: (1) data generator(s), (2) learning contents, (3) time, and (4) viewer of visualization. Also, necessary judgments of these points are proposed.

Their axes and granularity in Figure 1 and Table 1 were discussed in the laboratory, and not yet been discussed with real stakeholders. Therefore, future discussions and polish-up activities are necessary for more precise and useful framework details.

In the visualization of LA, combination of various graphs has been proposed as a Dashboard. However, the proposal of this paper dealt with a single graph, further consideration is needed for correctness and effectiveness of the Dashboard data combinations.

Table 1: Possible, unnecessary, and prohibited visualization of Learning Analytics

(a) Learners

	Quizzes	Units	A course	Courses	A subject	Subjects	All subjects
Unit time							
Day			LXGCS-	LX----	LXGCS-	LX----	
Date			LXGCS-	LX----	LXGCS-	LX----	
Term	LXGCS-	LXGCS-		LXGCS-		LXGCS-	
Year	LXGCS-	LXGCS-		LXGCS-		LXGCS-	

(b) Guardians

	Quizzes	Units	A course	Courses	A subject	Subjects	All subjects
Unit time							
Day			LXGCS-	LX----	LXGCS-	LX----	
Date			LXGCS-	LX----	LXGCS-	LX----	
Term				LXGCS-		LXGCS-	
Year				LXGCS-		LXGCS-	

(c) Teachers

	Quizzes	Units	A course	Courses	A subject	Subjects	All subjects
Unit time							
Day			-AGCS-	-A----	-AGCS-	-A----	
Date			-AGCS-	-A----	-AGCS-	-A----	
Term	-AGCS-	-AGCS-	-AGCS-		-AGCS-		
Year	-AGCS-	-AGCS-	-AGCS-		-AGCS-		

(d) Managers

	Quizzes	Units	A course	Courses	A subject	Subjects	All subjects
Unit time							
Day				----SS		----SS	
Date				----SS		----SS	
Term			----SS		----SS		----SS
Year			----SS		----SS		----SS

6 digits of each cells means: L (Learner), A (Another learner), G (Group), C (Class), S (Same grade), and S (School). X: Prohibited, -: Unnecessary

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 26282059.

References

[1] Ferguson, R.: “Learning analytics: drivers, developments and challenges”, International Journal of Technology Enhanced Learning, Vol. 4, No.5-6, pp.304-317 (2012).

[2] Ferguson, R., & Shum, S. B.: “Social learning analytics: five approaches”, Proc. 2nd international conference on learning analytics and knowledge, pp. 23-33 (2012).

[3] Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., & Schroeder, U.: “Design and Implementation of a Learning Analytics Toolkit for Teachers”, Educational Technology & Society, Vol.15, No.3, pp.58-76 (2012).

[4] Duval, E.: “Attention please!: learning analytics for visualization and recommendation”, Proc. 1st International Conference on Learning Analytics and Knowledge, pp. 9-17 (2011).

[5] Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L.: “Learning analytics dashboard applications”, American Behavioral Scientist, 0002764213479363 (2013).

Academic Performance Estimation by Visualization of Learning Patterns in Campus Environments.

Cheng-Wei Lee, Huan-Yi Pan, Liang-Chih Yu*, Chien-Lung Chan, Chih-Yueh Chou, and Shu-Fen Tseng
Yuan Ze University, Chung-Li, Taiwan
*lcyu@saturn.yzu.edu.tw

Abstract

Educational data mining provides a useful way to discovering hidden knowledge from the educational context. Mining educational data helps to clearly comprehend students' learning behaviors, estimate students' academic performance, identify weak students, decrease the drop-outs rate and adjust teaching strategies in a timely manner. The main purpose of this paper is to build a model to illustrate the academic performance of students by visualization of learning patterns in campus environments. In this research, the classification is implemented by Support Vector Regression (SVR). Information's like Academic Activities, Club Participation, Virtual Classroom Interaction Activities and Library Usage were collected from student portal system, to predict the performance, provide real time estimation and the most important, identifying students at risk to avoid drop-out.

1. Introduction

Academic performance estimation is important not only for students themselves but also for educators. Following the trend of big data and combined with advances in computation, educational data mining (EDM) becoming a growing research community that gained increasing popularity. International Educational Data Mining Society (http://www.educationaldatamining.org/) given the definition of educational data mining as follows: *Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique and increasingly large-scale data that come from educational settings, and using those methods to better understand students, and the settings which they learn in.*

There are many useful algorithms and methods, such like Clustering, Classification, Decision Trees, Association Rule Mining, Regression, Visualization, and many others, are used for exploring hidden knowledge from the educational context. Mandinach & Gummer (2013) noted that data-driven decision making has got a rich focus in education. All educators must understand how to use tangible evidence to inform their decisions, rather than use anecdotes, intuitions, or personal preferences. EDM could be applied to various research domains, including individual learning from educational software, computer supported collaborative learning, computer-adaptive testing, and the factors that are associated with student failure or non-retention in courses (Baker and Yacef, 2009).

In order to help people to highlight useful information and support decision making, graphic techniques were used to achieve information visualization. In educational environment, visual representations of course activities and usage records can help educators and students to get a general view of learning status (Remero, 2010).

The main purpose of this paper is to build a model to illustrate the academic performance of students by visualization of learning patterns in campus environments. In this research, the classification is implemented by Support Vector Regression (SVR). Information's like Academic Activities, Club Participation, Virtual Classroom Interaction Activities and Library Usage were collected from student portal system, to predict the performance, provide real time estimation and the most important, identifying students at risk to avoid drop-out.

2. Related works

Strecht et al.(2015) conducted a study to predict students' performance in courses and find out the causing factors related to success and failure. To facilitate this, classification and regression algorithms are used to analyze administrate data from 700 courses. The findings showed that classification yield good results in prediction of pass or fail in a course. Kabakchieva (2013) focused on student performance prediction based on students' personal, pre-university and university performance characteristics data. Various data mining algorithms of classification were used to achieve this goal. The findings may not remarkable (the prediction rate between 52-67%), but revealing the high potential of data mining applications for educational management. Remero et al. (2008) performed a study to compare different data mining techniques for extracting discovery-driven information from computer-based and web-based educational systems. Useful data mining techniques, such like classification, clustering, visualization, and association rule mining, could be applied to discover interesting patterns and tendencies in student's usage information. Baradwaj and Pal (2011) applied a decision tree model to extract knowledge that describes students' performance in end semester examination. The main objective is to identify dropouts earlier and help students and teachers to make adjustment. Huang et al. (2016) designed a research to discover students' learning performance and help educators to monitor students' learning behavior across different time-periods by using association rules across different semesters from 2009 to 2011.

Collectively, these studies show that EDM provides a useful way to discovering hidden knowledge from the educational context. Researchers used data mining techniques to extract knowledge from various data sets, including students' characteristics, academic activities, learning histories, exam scores, and grades. In order to improve learning effectiveness in more convenient way, our work provides real time estimation by visualization interface.

3. Data preparations and Research methods

The data set used in this study obtained 32 learning patterns from two groups of students enrolled on 2014 and 2015. There are 988 students belongs to 2015 and 862 students belongs to 2014. Those 32 learning patterns represent information about a student's characteristics or state, including Academic Activities, Club Participation, Virtual Classroom Interaction Activities and Library Usage. Descriptions of 32 learning patterns and their types are given in Table 1.

Table 1. Learning Patterns.

Category	Description	Types
Book Loans	amount of book loans, PR[1]	counts, %
E-Library	login record of My Library, PR	counts, %
Virtual Classroom	login record of Virtual Classroom, PR	counts, %
	click record of Grade, PR	counts, %
	click record of Information, PR	counts, %
	click record of Activity, PR	counts, %
	click record of News, PR	counts, %
	click record of Discussion, PR	counts, %
	click record of Learn Material, PR	counts, %
	click record of Homework, PR	counts, %
Club Participation	amount of Club Participation, PR	counts, %
Academic Activity	amount of Academic Activity, PR	counts, %
Library	amount of Library usage	counts
Grade	amount of Failed subject	counts
College Entrance Examination Score	the score of Chinese	grade
	the score of English	grade
	the score of Mathematics	grade
	the score of Science	grade
	the score of Social Studies	grade
	total score of College Entrance Examination	grade

We standardized part of our data, from raw counts to rates, to create percentage ratio (PR). It is helpful to make some learning patterns in two types. Both of raw counts and PR were used in analysis to enrich the predictive capability of our models. Besides, the PR value was integrated in visualization way, as shown in Fig. 1, to remind students their learning status.

Support Vector Regression (SVR) is applied in this study for the purpose of prediction of student performance. We used the trained regression model to calculate the drop-out ratio[2].

1 PR: Percentage Ratio.

2 We here define drop-out as the student grade under 70.

Figure 1. Dashboard in students' portal.

4. Experimental results

This section presents the results of our experiment.

4.1. Experimental results using all features

We used all features in our experiments to examine whether the student's grades would higher than 70 or not. The results listed in Table 2 and Table 3.

Table 2. Predict non-dangerous group using all features.

Non-dangerous group (grades>70)			
	Precision	Recall	F1-score
2014	0.87	0.94	0.9
2015	0.88	0.93	0.91

Table 3. Predict dangerous group using all features.

Dangerous group (grades<70)			
	Precision	Recall	F1-score
2014	0.79	0.58	0.66
2015	0.68	0.56	0.6

To compare the experimental results in Table 2 and Table 3, it shows that the trained regression model had better predictive ability in non-dangerous group. The average of Recall in 0.94 means we can correctly recognize students who get grade higher than 70.

4.2. Experimental results by feature selection

We used feature selection to get improvement. There are three different kinds of features in our data set: Grade, Counts, and PR. We tried to remove PR and Counts separately to compare the results. Table 4 and Table 5 are listed below to show results of two types of feature combination by 2015 data set (988 students). Experimental results demonstrated that remove PR (Type 1) or remove Counts (Type 2) could slightly improve our model in drop-out prediction.

Table 4. Non-dangerous group by features selection.

Non-dangerous group (grades>70)			
	Precision	Recall	F1-score
ALL	0.88	0.93	0.91
Type1	0.89	0.94	0.91
Type2	0.89	0.94	0.91

*ALL: Grade + Counts + PR(%); Type1: Grade + Counts; Type2: Grade + PR(%)

Table 5. Dangerous group by features selection.

Dangerous group (grades<70)			
	Precision	Recall	F1-score
ALL	0.68	0.56	0.6
Type1	0.74	0.55	0.62
Type2	0.74	0.57	0.62

*ALL: Grade + Counts + PR(%); Type1: Grade + Counts; Type2: Grade + PR(%)

5. Conclusions

In this paper, we are dealing with drop-out prediction to provide real time estimation of student learning performance in visualization way. Support Vector Regression (SVR) is applied to train the predictive model. Information's like Academic Activities, Club Participation, Virtual Classroom Interaction Activities and Library Usage were collected from student portal system, to investigate the effectiveness of our predictive model.

The experimental results demonstrated that our model have good predictive abilities of non-dangerous group in Recall(0.93), Precision(0.88), and F1-score(0.91). In order to achieve our goals, we deploy our model into students' portal system to predict the student performance, provide real time estimation and the most important, identifying students at risk to avoid drop-out.

6. References

[1] R.S.J.D Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions.", *Journal of Educational Data Mining*, 2009, 1(1), pp. 3-17.

[2] D. Kabakchieva, "Predicting Student Performance by Using Data Mining Methods for Classication.", *Cybernetics and Information Technologies*, 2013, 13(1), pp. 61-72,.

[3] P. Strecht, L. Cruz, C. Soares, J. Mendes-Moreira, and R. Abreu, "A Comparative Study of Regression and Classification Algorithms for Modelling Students' Academic Performance", *Proceedings of the 8th International Conference on Educational Data Mining*, 2015, pp. 392-395.

[4] C. Romeo, S. Ventura, and E. Garcia, “Data mining in course management systems: Moodle case study and tutorial”, *Computers & Education*, 2008, 51, pp. 368–384.

[5] B.K. Baradwaj and S. Pal, "Mining Educational Data to Analyze Students‟ Performance.", *International Journal of Advanced Computer Science and Applications*, 2011, 2(6), pp. 63-69

[6] T.C.K. Huang, C.H. Huang, and Y.T. Chuang, "Change discovery of learning performance in dynamic educational environments.", *Telematics and Informatics*, 2016, 33, pp. 773–792.

[7] E. Mandinach and E. Gummer, “A Systemic View of Improving Data Literacy into Educator Preparation.”, *Educational Research*, 2013,42(1), pp. 30-37.

[8] C. Romeo, “Educational data mining: a review of the state of the art”, *IEEE Transaction on Systems, Man, and Cybernetics*, 2010, 40(6), pp.601-618.

Measuring Engagement: Student Profiling and the Effects of Remedial Learning Counseling

Shu-Fen Tseng*, Liang-Chih Yu, Ho-Shing Wu, Po-Yao Chao
Yuan Ze University, Taiwan
*gssftseng@saturn.yzu.edu.tw

Abstract

Various data sources were collected from students' engagement activities, including logs of virtual classroom usage, library records, clubs, and participation of on-campus academic activities to predict students' learning outcome. Regression results suggested that logging in virtual classroom, going to library, and participating on-campus academic activity have the major positive prediction powers of learning performance. A total of five types of student engagement could be categorized by cluster analysis among first-year university students in this study. These types included: club only, learning engaged, apathetic, library specific, club and engaged groups. Engagement styles with those who participated more on learning activities and those who joined both club and learning activities received a better learning outcome. The group who showed apathetic attitude toward school activities performed poorly in their final grades. Inadequate time management and learning strategies were two main reasons for students' failing of their courses. Remedial interventions by teaching advisors showed a great impact on students' improvement of final grades. Better improvements could be found among students in the library specific and learning engaged groups. The least progress between midterm and final exams were students who belonged to the apathetic style of engagement. .

1. Introduction

Learning analytics is "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" [1] . Students leave their digital footprints whenever they interact with their universities and campus environment, such as going to the library, logging into the virtual classroom or submitting assignments online. Using the increased availability of big datasets around learner activities and digital footprints left by student activities in learning environments, learning analytics provide the tools and perspectives to improve learning and teaching in the pedagogical practices. Using educational data mining techniques, learning analytics can rely on the information culled from various data sources to determine the status of academic performance, identify possible risk factors, recognize intervention point, and predict possible outcomes [2].

Student engagement has been defined as "participating in educationally effective practices, both inside and outside the classroom, which leads to a range of measurable outcomes" [3]. The majority of existing literature on student engagement is concerned with improving students' learning outcome [4]. Rooted in constructivist approach, student engagement advocates stated that learning is influenced by how an individual participates in educationally purposeful activities [5]. Therefore, identify students' involvement with what activities and conditions that generate high quality learning become an important issue in learning institutions.

In this study, we collected multiple sources of students' engagement data from a private university in Taiwan to profile students' involvement of activities and to analyze learning engagement that highlights the keys to students' performance. A visualized tool of performance estimation is developed to signal students' learning outcomes. In this private university, students at high risks of failing courses were identified soon after their midterm exams, interview interventions by teaching advisors were then outreached to these students. We further examined the impacts of these interventions by calculating Z-score differences between midterm and final exams among the first year of students.

2. Methods

2.1. Regression Analysis

Data of engagement activities among 1,742 freshmen in a private university were collected, including logs of virtual classroom usage, library records, clubs, and participation of on-campus intellectual activities. In addition, students learning outcomes (i.e. grades) and counseling records were also collected to measure the effects of students' engagement. Regression analyses were employed to examine the extent of students' engagement on learning outcomes. A total of 119 out of 1,742 freshmen were reported to have received teachers' counseling interviews in this study. The effect of this remedial intervention was measured by calculating the differences of students' midterm and final grades.

2.2 Cluster Analysis

The Ward's hierarchical clustering method was employed in this study to determine number of engagement styles of students. The Ward's minimum variance clustering is useful for exploratory work when researchers do not have a preconceived number of clusters in the dataset. Main indicators of students' engagement were records of students' usage in virtual classroom, library related activities, club participation, and on-campus intellectual activities.

3. Results

3.1. Regression and Visualization

The extent of engagement activities of students can be used to predict students' learning outcome. Regression results suggested that logging in virtual classroom, going to library, and participating on-campus academic activity had the major positive predictive powers of students' final grades (table 1). Interestingly, students who checked homework and news more often on the virtual classroom, their final grades were lower in compared with others. In addition, those who borrowed more books from library also reported to have lower grades in this study.

Table 1: Regression Results of Learning Outcome by Engagement Activities

	Final grade	
	B	βeta
Constant	67.27	
Academic activity	0.39	0.12**
Club	0.70	0.05**
Library entrance #	0.03	0.14**
Library book #	-0.06	-0.07**
Library E-book #	0.12	0.05*
Virtual Classroom (VC) login	0.01	0.26**
VC Grade	0.01	0.07*
VC Class information	0.05	0.05
VC Activity	0.02	0.01
VC News	-0.03	-0.08**
VC Discussion	0.00	-0.01
VC Course material	0.02	0.08**
VC Homework	-0.02	-0.14**
R^2	.16	
N	1,742	

$$\begin{aligned}\widehat{\text{Final grade}} = &\underset{(125.16**)}{67.27} + \underset{(4.82**)}{0.39}\text{Intellectual} + \underset{(2.40**)}{0.70}\text{Club} + \underset{(5.85**)}{0.03}\,Entrance\ number \\ &- \underset{(-2.66**)}{0.06}\,Book\ number + \underset{(2.05*)}{0.12}\,Ebook\ number + \underset{(5.15**)}{0.01}\,Login + \underset{(1.99)}{0.01}\text{Grade} \\ &+ \underset{(1.50)}{0.05}\,Updates + \underset{(0.45)}{0.02}\,Activity - \underset{(-2.57**)}{0.03}\,News - \underset{(-0.18)}{0.01}\,Discussion \\ &+ \underset{(2.37**)}{0.02}\,Course\ material - \underset{(-4.11**)}{0.02}\,Homework\end{aligned}$$

This regression equation can be used to predict students' final grade of study. An easy to read signal system by employing regression equation was developed to help students' understand the impact of his/her activity engagement on learning outcome. Red color indicates student at higher risk of low grades, and green color means a better grade is foresighted.

(visualized system)

3.2. Student Profiling

By cluster analysis, we profiled freshmen students into five types of engagement styles (table 2):

(1)Club only group: a total number of 713 or 40% of 1,742 freshmen were grouped to this group. This group showed high proportion of club participation, but low in both library and virtual classroom usages, and inactive participation of on-campus academic activities.

(2)Learning engaged group: 53 students or 3% of freshmen were categorized in this group. This group scored the highest in participating on-campus academic activities, highest in using the library, most frequently used various functions in the virtual classroom, and moderately participated in club activities.

(3) Apathetic group: there were 497 students or 28.5% of freshmen were grouped in this category. This group scored the lowest in all kinds of activity records.

(4)Library specific group: 294 students or a proportion of 16.9% of freshmen was categorized in this group. Students in this group showed the highest number of library entrance and usage of personal e-book site. They actively participated in club activities and scored the second highest in participating on-campus academic activities. Students in this group showed a modest rate in the usage of virtual classroom.

(5)Club and engaged group: there were 185 students or 10.6% of freshmen belonged to this category. Students of this group actively participated in club activities. They used the library and joined on-campus academic activities in a moderate fashion. This group scored the second highest in the virtual classroom usage.

Table 2: Cluster Analysis by 4 Different Engagement Activities

Clusters		1	2	3	4	5
Academic Activity		1.6	3	0.9	2.5	2.2
Club Participation		0.7	0.6	0.6	0.7	0.7
Library Usage	Entrance #	40.8	45.9	20.2	48.6	35.6
	Book #	4.4	6.6	2.2	6.6	5.2
	E-book #	1.6	2.7	0.7	3	2.2
Virtual Classroom Usage	Login	334	982.4	172.7	493.2	656.6
	Grade	96.8	330.4	47.5	132.1	193.5
	Information	9.9	20.8	6.1	12	16.4
	Activity	9.5	18	6.3	11.1	14.2
	News	26.1	69.3	12.9	36.8	53.9
	Discussion	11.7	30.3	6.1	14.2	21.6
	Course material	49.8	113.7	23.3	67.9	100.3
	Homework	66.2	209.4	31.3	105.8	145.3

1st: Club only group (40%)
2nd: Learning engaged group (3%)
3rd: Apathetic group (28.5%)
4th: Library specific group (16.9%)
5th: Club & engaged group (10.6%)

3.3 Engagement Styles and Learning performance

Learning performance was measured by three elements: credit hours at risk by midterm exam, credit hours failed by final exam, and average final grades. Students with learning engaged (2nd cluster), and club and engaged (5th cluster) styles did show a better learning outcome than other groups of students. Their risks of flunking course were lower in midterm and final exams, and their final grades were higher than the others, at the average score of 79.9 and 78.2 respectively. Followed by the library specific (4th cluster) group with average final grade at 77.2. The apathetic group (cluster 3) of students showed the poorest performance in all three indictors. They received more risk warnings (average 4.2 credit hours) in the midterm exam, flunked more credit hours (average of 3.5) in final exam, and earned the lowest average final score (69.9) in compared to the other groups.

Table 3: Engagement Styles and Learning Performance

Cluster groups	Credit hours at risk by midterm	Credit hours failed by final	Final grade
Club only	2.7	1.7	75.4
Learning engaged	2.2	0.9	79.9
Apathetic	4.2	3.5	69.9
Library specific	2.6	1.3	77.2
Club & engaged	2.0	1.0	78.2

3.4 Remedial Learning Counseling

There were 1,742 freshmen in our study, 119 (12.2%) of them reported to have received remedial counseling from their teaching advisors. For those 119 students who received counseling interviews by teaching advisors, 63 (52.9%) of them reported inadequate time management was the most important factor for their poor performance of learning. A 45.5% of them indicated inadequate learning strategies (planning, reading skill, exam pressure management etc.) and 37% of them reported course factors (difficulty, comprehension problem, lack of pre-requite knowledge, teaching styles etc.) were the causes of their poor grades. Total of twenty nine students (24.4%) indicated their poor performance were due to personal factors (emotion, health, motivation, family problems etc.), while 19.3% reported English-taught courses (language difficulty) was one of the main reasons for their poor performance of final scores.

Table 4: Reasons for Poor Performance among High-risk Students

N=119 %	Personal	Time	Strategy	Course	English
Club only n=30	16.7	56.7	50.0	46.7	13.3
Learning engaged n=3	33.3	33.3	0.0	0.0	0.0
Apathetic n=65	26.2	47.7	46.2	35.4	27.7
Library specific n=14	35.7	64.3	50.0	28.6	0.0
Club & engaged n=7	14.3	71.4	28.6	42.9	14.3

Among different styles of student engagement, a total of 13.1% of students in apathetic group received remedial intervention, followed by learning engaged (5.7%), library specific (4.8%), club only (4.2%), and club and engaged (3.8%) groups. We further examined the impacts of this interview intervention by measuring student performances between midterm (standardized average credit hours at risk) and final exams (standardized average final score). Positive values meant students have improved their learning outcomes after midterm, while negative values indicated students have degenerated after midterm exam. The results suggested that remedial intervention by teaching advisors did show great impacts on students' improvement of grades. Particularly, huge improvements could be found among students in the library specific (4th cluster) and the learning engaged (2nd cluster) groups, at the improvement rates of 1.0 and 0.7 respectively. The least progress was found among students in the apathetic group (3rd cluster) with the improvement rate at 0.2.

Table 5: Engaged Styles and Remedial Counseling Effects

	Learning Counseling	Z-score of credit hours at risk by midterm (M)	Final grade Z-score (F)	Effects (F-M)
Club only n=713	No n=683	0.2	0.2	0.0
	Yes n=30	-2.0	-1.5	0.5
Learning engaged n=53	No n=50	0.4	0.6	0.2
	Yes n=3	-2.0	-1.3	0.7
Apathetic n=497	No n=432	0.0	-0.2	-0.2
	Yes n=65	-2.4	-2.2	0.2
Library specific n=294	No n=280	0.2	0.3	0.1
	Yes n=14	-2.2	-1.2	1.0
Club & engaged n=185	No n=178	0.4	0.4	0.0
	Yes n=7	-1.7	-1.1	0.6

4. Conclusion

In this study, we collected students' activity records to profile students' engagement among freshmen in a private university. A total of five types of student engagement could be identified by cluster analysis. They were club only, learning engaged, apathetic, library specific, club and engaged groups. Regression results suggested that logging in virtual classroom, going to library, and participating on-campus academic activity were the major predictors of students' learning outcomes. This regression equation can be used for predicting students' final grade of study. In addition, an easy to read signal system by this regression equation was developed.

In terms of the engagement styles and learning outcome, we found that those who engaged more on learning activities (2nd cluster) and those who joined both club and learning activities (5th cluster) received a better final grades, in contrast to those who show apathetic attitude toward school activities performed poorly in their studying. Inadequate time management and learning strategies were two main factors for students to fail their courses. Remedial interventions by teaching advisors showed a great impact on students' improvement of final grades. Particularly, better improvements were found among students in the library specific (4th cluster) and learning engaged (2nd cluster) groups, in compared to the least progress among students in the apathetic group (3rd cluster). Our findings not only delineated different styles of student engagements among the first year of college students, but also supported previous student engagement literature that better learning outcome can be found among those who actively engaged in intellectual, learning and club activities.

5. Acknowledgments

This study was supported by the Ministry of Education, Taiwan, IR Project and the Innovation Center for Big Data and Digital Convergence, Aim for the Top University Projects.

6. References

[1] Jones, S.J. (2012) " Technology review: The possibilities of learning analytics to improve learner-centered decision-making", Community College Enterprise, 18(1), 89-92.

[2] Mattingly, K.D., Rice, M.C., and Berge, Z. L. (2012) "Learning analytics as a tool for closing the assessment loop in higher education". Knowledge Management & E-Learning: An International Journal, 4(3), 236-247.

[3]Kuh, G.D., Kinzie, J., Buckely, J.A, Bridges, B. K. and Hayek, J.C. (2007) "Piecing together the student success puzzle: Research, propostions, and recommendations". ASHE Higher Education Report, V. 32, No. 5 San Francisco:Jossey-Bass.

[4]Trowler, V. (2010) Student Engagement Literature Review. The Higher Education Academy.

[5]Coates, (2005) The value of student engagment for higher education quality assurance. Quality in Higher Education, 11(1):25-36.

Monitoring Learning Attitude and Engagement from Students' Free-style Comments via Lexicon-based Analysis

Kazuharu Toyokawa
College of International Relations, Nihon University
Mishima, Japan, E-mail: toyokawa.kazuharu@nihon-u.ac.jp

Abstract

Observing students' emotions, learning attitude and engagement, a teacher adapts his/her teaching and provides appropriate feedback to them. In a large classroom setting, however, it is not easy to observe directly each student emotions.

In this paper, we report experiments for monitoring students' attitudes in learning from free style comments after each lesson. We used several lexicon based analysis methods. One method used a lexicon for affect analysis, and the other is that of listing characteristic words of course topics.

The experimental result shows that proposed methods can monitor engagement to learning of each student and a whole classroom. Also the methods can overlook final grade score of each student within 16 through 17% root mean square error.

1. Introduction

Observing learning attitudes and engagements of students is important for an effective teaching in a classroom. Positive attitudes mean motivation and success, and negative signals require a teacher to respond and adapt to students.

Observing students' attitudes during a whole semester may be possible in a small group. However, that is not easy job in a larger classroom. In larger settings, students' free-style classroom comments are expected to provide information about learning related emotions.

Makoto Abe applied sentiment analysis to students' reflective comments about their English reading performance, investigating learner's understanding levels using a text-mining method [1]. Nabeela Altrabsheh et. al. conducted experiments to predict learning related emotions and estimated risks of course drop-out from classroom feedback via Twitter [2]. Jingyi Luo et. al. analyzed free-style reflective classroom comments using Word2Vec and ANN techniques, predicting final student grades [3].

In the prior work, analysis of classroom comments were mainly focused on predicting final score or outcomes. More useful information for learning and teaching would be extracted from classroom comments.

In this paper, we report experiments for monitoring students' attitudes in learning from free reflective comments in classroom via affect lexicon and content-oriented lexicon based analysis, and discuss its effectiveness to observing students' learning process and predict final scores.

2. Experimental

2.1. Collecting comments

The comments were collected from a lecture on information media literacy taught at the undergraduate course in Nihon University. Encouraging each student to write free comment, request, opinion or question if any on a attendance flip, we collected these after each lesson in a course. 29 lessons were given during the first half-year term in 2015. 74 students were enrolled to the course.

We input written comments, attaching date and student ID into a computer. All comments were written in Japanese, and any written mistakes such as incorrect characters and/or grammatical errors were not corrected.

Table 1 shows number of collected comments in each lesson.

Table 1: Number of comments in each lesson

Lesson	Date	Num	Lesson	Date	Num	Lesson	Date	Num
1	6-Apr	40	11	18-May	24	21	22-Jun	30
2	9-Apr	8	12	21-May	52	22	2-Jul	30
3	13-Apr	10	13	25-May	15	23	6-Jul	24
4	16-Apr	9	14	28-May	51	24	9-Jul	23
5	20-Apr	35	15	1-Jun	18	25	13-Jul	13
6	23-Apr	8	16	4-Jun	13	26	16-Jul	20
7	27-Apr	13	17	8-Jun	37	27	20-Jul	52
8	30-Apr	47	18	11-Jun	15	28	23-Jul	42
9	11-May	12	19	15-Jun	36	29	27-Jul	21
10	14-May	37	20	18-Jun	15			

2.2. Comments data preprocessing

We prepared the dataset for analyzing students' learning attitude from the collected comments. We used a Japanese text-mining tool: KH Coder[1], which included a

[1] http://khc.sourceforge.net/

Japanese morphological analyzer, a term extractor and utilities such as statistics and graphics functions.

We extracted words and part of speech from each comment, took only noun, verb, adjective, adjective verb and adverb into accounts, and took a number of aggregation for each lesson. The average number of words in all comments was about 330 in each lesson, and the average number of words without duplication was about 180 in each lesson, and about 2,220 in all comments.

2.3. Features of emotion and engagement

We applied three methods to monitor students' emotion and engagement from comments. One is lexicon based affect analysis, and the other is a method to focus attention to characteristic words and terms in each lesson. The third is a method to measure length of comment text.

2.3.1. Lexicon based affect analysis. According to a method of an affect analyzer 'Nazuki'[2], we applied 5 affect categories and 81 affect expressions which is partially shown in Table 2.

Table 2: Affect Categories and Expression

No	感性表現	分類項目	Affect Expression	Affect Category
1	嬉しい	好評	Glad	Popular
5	安心		Peace of mind	
7	感謝		Thanks	
8	感動		Excitement	
9	満足		Satisfaction	
10	快い		Agreeable	
14	喜び全般		Joy General	
15	期待		Expected	
19	楽しみ全般		Fun General	
20	誉め･賞賛		Praise	
21	好き		Like	
24	好評･人気		Popular	
25	対応が早い		Quick action	
26	対応が適切		Appropriate action	
27	対応への賞賛		Praising action	
28	説明が良い		Description good	
29	良い		Good reputation	
77	買いたい		Want to buy	
30	怒り	苦情	Anger	Complaint
32	不満		Dissatisfaction	
39	効果が不満		Effect dissatisfaction	
41	対応が不適切		Inappropriate action	
42	対応への不満		Action dissatisfaction	
44	説明が悪い		Description bad	
50	諦め		Give up	
51	残念		Unfortunately	
59	苦しい		Painful	
61	恐怖		Fear	
62	不安		Anxiety	
63	嫌い		Hate	
64	困っている		You are in trouble	
71	お願い	要望	Please	Demand
70	要望		Demand	
73	疑問	質問	Question	Question
74	問合わせ		Inquiry	
69	驚き	その他	Surprise	Other
81	勧誘		Soliciting	

We constructed a lexicon registering specific words for each affect expression from all comments. We registered 80 unigrams and 1 bigram in 'Popular', 31 unigrams in 'Complaint', 1 unigram in 'Demand', 2 unigrams in 'Question', and 4 unigrams in 'Other'.

Using the lexicon, we counted number of affect expressions in each comment, observing a student's emotion in each class, and also in a whole class.

2.3.2. Analysis focusing on Characteristic Words. In university class in general, a lecturer talks and discusses various topics from a different point of views in each lesson. A students would consider covered topics and mention these in his/her comment of each lesson.

Some students may write down a comment like a routine such as "Clear explanation and I like it." We read such tendency in collected comments.

So we introduced a method to measure a degree of engagement to lesson by counting characteristic words in a comment.

The characteristic word of a lesson is selected by Jaccard coefficient of word A in n-th lesson:

$$Jac(A,n) = \frac{Freq(A,n)}{Freq(A,n) + Freq(A, all\ other\ lessons)}.$$

where $Freq(A, n)$ is the occurrence of word: A in comments of n-th lesson. We listed each word in the order of Jaccard coefficient and selected top 10 words as characteristic words in each lesson. Examples of characteristic words list in lessons are shown in Table 3.

The measure of engagement $E(c, n)$ in a comment c in lesson n is given as follows:

$$E(c,n) = \sum_{i=1}^{10} Jac(A_i, n) \cdot Freq(A_i, n),$$

where A_i is listed characteristic words.

[2] http://research.nttcoms.com/solution/textm/index.html

Table 3: Examples of Characteristic Words in Lesson

Lesson 7/13

Characteristic Word		Jaccard C.
ナレッジ	knowledge	.231
マネジメント	management	.188
知	intellectual	.154
ノウハウ	know-how	.154
暗黙知	tacit knowledge	.118
受ける	receive	.095
言葉	word	.095
経験	experience	.095
知識	knowledge	.094
難しい	difficult	.082

Lesson 7/16

Characteristic Word		Jaccard C.
黒川温泉	Kurokawa Onsen	.300
温泉	hot springs	.200
暗黙知	tacit knowledge	.130
行く	go	.107
旅行	travel	.100
教室	classroom	.100
暑い	hot	.095
親	parent	.095
上手	proficient	.083
昔	a long time ago	.080

Lesson 7/20

Characteristic Word		Jaccard C.
納豆	Natto	.423
マーケティング	marketing	.273
食べる	eat	.226
ユーザー	user	.151
商品	commodity	.147
ヘビーユーザー	heavy user	.115
セールス	sales	.109
気	care	.109
人	man	.108
ミツカン	Mitsukan	.096

3. Monitoring Learning Attitude and Result

3.1. Grade score prediction

Using these extracted features from comments, we applied linear regression model to predict final grade scores. We also add class attendance data of each student to independent parameters to determine grade scores.

We split all data into 3 almost equal size groups, and handled 2/3 data for learning, and 1/3 for verification, and cross-checked.

Root mean square error, RMSE, of predicted score vs. final score is shown in Table 4. Accuracies of prediction were slightly better for affect expression analysis and text length features than that of characteristic words.

Fig. 1 shows plots of final grade score and predicted score by affect expressions for each student in the order of final grade.

Table 4: Grade score prediction with linear regression

(RMSE %)

Feature	Learning data	Verification data
Affect expressions	0.162	0.174
Characteristic words	0.164	0.176
Length of text	0.162	0.174

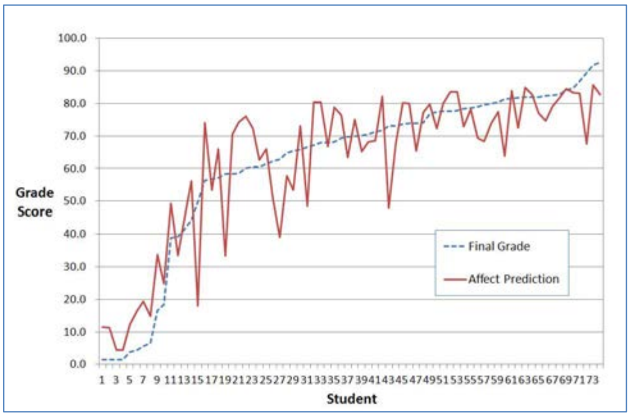


Fig.1: Final grade vs. predicted scores with affect feature

3.2. Monitoring learning attitude

We also found out that extracted features shows student's engagement to learning in a whole classroom. Fig.2 shows a plot of total measure of engagement in each lesson compared with number of attendances.

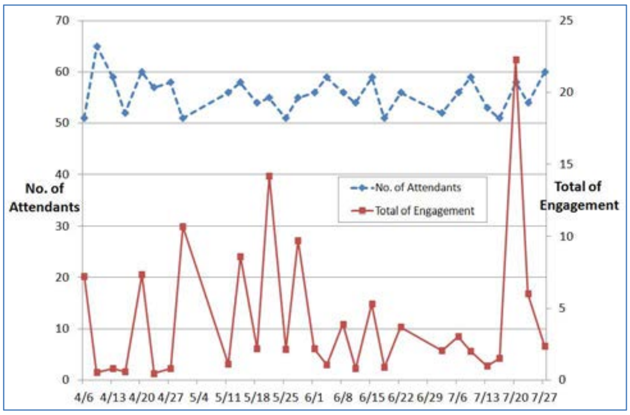


Fig.2: Observed total Engagement to Lesson

A Number of attendances to the class were around 50 through 60 during the semester, whereas total characteristic words were up and down, especially showed a peak on July 20th.

Actually, the lecturer talked business management using knowledge management scheme during 3 consecutive lessons on July 13th, 16th and 20th. Basic theory was given at first, and case studies were talked on the following two lessons. Students showed interesting on July 20th in the case study of Mitsukan Vinegar Ltd, a popular company for most students.

This fact shows that lecturer can monitor student interesting and engagement lesson to lesson in a whole classroom level.

The method can be applied to monitor each student too. As an example, Fig.3 shows plots of engagement of arbitrary selected 3 students during the semester. A plot of each student shows peaks at different lesson, and some coincidences too. So we can observe and compare engagement to lesson of each student.

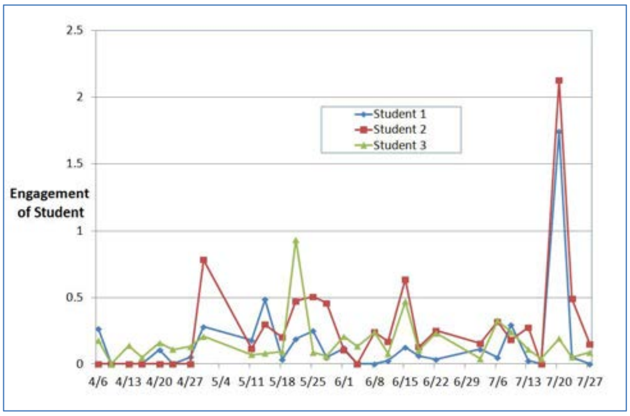


Fig.3: Engagement plot of students lesson to lesson

4. Conclusion and Future Work

In this paper, we analyzed free style written comment after each class, and realized that comment contains rich information of learning process and result of each student.

By using natural language processing method such as a lexicon based affect analysis, a method to focus on characteristic words on a lesson, and even a simple measurement of text length of comments, we can observe student engagement to learning. Therefore analysis of freestyle comments in class is found out to be effective method to monitor learning process and predict outcomes of learning of students.

A challenge is that there are some students who do not tend to write freestyle comments. For this reason, it was not simple to predict final grade score in good accuracy in this study.

Using Twitter instead of a written comment may one solution, however, appropriate educational consideration and classroom settings may be required to avoid for students to tempt to play with Twitter during face to face classroom.

5. References

[1] Makoto Abe. “An Analysis of Reflective Comments in English Reading Classes Using Text-Mining”, *Information Science Research Bulletin, Dokkyo University*, Jan. 2014, pp.164-168. (in Japanese)

[2] Nabeela Altrabsheh, Mihaela Cocea, and Sanaz Fallahkhair, “Predicting learning-related emotions from students’ textual classroom feedback via Twitter”, In *8th International Conference on Educational Data Mining (ICEDM 2015)*, 2015, pp. 436-439.

[3] Jingyi Luo, Shaymaa E. Sorour, Kazumasa Goda, and Tsunenori Mine, “Predicting Student Grade based on Free-style Comments using Word2Vec and ANN by Considering Prediction Results Obtained in Consecutive Lessons”, In *8th International Conference on Educational Data Mining (ICEDM 2015)*, 2015, pp. 396-399.

Collaborative learning in a knowledge integration task using PCAs: Investigation based on dyad eye movements

Yugo Hayashi
Ritsumeikan University, y-hayashi@acm.org

Abstract

The present study investigates the nature of dyad coordination in a knowledge integration task, where two Pedagogical Conversational Agents (PCAs) interfere to provide metacognitive suggestions. The goal of the study was to (1) investigate how the use of PCAs influences coordination during a collaborative integration task and (2) understand the nature of coordination. The study used two eye trackers to discover how a learner's coordination was successful when discussing and integrating two different concepts that were presented on the screen. In the experiment, two conditions (no agent condition vs. agent condition) were set and the data of coordination and task-related process were analyzed. The main results show that the relations of learners who are coordinating well perform well on task-related activities such as gazing towards the two concepts. In addition, the results show that such a tendency appears stronger in conditions where learners receive suggestions from the PCA.

1. Introduction

Studies in learning and cognitive science have shown that effective interaction strategies such as self-explanations [1,2] and taking different roles [3] are effective. Studies on more practice based learning have developed effective group based learning, such as jigsaw solving, and shown improved performance [4]. Research shows that learning through sharing knowledge with others can lead to conceptual changes that can bring new knowledge [5]. In addition, discussions based on different perspectives can bring an understanding of the content from a higher level [6]. These studies indicate that collaborative interaction for learning tasks such as integrating different knowledge and perspectives is effective. However, for novice learners, such activities are not easy to accomplish, and there is a need to investigate what kind of instructional and cognitive factors in interaction influence coordination.

From a cognitive viewpoint, language processing such as acquiring common knowledge is a critical process for learners. Studies in psycho linguistics have investigated the nature of how individual speakers establish common ground [7]. Studies in [8] manipulate a speaker's perspective in a collaborative task and show that knowledge integration is difficult. These studies indicate that collaborative learning activities such as integrating different knowledge and perspectives are difficult tasks to achieve, especially for novice learners.

Studies in intelligent tutoring systems have shown the effective use of Pedagogical Conversational Agents (PCAs) [9,10,11,12]. Few studies have investigated how a PCA can facilitate the collaborative learning of human-human interaction [13]. In [14], the authors investigated how affective feedback from the PCA can facilitate interactions in an explanation task. Learners formed an explanation of a key concept and the PCA joined their conversations by providing metacognitive suggestions. The present study focuses on how PCAs can facilitate the coordination of gazes in a dyad, which is known to be an effective method for understanding how well learners interact during a collaborative task [15,16]. In this study, an experiment was conducted by setting a concept integration task and investigating the nature of the coordination of learners and knowledge integration.

2. Goal and hypothesis

Our goal is as follows: (1) Goal 1: Investigate how the use of PCAs influences coordination during a collaborative integration task. (2) Goal 2: Understand the nature of coordination and how it relates to the knowledge integration process. To achieve our goals, the present study uses eye tracking methods to capture the nature of collaborative dyad interaction in this integration activity. Studies in learning science and CSCL(Computer Supported Collaborative Learning) have implied that eye-trackers are effective for understanding the nature of collaborative activities [17,18]. Communication studies such as [15,16] suggest that the degree of gaze recurrence between dyads (speaker-listener) is correlated with collaborative performance such as understanding and establishing common ground.

Using these eye tracking methods, we predict that (H1) learners will pay attention to the PCA when it interferes with the learners' collaborative activities. In addition, we predict that (H2) learners who coordinate well will also perform well on the integration task. It is supposed that both learners receiving suggestions from a PCA should notice what to discuss about and should perform better on the integration task; therefore, there should be a strong relationship between coordination and task performance.

3. Method

3.1. Task and settings

The present study focuses on the knowledge integration process of dyads (learners) in an experimental setting. The learners consisted of 52 students in a Japanese university majoring in psychology. Learners were formed into dyads and instructed to collaboratively explain two different types of sub-concepts and integrate them to gain a higher understanding of the concepts. Before the experiment, learners were required to study and be prepared to explain either one of the sub-concepts in the experiment. This was to create a situation where one of the concepts presented on their screen was familiar and the other unfamiliar. Learners sat together, facing each other with a computer monitor displayed in front of them.

On their screen, there were four areas: (1) sub concept 1, (2) PCA1, (3) sub-concept 2, and (4) PCA2.The areas were adjusted so that one of the learners could only see one of the sub-concepts. The sub-concept that was presented to the learner was the concept that the learner prepared for the experiment. A brief explanation was also presented for this concept. Such manipulation made learners to focus on this area, biased towards his/her familiar knowledge. If the familiar knowledge is presented in area1, it is predicted that the amount of gaze patterns will appear on area1 compared to area3. In addition, if the learner tries to consider about the other's concept (Integrate other's perspective), it is predicted that gaze may transit from area 1 to 3 rapidly. This point will be investigated in the gaze pattern analysis. During the task, learners were required to explain each of their sub-concepts and then discuss further to reach a more abstract understanding of the concepts. Learners were instructed to fist make explanations about each of their familiar sub-concepts and then try to integrate that knowledge to understand from a wider view. They were also instructed that two PCAs will appear on the screen and each will provide suggestion on how to make explanations and tips for communication during the activity. Learners received an average of 10 messages from the PCAs during the activity.

3.2. Experimental Design

In this study, we used two PCAs that were developed in a previous study [14]. Two PCAs were used because previous study shows it is effectiveness on producing social awareness [14]. The role of these PCAs were to provide meta-cognitive suggestions on making explanations about each of their concepts. Also, they were programmed to provide messages about finding the connections between the two concepts and develop an abstract knowledge about the concepts. The system was developed in Java and programmed for server-client networks. Information is collected through each client application that the learners are using and are sent to the server to process the responses by the PCA. The PCA consists of several related modules: a module that detects the input messages of each learner (Input analyzer), a rule-based engine that controls the type of response (Generator), and a module that sends synthesized speech and gestures to the learners (Output handler). Within a minute, each PCA will provide feedback based on the detected keywords. The system provided meta cognitive suggestion or back-channel feedbacks to facilitate their conversation during their activity. In the current experiment, as explained in the next section, we analyzed the eye movements of the learners. Thus, it was necessary to fix the learner's head movements. Hence, we did not use the text based chat that was used in the previous study. Instead, the experimenter input some of the key phrases into the system and the PCA responded automatically based on the rule. The rules were set for this study to respond to some of the important keywords the learners mentioned (e.g., long-term memory, episodic memory, or implicit memory). When the system detected some of these words, they provided suggestions that would facilitate their metacognitions. Learners were randomly assigned to two conditions (agent condition: n=26 and no agent condition: n=26). In the no agent condition, no PCAs were presented on the screen. Learner eye gazes and verbal protocols were all collected. The explanation time was 10 minutes.

3.3. Measures

In the present study, two eye-trackers (TobiiX2-30) were used for detecting the learner eye movements. Using these devices, we analyzed two types of gaze patterns: (a) task related individual transition and (b) dyad coordination. For (a), we counted and analyzed the frequency of gaze transitions between the four different areas in the screen. In this knowledge integration task, it is important to look at both concepts (area 1 and 3), thus we analyze the frequency of gaze patters between these two areas. In the no agent condition, the windows of areas 2 and 4 were not presented. In (b), based on [15,16], both learners' eye movement data were analyzed using recurrence analysis. This method enables to capture the proportion of fixations that are at the same location for both learners in a typical time state. The recurrence of *phi* observed between the two time-series (Learner A and B) is calculated on a specific time state *k*. The *phi(k)* coefficient increases with the frequency of matching recurrence on the same state *(k; k)* and decreases with the frequency of mismatching.

4. Results

4.1. Coordination

For each dyad, we calculated the recurrence of gaze for each area. A one-way ANOVA was conducted by dyad to investigate if gaze coordination changed because of the condition, although there were no differences ($F(1, 24) = 0.0989$, $p = .756$). Interestingly, even though learners in the agent condition had twice as many areas to look at (areas 1 to 4), they had the same gaze co-occurrence as those in the no agent condition. In the next section, we look at where learners were actually paying attention in the task.

4.2. Transitions toward each area

To investigate how a learner's gaze was influenced by agents, a statistical analysis was performed using a 2 (condition) X 12 (area) mix-way ANOVA. A significant interaction was found between the two factors ($F(1, 550) = 19.385$, $p = .000$). To see how learners were looking at the task-related important areas during the task, a focused analysis for a simple main effect was conducted based on area 1 (concept A) and area 3 (concept B). Results show that learners in the no agent condition showed higher transitions than those in the agent condition, indicating that learners were focusing on various points of the screen ($F(1, 50) = 28.640$, $p = .000$; $F(1, 50) = 33.683$, $p = .000$). In the next section, we see how these gaze patterns were related to learners' coordination performance and consider how PCAs can navigate a learner's coordination that is task-related to learning (in this case, to knowledge integration).

4.3. Correlation

To investigate our second goal of this study, here we focus on the correlations of the gaze recurrence on and transitions between the important areas. In particular, we look at the transitions between areas 1and 3 to see the relation of the learners 'integration process (considering the two concepts) and coordination process.

Figure 1 shows the relation between the two indices in the no agent condition for areas 1 to 3 and 3 to 1 (r=0.076, -0.082, respectively). This implies learners using PCAs facilitated coordination along with task related gaze.

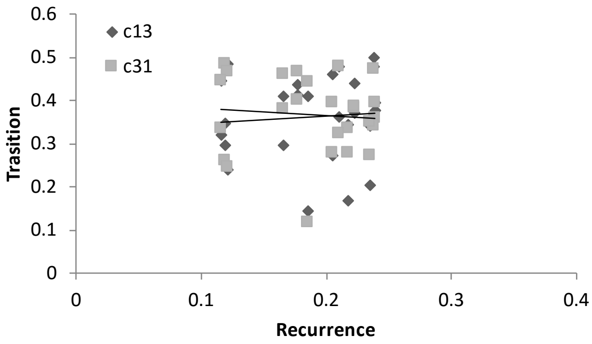


Figure 1. Results of the correlation in the no agent condition.

Figure 2 shows the relations for areas 1 to 3 and 3 to 1 in the agent condition (r=0.258, 0.227, respectively). The results indicate that learners in the agent condition had a stronger correlation between the two indices.

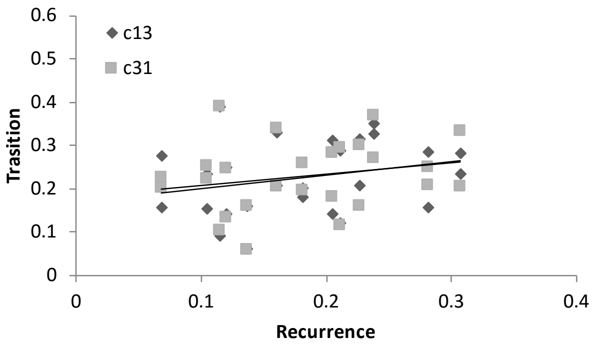


Figure 2. Results of the correlation in the agent condition.

5. Discussion and Conclusion

The present study focused on one of the efficient strategies of collaborative learning, i.e., knowledge integration. Our long-term goal is to develop intelligent tutoring systems that can support collaborative learning. To develop effective intelligent tutoring systems for collaborative learning, it is important to first understand how interactions change when we use such technology. The present study (1) investigated how the use of PCAs which provides facilitation of making explanations influence coordination during a collaborative integration task and (2) attempted to understand the nature of coordination and how it relates to the knowledge integration process.

The PCAs provided messages to the learners to think about the connections between the two different concepts and integrate their thoughts. We focused on how learners look at areas (area 1 and 3) which are important when integrating the two different concepts, and the areas of the PCAs (area 2 and 4). As predicted, the transition of gaze frequency between areas 1 and 3 reduced in the agent condition compared to the no agent condition. This is because the learners paid more attention to the two PCAs (areas 2 and 4). Interestingly, in the agent condition, even though they had more areas to look at, it did not obstruct coordination performance. Further analysis shows a positive correlation between coordination and transition (integration process) when learners used the PCAs. This indicates that learners who received metacognitive comments from a PCA were able to facilitate coordination that is task related such as looking at the concepts for integration. On the other hand, without PCAs, even if learners coordinated well, this fact did not relate to the gaze on the task-related concepts. Results show that when using PCAs with meta-cognitive suggestions, it not only help them address their gaze but can also facilitate their gaze synchronization.

As discussed in the beginning of this paper, it is known in language studies that gaze recurrence between dyads is important on establishing common ground during communication [15,16]. The results in this study provide possibility that the use of agent technology in a dyads communication can be used for coordination. However, there are needs to investigate what kinds of factors such as type of addressing and facilitations from the agent can be useful. Studies from human-human interaction can provide insight into the development of such systems in the future. As described above, result can be interpreted as an indication that the presence of the PCAs facilitates efficient interaction. However, it is not quite clear how coordination affected the integration process directly in this current analysis. Further analysis could be conducted by analyzing the verbal data in the future.

The present study provides new implications of how such technology can be used for coordination in knowledge integration tasks. In the preliminary study by [19], the authors have not yet discovered the effects of the suggested agent technology on learning performance. Therefore, using these implications in the current study, we plan to further investigate what kind of factors are effective for increasing performance.

Acknowledgments

Grant-in-Aid for Scientific Research (KAKENHI), The Ministry of Education, Culture, Sports, Science, and Technology, Japan (MEXTGrant), Grant No. 16K00219.

References

[1] Chi, M., Leeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3), 439-477.

[2] Miyake, N. (1986). Constructive interaction and the interactive process of understanding. *Cognitive Science*, 10(2), 151-177.

[3] Shirouzu, H., Miyake, N., & Masukawa, H. (2002). Cognitively active externalization for situated reflection. *Cognitive Science*, 26(4), 469-501.

[4] Aronson, E., & Patnoe, S. (1997). *The jigsaw classroom (2nd ed.): Building cooperation in the classroom.* New York: Addison Wesley Longman.

[5] Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *Journal of the Learning Sciences*, 2(3), 235-276.

[6] Schwartz, L. D. (1995). The emergence of abstract representation in dyad problem solving. *Journal of the Learning Science*, 4, 321-354.

[7] Clark, H. H. (1996). *Using language*. Cambridge University Press.

[8] Hayashi, Y., & Miwa, K. (2009). Prior experience and communication media in establishing common ground during collaboration. *In Proceedings of the 31th annual conference of the cognitive science society(cogsci2009)* (p. 526-531).

[9] Graesser, A., & McNamara, D. (2010). Self-regulated learning in learning environments with pedagogical agents that interact in natural language. *Educational Psychologist*, 45(4), 234-244.

[10] Kumar, R., & Rose, C. (2011). Architecture for building conversational architecture for building conversational agents that support collaborative learning. *IEEE Transactions on Learning Technologies*, 4(1), 21-34.

[11] Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review*, 6(1), 27-54.

[12] Moreno, R. (2005). *Multimedia learning with animated pedagogical agents.* In R. Mayer (Ed.), The Cambridge handbook of multimedia learning (p. 507-524). Cambridge University Press.

[13] Holmes, J. (2007). Designing agents to support learning by explaining. *Computers & Education*, 48(4), 523-547.

[14] Hayashi, Y. (2012). On pedagogical effects of learner-support agents in collaborative interaction. In *Proceeding of the 11th international conference on intelligent tutoring systems(its2012)* (p. 22-32).

[15] Richardson, D., C, & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, 29(6), 1045-1060.

[16] Richardson, D., C, Dale, R., & Kirkham, N., Z. (2007). The art of conversation is coordination. *Psychological Science*, 18(5), 407-413.

[17] Jermann, P., Mullins, D., Nuessli, M.-A., & Dillenbourg, P. (2011). Collaborative gaze footprints: Correlates of interaction quality. *In proceedings of the annual conference of computer supported collaborative learning (CSCL 2011)*, 184-191.

[18] Schneider, B., & Pea, R. (2014). Toward collaboration sensing. International *Journal of Computer-Supported Collaborative learning*, 4(9), 5-17.

[19] Hayashi, Y. (2016). Coordinating knowledge integration with pedagogical agents: Effects of agent gaze gestures and dyad synchronization. In *Proceeding of the 13th international conference on intelligent tutoring systems(its2016)* (p. 254-259).

The Analysis of Learning Effectiveness of Chinese Proficiency Test Tutorial System Including Capability Indicators

Chen-Yu Lee[1*], Gwo-Haur Hwang[1], Long-Siang Huang[1], Hsu-Yu Lee[2]
[1] Department of Information Networking and System Administration, Ling Tung University.
[2]General Education Center, Ling Tung University.
*cylee@teamail.ltu.edu.tw

Abstract

Language is an important tool of communication and cultural heritage in human society. However, Chinese language capability is generally poor for most undergraduate students in Taiwan; there are many typos and syntax errors frequently appeared in their homework and articles. In order to improve their Chinese language capability, many universities have organized the Chinese Proficiency Test to achieve this purpose. A traditional online question bank system typically provides only basic test function without providing any feedback; thus, the effect of improving Chinese language capability of learners is limited. Therefore, it is intended in this study to develop a "Chinese proficiency test tutorial system including capability indicators", and to explore the effect of the proposed system. The results indicate that learning effectiveness of learners can be increased by using the proposed system.

Keywords: Chinese language capability, learning feedback, learning motivation, learning effectiveness

1. Introduction

Along with popularity of internet and digital learning, Lipnevich and Smith found out that students prefer computer-based tests, and thus the question bank system has become a commonly used tool over the past years [1][2]. Umar pointed out that the simulation tests of question bank system allow students to be used to the setting of a formal examination during the process of repetitive practice, and then to improve their own skills [3]. However, it is important to increase students' learning motivation and learning effectiveness by using a question bank system. Devedzic considered that it can increase learners' learning effectiveness when applying the strategy of feedback [4]. Hence, multiple human factors including learning motivation and learning attitude are taken in account in this study; it aims to explore the impact of using Chinese proficiency test tutorial system including capability indicators on learning effectiveness.

2. Literature review

2.1. Chinese proficiency

Zhong suggested that language is the primary tool to learn and communicate and even the base of every subject [5]. In recent years, foreign language received more concerns than Chinese, and therefore national examination had ever revoked Chinese writing; it results in a significant decline of students' writing ability, and Chinese which represents their own culture is almost replaced by foreign language; it is an issue deserving more attention [6]. Confucius said: "They who know the truth are not equal to those who love it, and they who love it are not equal to those who delight in it."; students' motivation often affects learning effectiveness; thus, it becomes the most pressing issue to trigger students' motivation to learn, and so as to improve learning effectiveness.

2.2 Computer-based learning assessment

The study of Rudner suggested that although the procedure to build a question bank is similar to the one to edit a test, the former one need "calibration" to ensure the representativeness of questions and conforming to the developing principles of tests; moreover, it has to ensure the discrimination among questions and conforming to the principles of reliability and validity via preparatory test; hence, it needs quite a lot of labor, resource and time to build a question bank [7]. The steps of building a question bank mainly include: (1) to build a two-way breakdown, (2) to formulate questions based on proposition principles, (3) revise and review questions, (4) pre-test and question analysis, (5) screening questions and (6) input the content of questions and question parameters into the computer database. Wu has similar views for the question bank; he considers that the question bank is composed of a group of questions, and it can be classified by the content of question or the statistical properties; in other words, a question bank must be established through a basic specific procedure, such as making a two-way breakdown, preparation of questions, pre-test, analysis of questions, question selection and creation, and so on [8]. Chen suggested that a question bank is not just a collection of a large group of questions; the questions within the question bank need calibration, analysis, classification and evaluation, and they are composed by an appropriate process [9].

In recent years, along with the advancement of information technology; there is a rapid growth of development of computer software and hardware; interface design becomes increasingly affinity and ease of use. Fletcher and Collins found that students prefer to take computer-based tests [10]. Yang considered that a new generation of question bank is even developed to be a more effective and positive teaching assessing tool; it provides the required information to instructors, mentors and subjects; moreover, both of teaching and assessment are even bound closely together to make learners feel easier to understand the learning effectiveness [11]. Therefore, a computer-based question bank provides a lot of help for school teachers and test developers. However, there are many advantages for an assessment and test system, such as effectively increasing the efficiency and feedback, resource sharing, group discussion, and the freedom and flexibility to select the place for a test [12].

2.3. Learning motivation and learning effectiveness

Johnson, Chang, and Lord proposed that the assessment is an essential part for a teaching activity; the results of assessment will be helpful to improve the effect of teaching and the quality of learning via interpretation and analysis. The results of assessment allow teachers to understand students' learning effectiveness and their own teaching achievement for a subsequent guide of students' learning [13].

3. System feature

3.1. Question category

According to the question categories in Chinese Language Proficiency Test, there are four corresponding categories of questions in this system: word form, word pronunciation, word meaning and idiom; in addition, a comprehensive test includes all four categories of questions. When students have a test via this system, they can select single or multiple categories they want to test. A screenshot of selection of question categories is shown as Figure 1, and another screenshot of question answering in a test is shown as Figure 2.

Figure 1. Selection of question categories

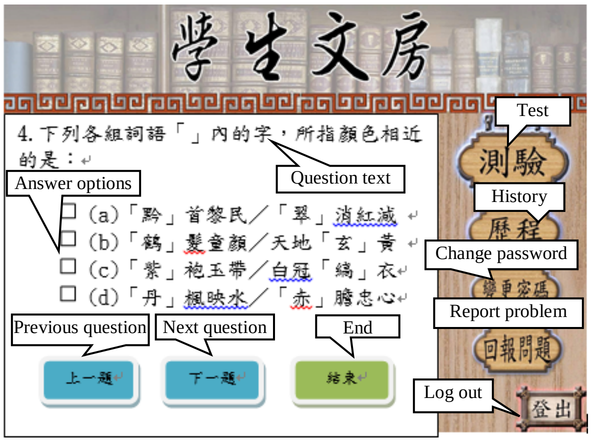


Figure 2. Question answering in a test

3.2. Capability indicator

Upon completion of a test, in addition to the display of a basic score, a radar chart which includes four capability indicators corresponding to four categories of questions is also provided; it allows students to make a full understanding of their own weakness, and then to enhance their capability with an intensive practice for those categories they have a poor score. It is expected to improve students' Chinese language proficiency. A screenshot of a test score with a Radar chart displaying capability indicators for a comprehensive test is shown as Figure 3.

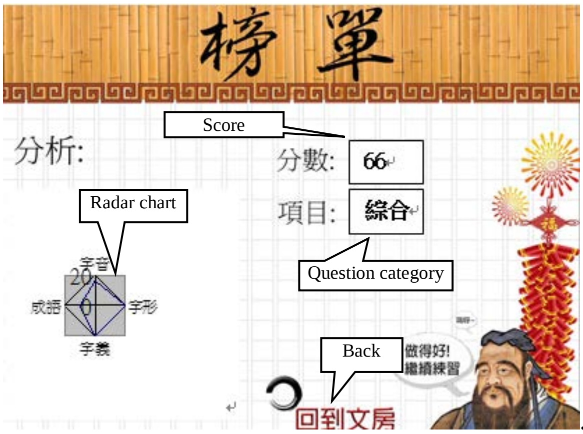


Figure 3. A test score with capability indicators

3.3 History of test

Students can inquire their own history of test to know the number of times to test and to pass, as well as category, score, start time and end time for each test. A screenshot of history of test is shown as Figure 4.

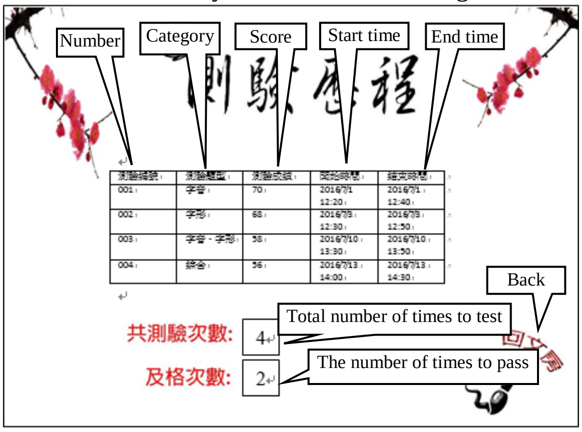


Figure 4. History of test

4. Experiment design

4.1. Subject

A total of 173 freshmen studying in four classes in a technology university at the Central Taiwan are served as the subjects by convenient sampling in this study; all of them are taught by the same teacher with the same textbook; their major is either business administration or information technology; considering the balance of major, two classes of 93 students are assigned to be the experimental group, and the other two classes of 80 students are assigned to be the control group. The effective sample size are 160 subjects with the entire process of participation, where 84 ones for the experimental group and 76 ones for the control group.

4.2. Tool

A questionnaire with Likert five-point scale is used for the evaluation of subjects' learning motivation and learning attitude [14][15]. In addition, a statistical software, SPSS, is used for the analysis of results.

4.3. Experiment process

At first, participated subjects take a Chinese proficiency pre-test, and fill out the questionnaire of learning motivation and attitude; then for eight weeks, the experimental group uses a Chinese proficiency test tutorial system including capability indicators, while the control group uses a traditional LMS question bank system. Finally, a Chinese proficiency post-test is implemented. The experiment process is shown as Figure 5.

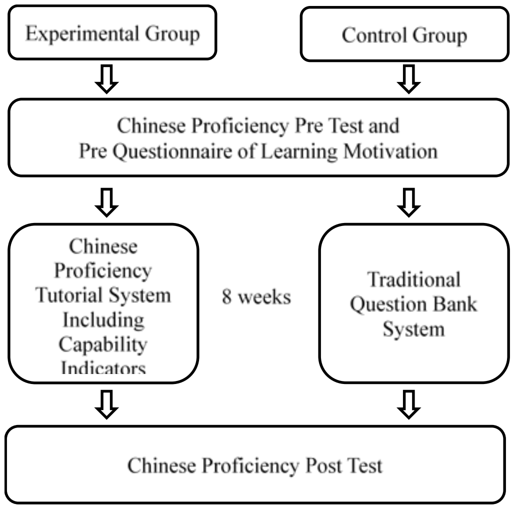


Figure 5. Experiment process

5. Result of analysis

This study use SPSS19 software to conduct ANCOVA analysis and discuss the effect of proposed system which includes capability indicators. The results indicate that the learning effectiveness of subjects using the proposed system which includes capability indicators is significantly better than that of subjects using traditional non-feedback question bank system; it is shown as Table 1. For those subjects with high learning attitude in the experimental group using the proposed system which includes capability indicators, their learning effectiveness is higher than that of those with high learning attitude in the control group, where it is shown as Table 2. However, ANCOVA is not performed for low attitude (Table 3) and low motivation (Table 5); there is no significance of effectiveness for high motivation (Table 4).

Table 1. Analysis of the scores of pre-test and post-test for the experimental group and the control group

Source	SS	df	MS	F	p
Covariate	2769.644	1	2769.644	51.467	0.000
Between	294.045	1	294.045	5.464	0.021
Within (Error)	8448.729	157	53.814		
Total	1016084.000	160			

*p**<0.05

Table 2. Analysis of the scores of pre-test and post-test for the experimental group and the control group with high attitude

Source	SS	df	MS	F	p
Covariate	407.848	1	407.848	8.969	0.004
Between	370.094	1	370.094	8.139	0.006
Within (Error)	3182.954	70	45.471		
Total	474416.000	73			

*p***<0.01

Table 3. Analysis of the scores of pre-test and post-test for the experimental group and the control group with low attitude

Source	SS	df	MS	F	p
Between x Covariate	276.273	1	276.273	5.150	.026
Error	4452.947	83	53.650		

Table 4. Analysis of the scores of pre-test and post-test for the experimental group and the control group with high motivation

Source	SS	df	MS	F	p
Covariate	768.562	1	768.562	13.213	0.001
Between	47.313	1	47.313	0.813	0.370
Within (Error)	3897.268	67	58.168		
Total	457256.000	70			

Table 5. Analysis of the scores of pre-test and post-test for the experimental group and the control group with low motivation

Source	SS	df	MS	F	p
Between x Covariate	523.618	1	523.618	11.822	.001
Error	3809.040	86	44.291		

6. Conclusion

The effect of Chinese proficiency test tutorial system including capability indicators is discussed in this study. The findings show that using the proposed system is helpful to improve the learning effectiveness of students, and it achieves even a higher significant level for those students with high attitude; hence, a test tutorial system with feedback (capability indicator) given is able to increase the learning effectiveness of students.

7. Reference

[1] A.A. Lipnevich, J.K. Smith, Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied, 15,* 2009, pp. 319-333.

[2] V.E. Dennis, D. Gruner, Computer managed instruction at Arthur Andersen & Computer: A status report. *Educational Technology, March,* 2009, pp.7-16.

[3] J. Umar, Item banking. *In G. N. Masters & J. P. Keeves (Eds.), Advances in 157measurement in educational and research and assessment,* 1999, pp.207-219.

[4] V.B. Devedzic, Key issues in next-generation web-based education. *IEEE Transactions on Systems, Man, and Cybernetics, Part C, 33*(3), 2003, pp. 339-349.

[5] P.L. Zhong, Improving Chinese language capability of students in Normal University by Chinese language activity. *The Report of Forum,* 2012.

[6] H.C. Liao, Improvement of reading ability by Chinese language. *The Report of Forum on "the Integrated Study of Planning, Implementation and Self-Evaluation of Curriculum Improving Chinese Reading/Writing Ability of Students in Normal University",* 2012.

[7] L. Rudner, The effect of type and timing of feedback on learning from multiple-choice tests. *Practical Assessment Research & Evaluation, 6*(4), 1998, pp. 1-7.

[8] H.Y. Wu, A multimedia computer question bank system: Design of editing system. *Distance Education (The Journal of R.O.C. Distance Education Society).* 1999.

[9] H.F. Chen, The research and development of on-line question bank system. *Journal of Secondary Education, 10,* 2003, pp. 301-326.

[10] P. Fletcher, M.J. Collins, Computer- administered versus written test- advantages and disadvantage. *Journal of Computer in Mathematics and Science Teaching, 60*(2), 1987. pp. 38-43.

[11] C.Y. Yang, M.J. Collins, A study on learning effect among different learning styles in a web-based lab of science at elementary schools. *IEEE International Conference on Advanced Learning Technologies,* 2000. pp. 80-82. *National Kaohsiung Normal University, Kaohsiung,* 2000. pp. 5-8.

[12] Y.L. Chou, A Study on Applying Cloud Technology to a Computer-Based Testing System, 2011.

[13] R.E. Johnson, C.S. Chang, R.G. Lord, Moving from cognition to behavior: What the research says. *Psychological Bulletin, 132,* 2006, pp. 381-415.

[14] R.R. Pintrich, E.V. DeGroot, Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82,* 1990, pp. 33-40.

[15] G.J. Hwang, H.F. Chang, A formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. *Computers & Education, 56*(1), 2011, pp. 1023-1031.

Computer Assisted Learning to Improve Educational Performance: A Study of Aboriginal Taiwanese Kids

Chyi-In Wu
sss1ciw@gate.sinica.edu.tw

Abstract

Recent consideration about E-learning has been placed on whether integrating Information Communication Technology into education can effectively improve educational performance. Still, the empirical evidence of the impact of programmes that adopt ICT in schooling is mixed [1]. Theory suggests it may be due to differences in whether or not the ICT programmes are integrated into a class. In order to understand the most effective way to design new programmes that attempt to utilize ICT to improve Math and Chinese learning, we conducted a clustered randomized controlled trial (RCT) with some classmates receiving ICT that was integrated into their learning behaviors; with some other classmates that did not receive ICT programme; and with other classes being used as controls. The RCT involved 1,407fourth and fifth grade students studying Math and Chinese in 86 aboriginal elementary schools in South and East parts of Taiwan. This ongoing study should be able to indicate that when the programme is integrated into the learning habits of a classmate it would be effective in improving student test scores relative to the control classes.

1. Introduction

Despite the high level economic and educational development in Taiwan, there is a persistent “development gap” between the aboriginal Taiwanese (or Yuanzhumin) kids and Han kids. According to a series of benchmarking studies, aboriginal Taiwanese kids are more likely to be from families under the official poverty line. Also, the academic achievement among aboriginal Taiwanese lags drastically behind Han kids.

What can account for this disparity? Although the government has invested significantly in the upgrading of Yuanzhumin schools, schools which are mostly attended by Han students still far outpace in quality their Yuanzhumin counterparts. As a consequence, Yuanzhumin students are at risk of being left behind. Evidence indicates that overall educational achievement of Yuanzhumin students is declining relative to Han students.

Besides the relatively poor state of facilities and teacher quality, lack of care outside of the classroom might be an important reason for the poor performance of Yuanzhumin students. Children from Yuanzhumin schools are at a disadvantage because they typically cannot access as many resources as the Han students—especially resources that may be able to help them if they fall behind. Unlike their Han counterparts, students enrolled in poor Yuanzhumin schools do not have before and after class review sessions in school by qualified teachers, cannot afford commercial tutoring classes, or do not have interested and well-educated parents to help them with their homework. For these reasons, although Yuanzhumin students often have plenty of time, they are less likely to get source of remedial care that could help them catch up or get ahead.

Integrating Information and Communication Technology (ICT) into the learning process is a promising approach to help disadvantaged students across the world [2][3]. Scholars who concerned this issue have tested various methods to use ICT for innovation and improvement of poor performance students. Despite the increasing use of ICT for helping disadvantaged students in learning, researchers have found large heterogeneity in the impact of ICT programmes on student academic achievement [4][5][6].

2. Objectives of this Study

What can be done to address the shortcomings in Yuanzhumin schools? We propose conducting a high quality, innovative program that has high potential to improve educational outcomes for Yuanzhumin children: A desktop PC based remedial learning program in the homes of Yuanzhumin students. The program target Chinese language skills, a key subject area where the gap is among the largest between Yuanzhumin students and Han students. Studies have also shown that the improvement of Chinese language may have important spill over effects on other subjects such as math. In other words, the improvement in Chinese may enhance the understanding and mastery of the knowledge of math. Thus, a remedial learning program that targets Chinese language skills may have positive impact on other subjects as well. As discussed, the intervention has particular promise, but it has not been rigorously evaluated in Taiwan before.

The goals of this proposed study can be divided into three categories: First, for intellectual objectives, provide clear, quantitative evidence that determines whether a desktop based CAL program at home can improve the educational performance of Yuanzhumin children, including non-cognitive effects (such as self-esteem and love of learning). Second, for educational objectives, develop a curriculum (and training manual for the students to learn how to use the software). The intervention will be carefully designed so that if it is found to be effective, the school system can implement it cost effectively and on a larger scale. Third, for policy objective, if found to be effective, convince relevant policy makers that such a CAL program can lead to better educational outcomes and should be scaled up in areas where Yuanzhumin students are concentrated or areas where poor student performance is chronic.

3. Approach

We propose to conduct a desktop PC based Chinese CAL program in Yuanzhumin kids' homes that aims to achieve the above goal. We randomly select 86 Yuanzhumin schools in which to run our evaluation of the CAL intervention arm. Among the classes in those schools, we randomly select half of the students in each class to receive intervention and the other half in the class to serve as control. The intervention target two grades, fourth and fifth grades. This arrangement will allow us to evaluate the impact of the intervention on student outcomes, such as test scores and non-cognitive traits.

During intervention, Yuanzhumin students use home PC to access the game-based remedial tutoring software. Before the intervention, we visit each treatment school to hold a training session among all treatment students and their parents. The training introduces how to use the online software at home to review and practice what students leant in class. The content of the software was matched with the curriculum in Chinese teaching at school. Students get an account to log into the software anytime over Internet at home. Their activities were recorded by the software in a way that allows us to analyse their learning progress.

The e-learning program have four core components: Hardware – wired desktop PCs at home; Software – educational gaming software; Curriculum – the software content will match exactly with mandate Chinese curricula, and Protocol – to include training regimen and exhaustive program handbook to organize trainings to teach how to use the software among children and parents, and 24/7 technical support. The combination of these four components ensures that regardless of prior computing background of the Yuanzhumin children, they can access and use the educational software to improve learning at home. We intend to make our new software to address the needs of all students (not just the poor performing). We will have two levels of difficulty in the software so that students can move at their own pace. Both poor-performing and higher-performing students will find the software attractive and helpful.

4. Evaluation

After the intervention is completed, we conduct an endline survey among all students in all schools (intervention and control). The survey will capture important data on student standardized test scores, family background, and economic standing. Differences that we observe between the outcomes of the treatment and control will be attributable to the intervention.

5. References

[1] Bai,Yu, Mo, Di, Zhang, Linxiu, Boswell, Matthew, and Rozelle, Scott. (2016). The impact of integrating ICT with teaching: Evidence from a randomized controlled trial in rural schools in China. *Computers & Education* 96:1-14.

[2] Lai, F., Luo, R., Zhang, L., Huang, X., & Rozelle, S. (2011). Does computer-assisted learning improve learning outcomes? Evidence from a randomized experiment in migrant schools in Beijing. REAP working paper.

[3] Mo, D., Zhang, L., Luo, R., Qu, Q., Huang, W., Wang, J., et al. (2014). Integrating computer-assisted learning into a regular curriculum: evidence from a randomized experiment in rural schools in Shaanxi. *Journal of Development Effectiveness*, 6(3), 300-323.

[4] Lai, F., Zhang, L., Qu, Q., Hu, X., Shi, Y., Boswell, M., et al. (2012). Does computer-assisted learning improve learning outcomes? Evidence from a randomized experiment in public schools in rural minority areas in Qinghai, China. REAP working paper.

[5] Lai, F., Zhang, L., Hu, X., Qu, Q., Shi, Y., Qiao, Y., et al. (2013). Computer assisted learning as extracurricular tutor? Evidence from a randomized experiment in rural boarding schools in Shaanxi. *Journal of Development Effectiveness*, 5(2), 208-231.

[6] Mo, D., Zhang, L.,Wang, J., Huang,W., Shi, Y., Boswell, M., et al. (2014). The persistence of gains in learning from computer assisted learning (CAL): Evidence from a

randomized experiment in rural schools in Shaanxi province in China. REAP working paper.

Response time analytics for evaluating the quality of survey data from school-age students

Leo A. Siiman, Margus Pedaste
University of Tartu, Ülikooli 18, Tartu 50090, Estonia, leo.siiman@ut.ee

Abstract

We investigated the use of response time data to evaluate the quality of data collected in a survey conducted electronically with 2649 sixth grade (M = 13.23 years old, SD = 0.49) and 803 ninth grade (M = 16.22 years old, SD = 0.36) students. The results indicate that although response time data did not prove to be useful in detecting response bias, it did provide some evidence that survey fatigue was not of substantial concern in the collected data. Overall, the quality of data collected in this survey was found to be of high quality.

1. Introduction

Surveys of young people's attitudes and self-reported behaviours are important means of gathering information about how they perceive themselves and the world around them. The large-scale educational assessment Programme for International Student Assessment (PISA) includes a student questionnaire section, in addition to the main sections testing subject matter competency, in order to explore connections between academic achievement and non-cognitive factors such as students' contextual characteristics, their attitudes about school and their approaches to learning [1]. Helping understand why and how students achieve certain levels of academic performance is the main reason the PISA assessment includes the student questionnaire section [2].

However, collecting valid data from young people is sometimes challenging because of response bias. Response bias is a systematic tendency for a participant to respond inaccurately or untruthfully. Response bias may be due to low motivation. For example, Silm, Must and Täht [3] found that test-taking motivation explains about 50% of variance in mental ability test scores. Osborne and Blanchard [4] reported that up to 40% of a sample of 540 middle school students engaged in random responding on a low-stakes test. In a non-testing taking situation, Morbitzer, Spröber, and Hautzinger [5] found that 31% of responses to a survey about bullying conducted with 208 students aged 10-12 years old (*M* = 11.03, *SD* = 0.51) were classified as "unreliable/invalid." Cornell et al. [6] identified 12% (918 of 7801) of students aged 10 to 19 years of age (*M* = 14, *SD* = 2.12) as invalid responders on a survey of bullying.

To ensure the quality of survey data, researchers have proposed various data screening techniques to detect and remove invalid responses [7-10]. Wise and Kong [11] found that response time data from a computer-based test correlated with examinee test performance and self-reported effort. A review of using response time in educational testing by Lee and Chen [12] showed that response time can be a useful parameter to differentiate between valid and invalid response behaviour.

In the present study we focused on applying response time analytics to evaluate the quality of survey data collected from school-age students. We wanted to examine whether fast response times are more likely to be associated with inconsistent responding (i.e. response bias), and whether similar items placed at the beginning and the end of the survey show substantially different response times, thereby indicating potential survey fatigue.

2. Method

2.1. Sample and procedure

A total of 3452 students from 146 schools in Estonia in the 6^{th} (*n* = 2649, 48.3% boys, 51.7% girls) and 9^{th} (*n* = 803, 47.3% boys, 52.7% girls) grades completed a survey of smart device usage and about other non-cognitive factors related to learning. The average age of 6^{th} graders was 13.23 years old (*SD* = 0.49) and the average age of 9^{th} graders was 16.22 years old (*SD* = 0.36). Students were informed that the survey was voluntary and that data confidentiality would be maintained. A signed parental consent form was also required for participation.

The survey instrument was administered electronically at schools using school computers with an internet connection or using tablet computers provided by a data collector responsible for implementing the survey at a school. Students typically completed the survey during regular school hours within the time constraints of a 45 minute class period.

2.2. Survey instrument

An electronic questionnaire was developed to survey students' use of smart devices for performing digitally competent activities [13], as well as their attitudes, interests, motivations, beliefs and behaviours related to learning in different contexts.

The questionnaire did not present all the items at once to students, but instead required students to answer a few items and then click a button to continue answering new items on a new page. Each time a student clicked the "next" button the response time was calculated as the total amount of time spent on that page (i.e. the time when the button was clicked on the current page minus the time when the button was clicked on the previous page). The number of items presented on a page was determined so that tablet computer users would not have to unnecessarily scroll down to respond to the items.

One construct the survey measured was students' intrinsic motivation in four different learning contexts: learning in general, learning with smart devices, learning mathematics, and learning science. The intrinsic motivation construct was measured using four items selected from the Intrinsic Motivation Inventory [14]. Students were asked to indicate their level of agreement on a five-point Likert scale (1 = disagree, 2 = somewhat disagree, 3 = neither agree nor disagree, 4 = somewhat agree, 5 = agree) to the following statements:

- I enjoy doing [*this activity*] very much.
- [*This activity*] is fun to do.
- [*This activity*] is boring.
- [*This activity*] is very interesting.

This set of intrinsic motivation items appeared at different locations in the survey, each time with the activity under study referring to a different learning context. The regular use of these intrinsic motivation items in different parts of the survey inclined us to use them as a check of respondent consistency, as will be explained in the next subsection. Moreover, since the structure of these items remained the same in different parts of the survey, it was assumed that students should take about the same amount of time to respond to them. A substantial difference in response time would suggest a discrepancy and potential response bias.

2.3. Measure of response bias

In order to identify response bias we chose to use the inter-item standard deviation (ISD) as a measure of within-person consistency. Marjanovic et al. [15] proposed the ISD as a statistic to detect random responding and showed that it was effective in identifying inconsistent responders on a five-item Conscientious Responders Scale (CRS). The authors defined ISD as "identical to the unbiased estimate of standard deviation except that a single respondent's mean score is used in the place of the group's means score." [15, p. 80].

In the current study, the inter-item standard deviation was calculated from responses to items measuring intrinsic motivation. These items came from the well-known Intrinsic Motivation Inventory and research confirms that it generates reliable results [16]. It is also noted that the third intrinsic motivation item selected is reversely worded and requires opposite scoring for data analysis. For example, a respondent who carelessly agreed to all four intrinsic motivation items would generate an ISD value of 2.0, whereas a perfectly consistent value of 0.0 would have required the respondent to disagree with the third item. It is also useful to note that according to probability theory, randomly selecting values to the four items results in an expected ISD value of 1.34. This threshold value is used to estimate the amount of overall random responses in the collected data.

Because the set of intrinsic motivation items appeared in four different parts of the questionnaire, an aggregate ISD value was calculated as an average of the four individual ISDs.

2.4. Measure of survey fatigue

Long surveys can be a possible source of fatigue and response burden for participants. Wise and Kong [11] showed that response time data can indicate if respondents become potentially fatigued and start to abruptly respond at a much rapid rate than their previous pattern of responding.

In the current study, survey fatigue was evaluated by looking at the average response time per item for each set of intrinsic motivation items. An abrupt decrease in response times would be a sign of potential survey fatigue and a cause of concern for the quality of the data collected. In addition, the internal reliability (Cronbach's alpha) of the set of intrinsic motivation items from different sections of the questionnaire was calculated to compare if it abruptly declined in later parts of the questionnaire.

3. Results

3.1. Response time and response bias

The mean survey completion time (i.e. the sum of response times from all questionnaire items) for 6^{th} grade students was 31.8 minutes (SD = 8.9) and 27.1 minutes (SD = 8.8) for 9^{th} grade students. Figure 1 presents scatter plot results of the aggregate inter-item standard deviation versus survey completion time for students in both grades. In both cases no correlation was found between the aggregate ISD and survey completion time (r = -.062 for 6^{th} graders, r = -.002 for 9^{th} graders).

The threshold ISD value for completely random responding was calculated to be 1.34, and in the collected data the percentage of aggregate ISD values above 1.34 was only 1.13% for 6^{th} graders and 0.12% for 9^{th} graders, indicating a very low overall response bias in the collected data.

3.2. Response time and survey fatigue

Table 1 presents the average response time per item for each set of intrinsic motivation items for 6^{th} and 9^{th} graders.

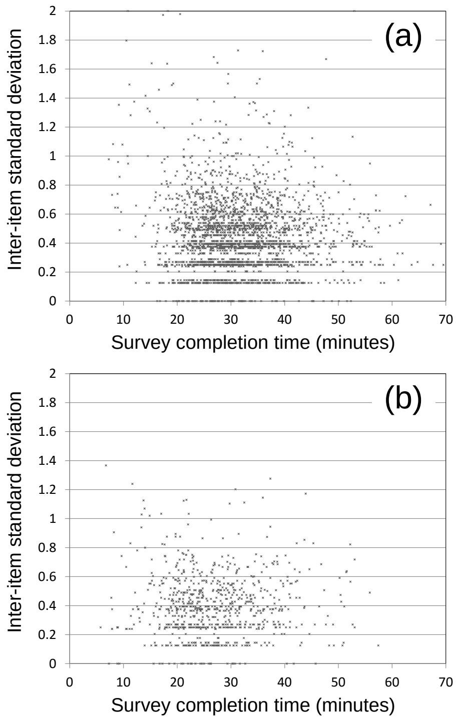


Figure 1. Scatter plots of the aggregate inter-item standard deviation versus survey completion time for (a) 6th graders, and (b) 9th graders.

It can be seen from Table 1 that although the average response time declines over time, there are no abrupt changes in the average response time when comparing adjacent sets of intrinsic motivation items. Furthermore, the lowest average response times (5.46 seconds for 6th graders and 4.59 seconds for 9th graders) are still above the 1-2 second response time levels characterized by Wise and Kong [11] to be indicative of rapid guessing.

To confirm that student responses to the set of intrinsic motivation items were consistent in different parts of the questionnaire, the internal consistency (Cronbach's alpha) was calculated. Table 2 shows that values of Cronbach alpha for both 6th and 9th grade students remained high for the set of intrinsic motivation items located in different parts of the questionnaire.

Table 1. Average response time per item for each set of intrinsic motivation items located in different parts of the questionnaire.

Starting location of the set of intrinsic motivation items	Average response time per item (in seconds)			
	6th grade		9th grade	
	M	*SD*	*M*	*SD*
Page 3 of 60	8.35	4.71	6.93	4.39
Page 17 of 60	7.62	3.11	6.67	3.05
Page 35 of 60	6.21	2.46	5.40	3.45
Page 43 of 60	5.46	2.32	4.59	2.36

Table 2. Internal consistency (Cronbach's alpha) of the set of intrinsic motivation items located in different parts of the questionnaire.

Starting location of the set of intrinsic motivation items	Cronbach's alpha	
	6th grade	9th grade
Page 3 of 60	.896	.885
Page 17 of 60	.867	.862
Page 35 of 60	.928	.916
Page 43 of 60	.925	.924

4. Discussion and conclusion

Our results showed that response bias did not correlate to response time. This means that fast response times were no more likely to be associated with inconsistent responding than slow response times. Thus, this result indicates caution when applying data screening techniques based solely on response time data. Lee and Jia [17], in the context of interpreting potentially invalid responses on educational tests, warn that "Short RTs [*response times*] alone do not indicate rapid guesses because students with pre-knowledge on items may have short RTs but high accuracy. However, long RTs might result from no engagement with the items and distraction by unrelated activities, rather than high engagement." [17, p. 22]. In self-report surveys, some individuals may simply demonstrate various speeds of reading skills and familiarity with survey items. Thus, it appears that response time by itself is not a sufficient indicator to detect response bias.

In terms of evaluating survey fatigue, the results of a set of similar items located in different parts of the questionnaire showed no abrupt response time differences between adjacent item sets, and that response times remained above levels indicative of rapid guessing. Thus substantial survey fatigue was not evident in the studied dataset.

In conclusion, we studied the use of response time analytics to evaluate the quality of survey data from school-age students. Although response time data did not prove to be useful in detecting response bias, it did provide some evidence that survey fatigue was not of substantial concern in this dataset. Further investigation of how response time varies with survey conditions may provide additional information useful for evaluating the quality of survey data.

5. Acknowledgements

This study was funded by the Estonian Research Council through the institutional research funding project "Smart technologies and digital literacy in promoting a change of learning" (Grant Agreement No. IUT34-6).

6. References

[1] Programme for International Student Assessment (PISA). (n.d.). *PISA FAQ*. Retrieved from https://www.oecd.org/pisa/aboutpisa/pisafaq.htm (accessed June 29, 2016).

[2] Organisation for Economic Co-Operation and Development (OECD). (2014). Context questionnaire development. In *OECD, PISA 2012 Technical Report.* Paris: OECD Publishing.

[3] Silm, G., Must, O., Täht, K. (2016). What accounts for low-stakes testing results? Test-taking motivation explains about 50% of the variance in scores. Manuscript submitted for publication.

[4] Osborne, J. W., & Blanchard, M. R. (2011). Random responding from participants is a threat to the validity of social science research results. *Frontiers in Psychology, 1*, 1-7.

[5] Morbitzer, P., Spröber, N., & Hautzinger, M. (2009). Wie zuverlässig sind Selbsteinschätzungen von Schülern zum Vorkommen von Bullying? [How valid are student self-reports of bullying in schools?]. *Prax Kinderpsychol Kinderpsychiatr, 58*(2), 81-95.

[6] Cornell, D., Klein, J., Konold, T., & Huang, F. (2012). Effects of validity screening items on adolescent survey data. *Psychological Assessment, 24*(1), 21-35.

[7] Meade, A. W. & Bartholomew Craig, S. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437-455.

[8] Curran, P. (in press). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*. [Available online 1 August 2015].

[9] Huang, J.L., Bowling, N.A., Liu, M., & Li, Y. (2014). Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions. *Journal of Business and Psychology, 30*, 299–311.

[10] Maniaci, M.R., & Rogge, R.D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality, 48*, 61–83.

[11] Wise, S., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18*(2), 163-183.

[12] Lee, Y-H., & Chen, H. (2011). A review of recent response-time analyses in educational testing. *Psychological Test and Assessment Modeling, 53*(3), 359-379.

[13] Siiman, L. A., Mäeots, M., Pedaste, M., Simons, R. J., Leijen, Ä., Rannikmäe, M., Võsu, K., & Timm, M. (in press). An instrument for measuring students' perceived digital competence according to the DIGCOMP framework. In *Proceedings of the 18 International Conference on Human-Computer Interaction, HCI International 2016, Toronto, Canada, July 17-22, 2016*. Springer International Publishing.

[14] Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology, 43,* 450-461.

[15] Marjanovic, Z., Holden, R., Struthers, W., Cribbie, R., & Greenglass, E. The inter-item standard deivation (ISD): An index that discriminates between conscientious and random responders. *Personality and Individual Differences, 84*, 79-83.

[16] McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: a confirmatory factor analysis. *Research Quarterly Exercise and Sport, 60*(1), 48-58.

[17] Lee, Y-H., & Jia, Y. (2014). Using response time to investigate students' test-taking behaviors in a NAEP computer-based study. *Large-scale Assessments in Education, 2*(1), 1-24.

Implications of the European Data Protection Regulations for Learning Analytics Design

Tore Hoel & Weiqin Chen
Oslo and Akershus University College of Applied Sciences,
{tore.hoel,weiqin.chen@hioa.no}

Abstract

By May 2018 the new European Union's data protection reform will become enforced in most European countries. The reform claims to ensure that the citizen receives clear and understandable information when his or her personal data is processed. The new rules also strengthen the individual's rights to be forgotten. Learning analytics (LA) systems must operate within the confines of the law. However, there is not much evidence in the LA research literature that data protection and privacy constraints have played an important role in systems designs. This paper outlines the new data protection requirements put in place in Europe and investigates the implications for the design of learning analytics systems.

1. Introduction

Is privacy a show-stopper for learning analytics (LA)? This question is the title of a recent review report from the European LACE project [1]. The report, however, does not give a definite answer to the question given the complexity of the panorama of issues presented by the new and emerging LA technologies and practices. The technical environments are increasingly complex, sometimes putting the users in control of managing their data, sometimes keeping them completely out for the loop. The range of data sources involved makes learning analytics enmeshed with multiple personal and societal issues that are not yet fully analysed. What makes the situation even more convoluted is the fact that the rise of learning analytics "is not presented simply as a more effective way to carry out educational activities, but also as a means to transform the context in which the new methods are embedded" [1]. In this situation educational authorities may be tempted to ask for a time-out in order to sort out legal and other issues related to large-scale implementation of LA.

Maybe it was the emergency break that was pulled when a school agency, The Norwegian Centre for ICT in Education, concluded that most probably the application of LA would be against the law unless a number of principles were not adhered to [2]. In the guidelines from the Centre, the principles of data protection were summarised under the headlines of "lawfulness, purpose limitation, data minimisation, data quality, storing and deletion, right to know what information is stored, and information safety". These principles are derived from the Norwegian *Personal Data Act* of April 2000 [3], which in turn builds on the *European Data Protection Directive* (Directive 95/46/EC). The text of the Act describes these principles succinctly: The data controller shall ensure that personal data are processed only if the data subject has consented; "b) are used only for explicitly stated purposes that are objectively justified by the activities of the controller, c) are not used subsequently for purposes that are incompatible with the original purpose of the collection, without the consent of the data subject, d) are adequate, relevant and not excessive in relation to the purpose of the processing, and e) are accurate and up-to-date, and are not stored longer than is necessary for the purpose of the processing" [3].

What is interesting to note in the guidelines from the Norwegian school agency is the interpretation of the principles in the *Personal Data Act*, which lead to the assumption that "the school owner is not able to maintain the most important principle of data protection: The data subject should be in control of and agree to how their own data are used" [2]. The agency asks "how will the school owner make sure that information only are used for learning and not for other purposes, for example to control pupils and teachers? (..) What is the boundaries between information that are relevant for learning and information that are not relevant, but nevertheless are of interest for registration and analysis" [2]. It is evident to anyone in the field of LA, that there is no clear answer to these questions.

What makes it even more necessary to explore how the emergent LA practices are grounded in current laws is the new European data protection reform that will become law by May 2018. According to a factsheet from the European Commission (EC), the "new General Data Protection Regulation (GDPR) will ensure that you receive clear and understandable information when your personal data is processed. Whenever your consent is required, it will have to be given by means of a clear affirmative action before a company can process your personal data. The new rules will also strengthen individuals' right to be forgotten, which means that if you no longer want your personal data to be processed, and there is no legitimate reason for a company to keep it, the data shall be deleted" [4].

This paper raises the question of how GDPR will influence the design of LA systems and practices. Based on a review of LA research literature, the new data protection regulations, and current design efforts the authors develop requirements for a development of LA based on the principles of "data protection by design" and "data protection by default".

2. Related Work

A 2015 survey of European citizens' attitudes to data protection [5] concluded that only 15% felt they had complete control over the information they provided online; one in three people (31%) thought they had no control over it at all. Nine out of ten Europeans expressed concern about mobile apps collecting their data without their consent, and seven out of ten worried about the potential use that companies may make of the information disclosed. Given this massive concern about data protection it should be noted that data protection did not appear in the proceedings of the main conference of the LA research community[1] in 2014 and 2015, and only one time in 2016 [6]. This might imply that the principles of Data protection by design and by default inscribed in the GDPR would have a way to go in order to influence LA design and practices. However, other research and community exchange have put the issues of ethics, privacy and data protection on the international agenda.

In two papers Mason, Chen and Hoel [7, 8] have found that issues related to ethics and privacy are on the top of the list of concerns that researchers and practitioners in the emergent field of LA want to be addressed. “Examples of some of the major questions are related to the ownership and protection of personal data, data sharing and access, ethical use of data, and ethical implications of the use of learning analytics in education”, observes the editors of Journal of Learning Analytics, which featured a special issue on ethics and privacy in 2016 [9]. In a guest editorial [10] Ferguson et al. examined the learning analytics challenges with ethical dimension identified within this special issue. They found 21 challenges, of which six related to the *duty to act*; one addressed *informed consent*; three concerned *safeguarding individuals’ interests and rights*; two were about *equal access to education* and a *just society*; seven dealt with *data protection*; and the last two were related to the *privacy* as socio-cultural concept.

A review of recent literature on ethics and privacy confirms the identified gap between concerns and challenges and proposals for design to address these issues.

In a number of papers Hoel and Chen [11, 12, 13] have explored what technical solutions a privacy-driven design of LA might lead to. In [14] Hoel, Cho and Chen researched how privacy and data protection requirements would affect all processes of the LA cycle.

In analysing the design implications of the GDPR we will use the same process model (Figure 1) as in [14]. The research question is to do a first exploration of how GDPR will influence design of the different processes of LA.

3. GDPR – new regulations for the digital age

European Union legislation on data protection has been in place since 1995. The objectives and principles of this directive (95/46/EC) remain sound, according to the new GDPR [15]. The existing laws were, however, drafted before the advent of cloud computing, social networking sites, location-based services and smartphones, so there was a need to update the laws to "make sure people's right to personal data protection (…) remains effective in the digital age". The aim was "to give people more control over their personal data and make it easier to access it" [16].

These are key changes of the new regulations [15], as the European Commission explains it[2] from the point of view of the citizens:

Consent for processing data: When your consent is required, you must be asked to give it by means of a clear affirmative action. More transparency about how your data is handled, with easy-to-understand information, especially for children.

Easy access to your own data: Free and easy access to your personal data, making it easier for you to see what personal information about you is held by companies and public authorities, and making it easier for you to transfer your personal data between service providers (data portability).

Data breaches: Without undue delay you have the right to know when your data has been unrightfully accessed or hacked.

Right to be forgotten: If you no longer want your personal data to be processed, and there is no legitimate reason for an organisation to keep it, it must be removed from their system. Data controllers must prove that they need to keep the data rather than you having to prove that collecting your data is not necessary.

Data protection by design and Data protection by default: Data protection safeguards should be built into products and services from the earliest stage of

[1] The Learning Analytics & Knowledge conferences organised by the Society for Learning Analytics Research

[2] Factsheets, directives, and regulation of EU reform on Data Protection is found at http://ec.europa.eu/justice/data-protection/reform/index_en.htm

development; the default settings should be those that provide the most privacy. Companies will be obliged to inform you as clearly, understandably and transparently as possible about how your personal data will be used, so that you are in the best position to decide what data you share. This information may be provided in combination with easy to understand standardised icons.

The GDPR does not concern processing of anonymous data. However, real anonymisation (removing personally identifiable information where it is not needed) is hard to achieve, therefore techniques of pseudonymisation (replacing personally identifiable material with artificial identifiers), and encryption (encoding messages so only those authorised can read it) are often applied, and encouraged in the GDPR, to protect personal data. According to the EC this will encourage the use of "big data" analytics, which can done using anonymised or pseudonymised data [15]. However, pseudonymisation or encryption do not exclude any data processor to conform to the GDPR.

4. Requirements for LA systems

Figure 1 depicts the main processes of a LA system [14] as been specified in an ISO/IEC standard (ISO/IEC 20748:2016). All processes are affected by the GDPR, which foresees that the principle of data protection by design and by default will be effective tools to create technological and organisational solutions [16]. In the following we will go through the LA processes one by one to solicit requirements from the new GDPR for system design.

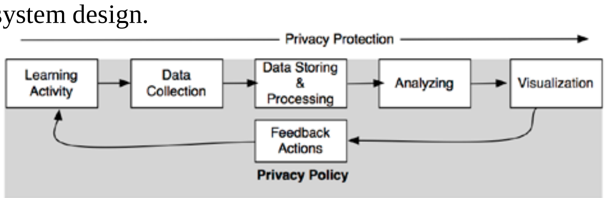


Figure 1. Learning Analytics processes [14].

Learning Activity: This is the process where the LA is set up, aims are decided, metrics for data collection defined, etc. GDPR requires that a dedicated conversation is set up between the LA processor and the student, parent, teacher or any other user of the system so that the data subject can "be informed of the existence of the processing operation and its purposes" [15, recital 60]. The information provided should take into account "the specific circumstances and context in which the personal data are processed. Furthermore, the data subject should be informed of the existence of profiling and the consequences of such profiling" [15, recital 60]. Given that context qualifies most things then pedagogical approach and the use of predictive models (profiling) should be part of this conversation. It is clear that this cannot be left to the machine to communicate alone; however, the outcome of such discussions must be documented by the system.

Data Collection: The process of gathering and measuring information on variables of interest is dependent upon consent by the learner. It is not enough to take the consent for granted when the learner enrolls in a course. GDPR requires "a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data relating to him or her. (..) Silence, pre-ticked boxes or inactivity should not (..) constitute consent [15, recital 32]. This requirement will lead to development of unprecedented software [12], as the consent should cover every purpose set out for the data collection.

Data Storing & Processing: This is the process of preparing and storing data from heterogeneous sources for transport and preparation to data analysis. In designing this process developers must develop mechanisms that allow the data subjects "to request and, if applicable, obtain, free of charge, in particular, access to and rectification or erasure of personal data" [15, recital 59]. In addition the end-users of the LA system have the right to object to information collected, and there need to be some provisioning of their exercise of this right. Again, there needs to be put in place log systems that will record actions taken for data protection audits. The GDPR enforces the principles of fair and transparent processing. The regulation introduces the idea of "standardised icons in order to give in an easily visible, intelligible and clearly legible manner, a meaningful overview of the intended processing. Where the icons are presented electronically, they should be machine-readable" [15, recital 60]. This provides requirements for a personal LA data monitoring tool, which gives the learner the opportunity to check at any time what data is gathered and stored about him or her, and to initiate actions. These actions could be rectification or erasure of personal data, or download their full personal dataset for storage in a personal data record store, or for transmission to another LA provider [15, recital 68].

The right to be forgotten gives the data subject the right to have his or her personal data erased and no longer processed "where the personal data are no longer necessary in relation to the purposes for which they are collected or otherwise processed, [or] where a data subject has withdrawn his or her consent or objects to the processing of personal data concerning him or her" [15, recital 65]. The system needs to communicate with other systems that have "links to, or copies or replications of those personal data" [15, recital 66]. However, this right is not absolute, and the system should be able to "handle conflicts", e.g., "temporarily moving the selected data to another processing system, making the selected personal data unavailable to users, or temporarily removing published data from a website" [15, recital 67].

In this LA process steps should be taken towards pseudonymisation and/or encryption of personal data. The GDPR encourages use of measures of pseudonymisation [15, recital 29]. The solution needs to specify how the keys, the "additional information for attributing the personal data to a specific data subject" [15, recital 29], are managed either within the organisation running the LA or outside. There is also a need to set up procedures for risk evaluation to establish "whether data processing involve a risk or a high risk" [15, recital 76].

Analysing: The process of systematic examination of learning data in order to extract descriptive and possibly predictive knowledge about the learners and their contexts is the core of a LA system. This is a kind of profiling. GDPR defines profiling as "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements" [15, Article 4 (4)]. The models used for analysis are often impenetrable for the end-user; the GDPR, however, give clear incentives for the LA system developers to give "meaningful information about the logic involved" [15, Article 13, (f, g, h)]. It would be interesting to find out how far the learner's rights to information about algorithms and predictive models in LA system go with this new European law. The strong requirement gleaned from a design perspective is the need to plan for transparency and possibly some kind of open sharing of these technologies.

Visualisation: This is the process of interpreting and presenting the analysis result of LA data in a (mainly) visual form that contributes to the understanding of the meaning of the data. All the general requirements of the GDPR about transparent information, communication and modalities for the exercise of the rights of the data subject come into play here. However, we don't find that GDPR gives any LA specific requirement for this process.

Feedback Actions: Such actions serve the results of a cycle of learning analysis back to the learners and their contexts so that corrective actions can be taken. The GDPR regulates that the learner or the teachers "should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing" [15, recital 71]. The automatic learning machine does not seem to be favoured by the GDPR; on the contrary, the regulations provide a strong incentive to design for human mediation in feeding back the results from LA. The design must give the learners a possibility to object to being targeted as a specific type of learner, if based predominantly on invisible statistical inferences. The principle of profile transparency gives the learners a right to be informed about how they are anticipated.

However, the GDPR goes further: If there is any automated individual decision-making, including profiling, the data subject have the right to be informed about "the significance and the envisaged consequences of such processing" [15, Article 14, g]. In order to establish some common ground for discussing envisaged consequences, the strict minimum period for which personal data could be stored [15, recital 39], specific circumstances and contexts in which personal data are processes [15, recital 60], etc. it is clear that the system design coming out of the GDPR requirements point in the direction of open systems with extensive conversational capabilities.

5. Discussion

Big data – the computational analysis of large dataset to reveal patterns, trends, and associations – can be used for many good purposes, and education is one of them. Big data entails a new way of looking at data, where data are assigned value in itself. The value of the data lies in its potential future uses. Big Data's business model is the antithesis of data minimisation and purpose limitation, which are key principles of privacy protection [16]. Previously, the data protections risks could be made to go away, almost by definition, by applying anonymisation of personal data. With big data new challenges arise with the risk of re-identification, which makes anonymisation less effective as a method for preventing the privacy disadvantages associated with profiling and other data analysis.

The GDPR is meant to strengthen data protecting while making it easier to run digital enterprises. For education it makes it necessary to go back and ask what the core values and ideas of teaching and learning as a business. It is not only about selling a course, but to build capabilities together with the students in a dialogue that is more complex than any consumer store transaction. This invites to more openness and transparency on what goes on with the use of learners' data for analytics. It also opens up for a dialogue with the learner that brings other principles of GDPR into play, e.g., the principle of proportionality, the principle of fair and transparent processing, etc. If data is collected for one purpose, it does not automatically prohibit processing for a different purpose or restrict raw data for use in analytics. "A key factor in deciding whether a new purpose is incompatible with the original purpose is whether it is fair. Fairness will consider factors such as; the effects on the privacy of individuals (e.g. specific and targeted decisions about identified persons) and whether an individual has a reasonable expectation that their personal data will be used in the new way" [17].

In the requirements this paper has solicited from GPDR (Section 4) a strong message comes through, and that is the need to involve the end-users of the systems in a dialogue that encompasses all the different processes of LA systems. This is not only a question of design of the LA system as such, but also concerning the execution of each individual process. It will not be possible to enter the discussion about the use of LA system only at enrollment and agree upon conditions of use by presented a list of check boxes. The negotiation about use of personal data and their analysis has to be a continuous process that will need tools that still have to be developed. If these tools are provided, there should not be limits to what could be achieved by LA. As the EC explains "companies are free to base processing on a contract, on a law or, on, in the absence of other bases, on a "balancing of interests". These 'formal requirements', such as consent, are set out in the rules to provide the necessary control by individuals over their personal data and to provide legal certainty for everyone" [17].

6. Conclusions and further work

This paper has explored what requirements for LA design come out of the new legislation passed in the European Union regarding protection of personal data and free movement of such data. Even before the GDPR was known there has been a considerable uncertainty to whether emergent LA practices were according to national and European data protection laws. GDPR maintain the principles from the 1995 directive, but clarifies a number of requirements that have to be built into LA architectures and systems now being designed. Openness, transparency and continuous negotiation between data subjects and data processors (i.e., school owners, universities, and vendors) are the principles the authors of this paper would highlight as the take away for further research and development.

The paper is a first exploration of how new European legislation will impact LA system design. Further research is now necessary to identify how other jurisdictions might deliver requirements for privacy and data protection in LA systems, and how the global market will be influenced by this European legislation.

7. References

[1] Griffiths, D., Drachsler,H., Kickmeier-Rust, M., Steiner, C., Hoel, T., Greller, W. (2016). Is Privacy a Show-stopper for Learning Analytics? A Review of Current Issues and Solutions. Learning Analytics Review 6. Published by the LACE project. ISSN:2057-7494 http://www.laceproject.eu/learning-analytics-review/privacy-show-stopper

[2] Dahl, M. (no date). Læringsanalyse. Memorandum published at the Norwegian Centre for ICT in Education website http://iktsenteret.no/ressurser/notat-laeringsanalyse. Accessed 2016-06-01

[3] Datatilsynet (n/d) Personal Data Act. Act of 14 April 2000 No. 31 relating to the processing of personal data (Personal Data Act) Online at https://www.datatilsynet.no/English/Regulations/Personal-Data-Act-/ Accessed: 2016-06-01

[4] European Commission (2016). How does the data protection reform strengthen citizens' rights? Online http://ec.europa.eu/justice/data-protection/document/factsheets_2016/factsheet_dp_reform_citizens_rights_2016_en.pdf

[5] Special Eurobarometer 431, Data protection: http://ec.europa.eu/public_opinion/archives/eb_special_439_420_en.htm#431.

[6] Drachsler, H., & Greller, W. (2016). Privacy and Learning Analytics – it's a DELICATE issue. LAK '16, April 20-24, 2016, Edinburgh, UK, 1 ACM 978-1-4503-3417-4/15/03.

[7] Mason, J., Chen, W., & Hoel, T. (2016). Questions as data: illuminating the potential of learning analytics through questioning an emergent field. Research and Practice in Technology Enhanced Learning, 1–14. http://doi.org/10.1186/s41039-016-0037-1

[8] Hoel, T., Mason, J. & Chen, W. (2015) Data Sharing for Learning Analytics – Questioning the Risks and Benefits. In Ogata, H. et al. (Eds.) (2015). Proceedings of the 23rd International Conference on Computers in Education. China: Asia-Pacific Society for Computers in Education

[9] Gasevic, D., Dawson, S., & Jovanovic, J. (2016). Ethics and privacy as enablers of learning analytics. Journal of Learning Analytics, 3(1), 1–4. http://doi.org/10.18608/jla.2016.31.1

[10] Ferguson, R., Hoel, T., Scheffel, M., & Drachsler, H. (2016). Guest editorial: Ethics and privacy in learning analytics. Journal of Learning Analytics, 3(1), 5–15. http://doi.org/10.18608/jla.2016.31.2

[11] Hoel, T. & Chen, W. (2015). Privacy in Learning Analytics – Implications for System Architecture — In Watanabe, T. and Seta, K. (Eds.) (2015). Proceedings of the 11th International Conference on Knowledge Management. ISBN 978-4-9908620-0-8 Presented at ICKM 15 in Osaka, Japan, 4 - 6 November 2015 ickm.kis.osakafu-u.ac.jp

[12] Hoel, T., & Chen, W. (2016). Privacy-driven design of learning analytics applications: Exploring the design space of solutions for data sharing and interoperability. Journal of Learning Analytics, 139–158. http://doi.org/10.18608/jla.2016.31.9

[13] Hoel, T., & Chen, W. (2014). Learning Analytics Interoperability– looking for Low-Hanging Fruits. In Liu, C.-C. et al. (Eds.) Proceedings of the 22nd International Conference on Computers in Education. Japan Asia-Pacific Society for Computers in Education.

[14] Hoel, T., Chen, W., & Cho, Yong-Sang (2016) Privacy Requirements for Learning Analytics – from Policies to Technical Solutions. Paper presented at Workshop on Ethics and Privacy for Learning Analytics, Monday, April 25th 2016 at the 6th International Conference on Learning Analytics and Knowledge (LAK '16), Edinburgh, United Kingdom

[15] REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

[16] Datatilsynet. (2013). Big Data - privacy principles under pressure. https://www.datatilsynet.no/globalassets/global/04_planer_rapporter/big-data-engelsk-web.pdf

[17] European Commission (2015). Questions and Answers - Data protection reform. Online at http://europa.eu/rapid/press-release_MEMO-15-6385_en.pdf. Accessed 2016-06-01

ISBN: 978-82-8364-023-6