

# DATA AND TEXT MINING FOR ANALYSING THE PEDAGOGICAL EFFECTS OF DIVERSE DESIGN PROJECTS

**Takao ITO, Masako SHIN, Keisuke MIYAZAKI and Keiko MATSUMOTO**  
Kanazawa Institute of Technology, Nonoichi, 921-8501, Japan

## ABSTRACT

This paper reports on the results of using data mining and text mining approach to analyse the pedagogical effects of the design and engineering PBL (Project-Based Learning) programmes at Kanazawa Institute of Technology (KIT) in Japan. It is difficult to measure the educational effects of PBL courses compared with the other courses that use paper tests for assessment. In this paper, we introduce questionnaires which include multiple-choice and free description for two courses, Project Design Introduction (PD Intro) and Project Design Implementing and Operating (PDIO). The selective responses collected are analysed using the data mining technique and the text mining is used for analysing the free description responses. The improvement of students' communication skills and report preparation skills is observed. The results show that whether students feel the experiment in the team is pleasant or not affected their self-growth evaluation. Using data mining and text mining is found useful for analysis of the pedagogical effects of the PBL programme.

*Keywords: PBL, Project Design, text mining, data mining, pedagogical effects, assessment*

## 1 INTRODUCTION

Kanazawa Institute of Technology (KIT) applies PBL to its design and engineering programme as Project Design Programme. The programme aims to grow up independently minded, actively engaged engineers. The measurement of the educational effect of using diverse themes in PBL courses themes is difficult compared with the other courses which can be assessed by paper tests. Although we score the achievement by evaluation of the assignments, presentations, reports, and notes, it is hard to measure if students grow up as independently minded, actively engaged engineers. Self-valuation questionnaires are usually carried out in class to evaluate the educational effectiveness regarding the students' growth degree.

The spiral improvements of skills of students have been reported based on the results of the multiple-choice self-valuation questionnaire for two courses, Project Design 1 (PD1) and Project Design 2 (PD2) at KIT [1], [2]. The questionnaire asks about presentation skills, communication skills, reporting skills, literature survey skills, and idea creation skills. Descriptive columns in questionnaires are prepared and teachers can read them and education effects can be judged qualitatively. But they cannot be used as quantitative data for class improvement. For this reason, it is difficult to judge how students feel their growth from such questionnaires.

However, with progress of IT technique, utilization of data mining or text mining is getting more popular. Data mining as the meaningful extraction from big data, such as customers' buying history has already been used in the commercial fields. Text mining is also widely used in order to analyse the customers' claims and demands. The application of data mining or text mining has been reported in education fields [3], [4]. However these cases still require some IT skills, which makes it hard for many teachers in analysing the questionnaires effectively.

In this paper, we first state the education objectives of the PD programme at KIT and questionnaires used for this programme. Questionnaires include multiple-choice and free description for PD Intro and PDIO. Next, we introduce the commercial based data mining and text mining tools which are used to analyse the questionnaires. The selective responses are analysed using the data-mining technique and

the text-mining technique is used for free description sentences. Finally, we show the analysed results using those tools.

## 2 PBL PEDAGOGICAL DATA ACQUISITION

### 2.1 PBL Programme in KIT

Kanazawa Institute of Technology (KIT) applies PBL to the backbone of the curriculum, and names it Project Design (PD) Programme. PD Programme is consisted of five courses as shown in Table 1. They are compulsory for freshmen, sophomores, and seniors in all four colleges at KIT (College of Engineering, College of Informatics and Human Communication, College of Environmental Engineering and Architecture, and College of Bioscience and Chemistry). Each project team in each PD course is consisted of four to six students and sets up their own project theme.

Table 1. Project Design Programme Courses and Main Contents

Grade	Freshman		Sophomore		Senior
Course	PD Introduction	PD 1	PD 2	PDIO	PD 3
Main Contents	Experiment methods	Idea creation	Idea into shape	Verification experiments	Graduation thesis

In PD Intro, students freely set up the experiment theme of a team by themselves. The main objective of this course is to give a chance for students to carry out experiments safely in team towards their established goal by choosing the appropriate equipment for their own experiment plan. PD2 and PDIO are programmed as a set. PDIO is a follow-up step of PD2. In PDIO, students create ideas and make plans to verify the function of ideas created in PD2. They also set up models and verify some of their functions. In 2017, the themes to tackle were 288 in PD Intro, 263 in PD1, and 265 in PD2 and PDIO.

### 2.2 Questionnaires in PBL courses

For each course, questionnaire-based evaluation is carried out in order to measure pedagogical effects in students' growth. In this paper, we introduce the results of data mining and text mining analysis for PD Intro and PDIO courses held in 2017.

#### 2.2.1 Questionnaire for PD Intro

Survey questionnaires were administered in the 15<sup>th</sup> class of PD Intro (each PD course is held in 16 classes) The questionnaires consist of three items; five selective responses asking about activities in PD Intro, self-evaluation of ten capabilities, and one free description asking about one's growth. They are listed in Table 2. 1,504 freshmen students of PD Intro answered the questionnaire. Answers were entered into the excel sheet including descriptions.

Table 2. Questionnaire of PD Intro

No.	Questions	Question type
1-1	Pleasantness of experiments	4 level score
1-2	Effect of note taking	5 categories
1-3	Improved skill in PD Intro	7 categories
1-4	Skill need to improve more	8 categories
1-5	Approach for team assignments	5 categories
2-1	Self-evaluation of communication skill	5 level score 5 can do 3 neutral 1 cannot
2-2	Self-evaluation of teamwork	
2-3	Self-evaluation of scientific perspective	
2-4	Self-evaluation of observation	
2-5	Self-evaluation of creativity	
2-6	Self-evaluation of setup	
2-7	Self-evaluation of drawing ideas	
2-8	Self-evaluation of writing ideas	
2-9	Self-evaluation of collecting information	
2-10	Self-evaluation of homework and review	
3-1	Growth and ambition	Free description

### 2.2.2 Questionnaires for PDIO

Sophomore students were asked to answer web-based questionnaire of PDIO. The questionnaire asks them to refer to their experience in learning PDIO and describe how they would utilise those experiences for future use. There are seven questions in the questionnaire (Table 3).

1,283 students answered the questionnaire. The answers were categorised according to departmental classes. All seven descriptions were then entered into the excel sheet; each description was in each cell.

Table 3. Questionnaire of PDIO

No.	Question
1	What learning activities of project design did you take passionately?
2	What was the most difficult learning activity of the project design course? How did you overcome it?
3	What do you think has been achieved by group activity?
4	What role do you think you played in group activity? To what extent do you think you have contributed?
5	What skills and capabilities do you think you have improved? What was the cause of the growth?
6	How do you think you will utilize learning experience for future studies or for future work?
7	Please refer to your experiences of learning activities in two years and describe how you can advise new students to take the courses.

## 3 DATA AND TEXT MINING METHODS

The statistics analysis soft JMP® 11 (SAS Institute Inc., Cary, NC, USA) is used for data mining of the selective answer contents and text-mining results (JMP® is a registered trademark of SAS Institute Inc.) [5]. Text mining software Mieruka-Engine™ is used for analysing the free descriptive answers (Mieruka-Engine™ is a registered trademark of Plus Alpha Consulting Co., Ltd.) [6].

### 3.1 Data Mining

JMP® 11 (SAS Institute Inc., Cary, NC, USA) is the data analysis tool which has many features in exploring data with statistics. It is easy to use for analysing and visualising multivariate data.

The results of selective answer questionnaire of PD Intro were analysed using the data mining. Results of five questions about learning activities and 10 self-evaluation questions of 1,504 students from 14 departments were analysed. To identify the significant activities that might affect the self-evaluation scores, we selected Analyse > Modelling > Response Screening and received the PValues data table. Before screening the responses, the incomplete self-evaluation answers were eliminated because they might affect the analysing methods. The table showed *p*-value for the significance test corresponding to the pair of self-evaluation and learning activities.

The pairs which showed *p*-value < 0.01 were analysed to extract the response items that had major effects in self-evaluation scores. The Fit Y by X platform was used to analyse the pairs.

### 3.2 Text Mining

A Japanese sentence does not have a partition of a word and the same word has a notation in Chinese character, hiragana, katakana and alphabet. Therefore, Japanese peculiar treatment is needed in text mining and requires specialized ICT knowledge. Researchers have used the morphological analysis engine MeCab to extract individual words combined with their customised statistical tools [7], [8]. Text mining has been hard to use for many teachers. However, there are some commercial based text mining tools available that can be used to analyse voices of customers from surveys or social media. In this paper, we use Mieruka-Engine™ for analysing the free descriptive answers. Mieruka-Engine™ claims as No.1 text mining tool in Japan.

It has numerous analysis functions which are easy to use. Word frequency is automatically calculated and the word mapping shows the relationship between words visually. The size of the word circle represents the word frequency. The relationships between words are shown as connecting lines. We can compare word frequency or word mapping using categories like departments or project themes. It also helps to create and save the output results in PowerPoint or Excel format. An example of text mining report is shown in Figure 1.

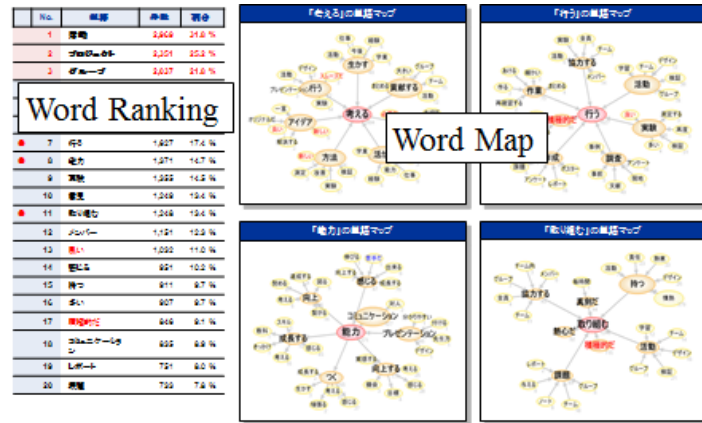


Figure 1. Example of text mining report

## 4 RESULTS AND DISCUSSION

### 4.1 PD Intro

The results of questionnaire of 1-\* and 2-\* in Table 2 were analysed using JMP® 11. The response screening showed strong relations ( $p$ -value < 0.01) between all ten self-evaluation scores and 1-1 (*pleasantness*), seven self-evaluation scores and 1-2 (*note taking*), 2-7 (*drawing ideas*) and 1-3 (*improved skills*), 2-1 (*communication skills*) and 1-5 (*team assignment*), and 2-2 (*team work*) and 1-5. The mosaic plots showing the Contingency Table of 2-2, 2-1, 2-5, 2-4 vs. 1-1 are shown in Figure 2. The proportions on the x-axis represent the number of observations for each level of the 1-1 variable. For each plot, 1-1 axis indicates *pleasantness* of experiments from the *very pleasant* (left) to *not so much* (right). The proportions on the y-axis represent the self-evaluation score of 3 to 5. These plots show that the self-evaluation score is high when students feel *very pleasant* about the experiments.

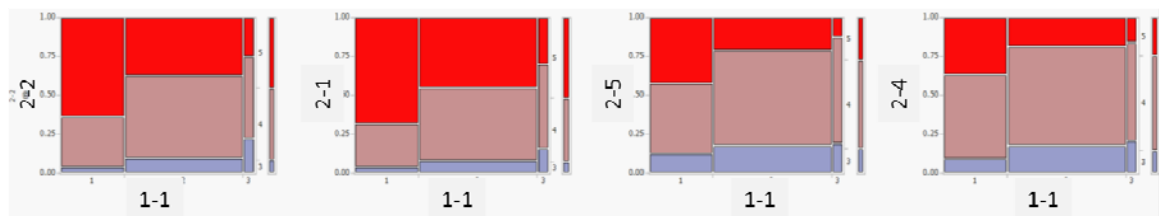


Figure 2. Mosaic Plots of 1-1 vs. 2-2, 1-1 vs. 2-1, 1-1 vs. 2-5 and 1-1 vs. 2-4

In the same way, all of the self-evaluation score is high when students achieve a feeling of accomplishment in note taking. The self-evaluation scores of 2-1 (*communication*) and 2-2 (*teamwork*) are affected by 1-5 (*approach for team assessments*). The scores are high when all members have some contributions to team assessments.

Free description answers of 3-1 (*growth and ambition*) were analysed by the text mining tool. Extracted top ten word counts and word mapping are shown in Figure 3. The words of “team” “cooperate” and “ability” are top three words which share 13.8%, 10.1%, and 9.8% respectively. Word mapping shows a close connection of “team” and “cooperate.” The mapping also shows that “team” is connected with “assignment.” “Ability” is connected with “report”, “communication” and “presentation.”

From the results of the free description analysed by text mining, the relationship between 2-1 (*communication*), 2-2 (*teamwork*) and 1-5 (*team assessments*) is verified. The word mapping shows that students felt the importance of reporting, communication and presentation skills.

These analysing methods may be useful to compare the pedagogical effects when we change the contents of the course.

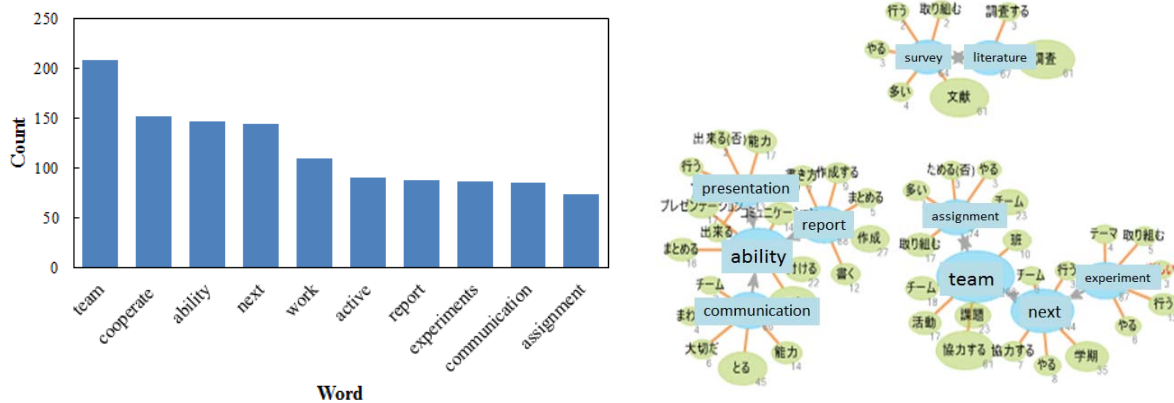


Figure 3. Word count and word mapping of PD Intro

#### 4.2 PD I/O

Free description answers of all seven questions were analysed by the text mining tool. Cross analysis result of extracted top 10 word-ranking in all seven questionnaires is shown in Figure 4. The vertical axis shows department classes; EM as ‘Mechanical Engineering’, EA as ‘Aeronautics’, ER as ‘Robotics’, EEET as ‘Electrical and Electronic Engineering’ and ‘Electronics, Information and Communication Engineering’, EP as ‘Information and Computer Science’, FMFSFP as three departments of College of Information and Human Communication, VA as ‘Architecture’, VS as ‘Architectural Design’, VE as ‘Civil and Environmental Engineering’, and BBBC as two departments of College of Bioscience and Chemistry. The horizontal axis expresses the percentage of the used frequency in top 10 words listed as explanatory notes.

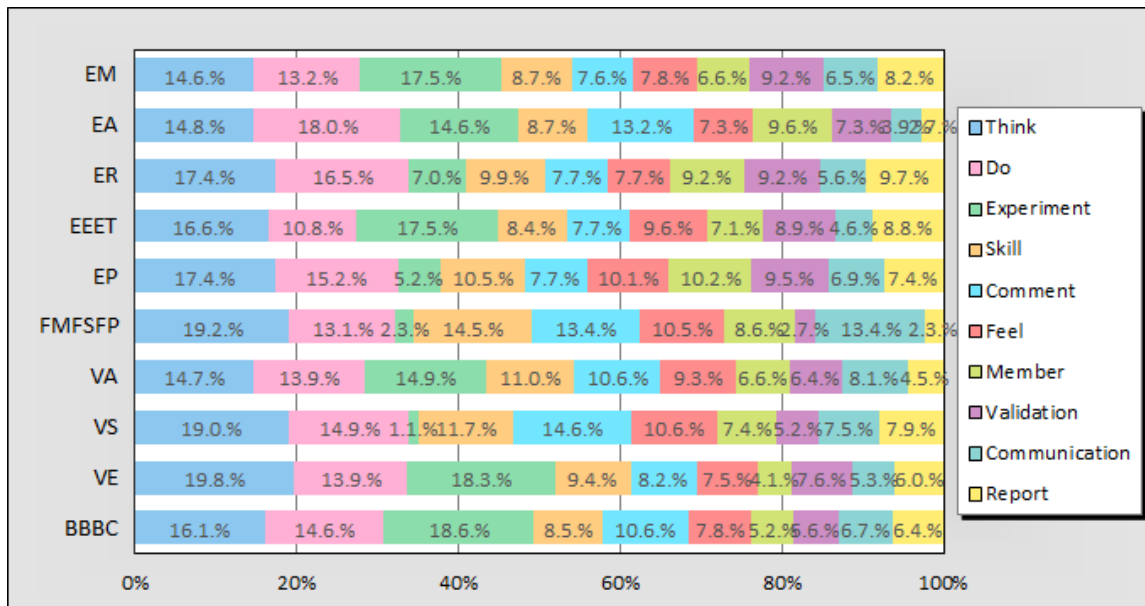


Figure 4. Cross analysis (Department vs. Over All Word Ranking Top 10)

FMFSFP and VS classes show less word counts of “experiment” because these classes focused more on exploration than experiments, or only a few teams tackled experiments. Instead, FMFSFP class focused on presentation and communication, VS class on doing assignments and designing posters. Word mapping in Figure 5 shows difference in answers for No. 5 questionnaire in Table 3 between

these departments. EM students acquire skills in experiments, literature survey and leadership through team work. FMFSFP students improve skills in communication and presentation. VS students feel that the assignment helped to improve their skills in survey and poster presentation.



Figure 5. Word mapping of answers of three departments (No. 5 questionnaire in Table 3.)

1,283 students answered seven questionnaires and there are total of 8,981 free descriptions. It is difficult for instructors to extract pedagogical effects even if they read all of them. The text mining tool helps to extract key words from the descriptions and to compare the difference between departments. The word mapping is useful to judge what skills students acquire and how they grow through the classes.

## 5 SUMMARY

The methods using data mining and text mining have been found useful for analysing the pedagogical effects of the design and engineering PBL programmes. The results of questionnaires conducted to 1,504 students of PD Intro were analysed using the data analysis and text mining tool. Whether students felt the experiment in the team was pleasant or not affected their self-growth evaluation. The growth of communication skill and report preparation skill was observed from the text mining results. 8,981 descriptions in the questionnaires used for PDIO were analysed using the text mining tool. Cross analysis of key words and departments showed difference in acquired skills between departments. The standard statistics analysis soft was used to analyse selective answers and commercial based text mining tool was used to extract key words from free descriptions. Combined use of these tools helped to extract statistically what the students felt about their growth and what kind of educational contents were effective.

## REFERENCES

- [1] Ito T., Shin M., Miyazaki K., Iwata S. and Sentoku E. The Effects of Spiral Educational Method through PBL: KIT Project Design Programme. In *Proc. of the 43rd Annual SEFI Conference*, 2015.
- [2] Ito T., Shin M., Miyazaki K., Iwata S. and Sentoku E. The Project Design Education Collaborating with City Governments and Communities. In *Proc. of the 18th International Conference on Engineering & Product Design Education*, 2016.
- [3] Takechi S. and Rynearson L., Understanding Student Performance Diversity Using Clustering Techniques. In *Proc. of Joint International IGIP-SEFI Annual Conference*, 2010.
- [4] Agrawal R. and Batra M. A Detailed Study on Text Mining Techniques, *IJSCE*, 2013, 2(6), 118-121.
- [5] *JMP*, Available: [https://www.jmp.com/en\\_us/home.html](https://www.jmp.com/en_us/home.html) [Accessed on 2018, 22 Feb.], (2018) 22 Feb.
- [6] *Mieruka-Engine*, Available: <http://www.pa-consul.co.jp/mieruka/en/> [Accessed on 2018, 22 Feb.], (2018) 22 Feb.
- [7] *MeCab*, Available: <https://sourceforge.net/projects/mecab/> [Accessed on 2018, 22 Feb.], (2018) 22 Feb.
- [8] Ohsako S., Kamei Y., Hosoi S., Ishida S., Ubayashi N., Fukuda A. Text Mining for Analyzing the Effect of the Value-Creation Oriented Educational Curriculum, *IPSJ SIG Technical Report*, 2014, 2014-SE-186(8), 1-8.