JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 21, 973-984 (2005)

The Analysis of Learning Records and Learning Effect in Blended e-Learning

DOWMING YEH, CHUN-HSIUNG LEE* AND PEI-CHEN SUN
Institute of Computer and Information Education
National Kaohsiung Normal University
Kaohsiung, 802 Taiwan
E-mail: {dmyeh, sun}@nknucc.nknu.edu.tw
**Department of Information Management*
National Taiwan University of Science and Technology
Taipei, 106 Taiwan
E-mail: D9309105@mail.ntust.edu.tw

Some teachers adopt a blended learning model that combines traditional classroom teaching and an e-learning system. In this model, a teacher may teach the first few sessions in a classroom. After the students have established a general idea of the course, they can then proceed to on-line learning and interaction. This study aimed to discover the relationship between learning records and the learning effect in a blended e-learning environment through multiple regression analysis. The learning records considered included the grades for online assignments, reading time, the total number of login times, and the total number of online discussions. The learning effect was defined as the total grade for two monthly exams and one final exam. To collect learning record data, an e-learning system was designed that integrates the data collection functionality of learning activities with a teaching material managing module so that the learning records of all the learners are recorded automatically. With this system, an experiment was conducted on a program design course in a local high school. The results differed from those obtained in a 'pure' e-learning setting, and the online homework performance was the only item that significantly accounted for the learning effect, which is a natural result of learning procedural knowledge.

Keywords: distance education, e-learning, learning record, learning effect, regression analysis

1. INTRODUCTION

With the rapid development in information technology, information and knowledge can be disseminated faster than ever before. This has brought enormous changes to the traditional roles of teachers and students. The role of a teacher is transformed from that of a leading character to that of a supporter. The teacher now assists students instead of dominating the whole learning process. On the other hand, the style of learning has shifted from passive to active.

Traditional distance education also needs to move toward student-oriented e-Learning [21]. The process of change of distance education is shown in Fig. 1. Although

Received July 1, 2004; revised March 30, 2005; accepted May 18, 2005.
Communicated by Robert Lewis.

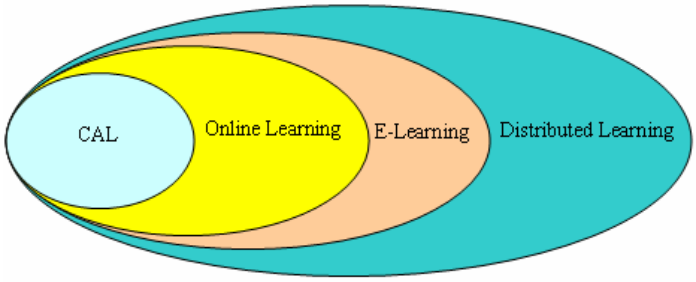

Fig. 1. The process of change of distance education.

distance education uses the Internet to teach, it does not allow interaction between teachers and students, and it cannot keep records of learners' learning [12]. Therefore, we will use the term e-Learning instead of distance education in the rest of this paper.

E-learning, also known as digital learning puts emphasis upon providing simultaneous discussion and collaboration among participants. This is the latest on-line learning method that has appeared in recent years [17]. E-learning systems use network technology and the World Wide Web as their platforms and combine multi-media information, such as audio, video, scripts and graphs [7]. Through hyper-links organized in Web pages, students can browse related materials for information about a given subject. By tracing the history data commonly found in e-learning systems, students can observe their own learning progress. As for teachers, this approach enables them to review and track the learning portfolios of all the students so to achieve the educational goals. The learning portfolio usually consists of the personal records, learning records, and historical records of a student [16].

Some teachers adopt a blended learning model that combines traditional classroom teaching and an e-learning system [18]. In this model, a teacher may teach the first few sessions in a classroom. After the students have established a general idea of the course, they can then proceed to on-line teaching and interaction. Ideally, if we can combine the advantages of classroom teaching and e-Learning, the learning effects will be enhanced and extended in a blended model.

This research investigated the relationship between learning records and learning effects in a blended e-learning environment. The learning records included the duration of reading, on-line assessments, interaction with classmates, uploaded homework, and collaborative learning. A linear regression model was constructed from the learning record data and evaluation data of a class taught by one of the researchers in a blended e-learning setting. Section 2 reviews the related learning theory literature. Section 3 presents the research design, and the results are discussed in section 4. A conclusion is given in section 5.

2. LEARNING THEORY RELATED TO E-LEARNING

The origin of e-learning can be traced back to distance education and Web-Based Instruction (WBI). WBI is actually a course plan that uses the WWW and its resources to

create a meaningful learning environment [8, 20]. The main goal of e-learning is to use the Internet to deliver the course content to learners so that the latter can study via the Internet. The goal is to equip learners with active learning skills and support the learning activities of learners.

E-learning can provide diverse sources of knowledge and rapid learning. It can also lower costs, improve learning quality, and encourage participation. Besides these advantages, e-learning can also teach learners how to be responsible for their own learning effects. E-learning achieves its goals of sharing and delivering knowledge through various media, and it is especially helpful for learner-oriented learning activities. It can increase motivation, accelerate the learning speed, and lower costs. The learning theories underlying e-learning are discussed below.

The goal of constructivism is to pass the core of the teaching procedure from knowledge instructors to learners [14]. Constructivism focuses on the following four points [4]:

(1) Knowledge is learned by learners in an active way.
(2) Knowledge is the rationalization and confirmation of a learner's experience. It is not a memorized facts or truth.
(3) Knowledge is the consensus formed by a learner's interaction with others.
(4) A learner will revise and adjust the knowledge he possesses and integrate it together with new and old experience to develop new knowledge.

The most successful application of this theory to e-learning is the LOGO programming language developed by Seymour Papert [13].

Learners in an e-learning system often require extra support and motivation to engage in unfamiliar tasks. Such needs suggest the use of scaffolding in e-leaning. Wood, Bruner and Ross invented scaffold instruction in 1976 [22]. The main feature is the interaction between instructors and learners. Interacting with others can increase the level of understanding and the ability to solve problems. They listed six tutoring features of scaffolding: "an interactive system of exchange in which the tutor operates with an implicit theory of the learner's acts in order to recruit his attention, reduces degrees of freedom in the task to manageable limits, maintains 'direction' in the problem solving, marks critical features, controls frustration and demonstrates solutions when the learner can recognize them."

Bandura argued that the mental progress of a human being occurs through consecutive interactions with the environment [1]. The outside surroundings have to be subjected to one's cognition process before they affect one's behaviour. An individual has the ability to adapt himself or herself to the surroundings and the use of signifiers. One does not receive excitement passively but actively responds to those excitements, such as by making choices, organizing, exchanging, and giving meanings to them so as to adjust one's behaviour.

The situated theory was first mentioned in 1989 by Brown, Collins and Duguid [2]. The school of task-based learning theory maintains that during the process of forming knowledge, learners are eager and active. They do not only accept excitements from the outside world. They want to dig and discover other useful information hidden in the network of learning resources and build up an individualized cognitive system [9].

A portfolio is a systematic and organized collection of data used to examine a student's knowledge, skills, and attitudes [19]. Portfolio assessment focuses on the collection of multidimensional data in order to understand the learner's learning and changes [3]. Shores and Grace classified portfolios into three types, private portfolios, learning portfolios and pass-along portfolios [16]. A private portfolio consists of background data and historical learning records. A learning portfolio contains learning records made during the learning process, such as material reading hours, formative evaluation, and summary evaluation. A pass-along portfolio combines the above two portfolios to form historical learning records for the next learning level.

3. EXPERIMENTAL DESIGN

Most e-learning systems can only keep basic learning portfolios [5]. The complete data of learning activities are sparsely scattered in web logs. Teachers must perform tedious processing to get a clear picture of the learning activities of his or her students [10]. In fact, to perform such analysis requires certain technical capabilities that can pose a difficulty for many teachers.

To solve this problem, we integrate the data collection functionality of learning activities with the teaching material managing module in our e-learning system to record the learning paths and learning records of all learners simultaneously and automatically. The system is implemented in ASP.NET. Another tool, the learning path viewer, recalls the learning path and records of any specific student. The learning records include the time spent reading online materials, the number of login times, and the number of online discussions. The structure of our e-learning system is shown in Fig. 2. The details of the system can be found in [11].

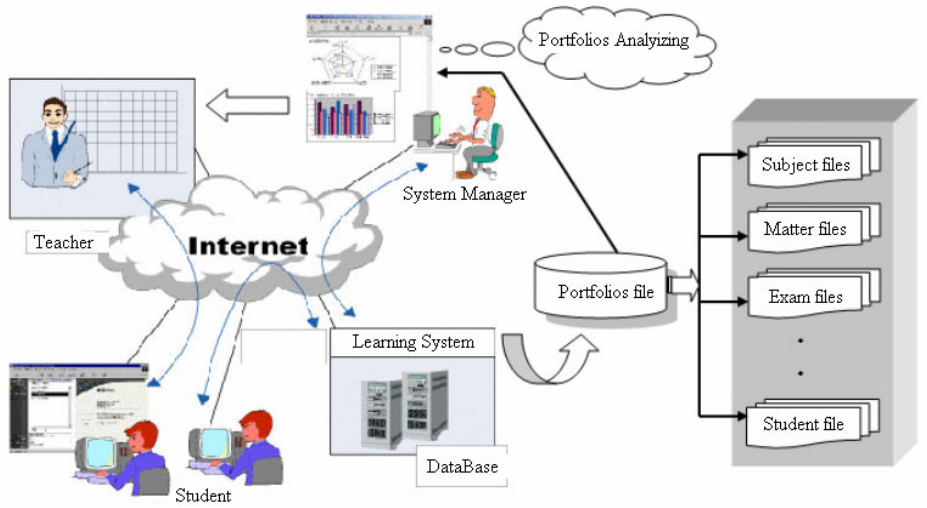

Fig. 2. E-learning system structure.

To evaluate the influence of learning records on the learning effect, we applied multiple regression analysis to study the relationship between the learning records of a learner and his or her learning effect. There are four steps in establishing a regression model:

1. Define the independent variables and dependent variables of the model.
2. Collect data of these variables.
3. Apply regression analysis to calculate the weighting of the independent variables, also known as indicator variables.
4. Establish the regression model.

The dependent variable *Y*, representing the learning effect, is the total of the grades of two monthly exams and one final exam. There are four independent variables: the grade of the online assignments (X_1), the reading time (X_2), the total number of login times (X_3), and the total number of online discussions (X_4). Our initial regression model, therefore, takes the following form:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4.$$

The sample for this research consisted of 47 first-year students in the data processing department at Fu-Hua Secondary Extension School. The gender distribution was 33 female and 14 male students. Most of the students, 87 percent, possessed one year of online experience. A four-credit course, the program design, was provided for these students in the blended e-learning setting, starting on September 15th, 2003, and ending on January 15th, 2004.

The tools used in this research were as follows:

A. On-line materials and video: 12 chapters of Programme Language teaching materials and videos. Some of the materials could not be found in the textbook.
B. E-learning System: this keeps track of the learners' learning records.
C. Evaluation of the learning effect: based on 2 monthly exams and 1 final exam.

In order to get a clear picture of the learners' learning record, our system records and saves the learner's learning path automatically in the database while he/she is browsing online materials. SQL commands in ASP.NET programs automatically accumulate the total reading time in a database. The learning activity data of each learner can be obtained by choosing and analyzing the learners' learning record. We adopted SPSS for Windows 10.0 as an analysis tool. To prove and analyze the research purpose, the analysis methods used included correlative analysis and multiple regression analysis.

4. DATA ANALYSIS

Before stepwise regression analysis was conducted, Pearson analysis was performed to check whether the independent variables were significantly related to the dependent variables. The analysis results shown in Table 1 indicate that all the interdependency

Table 1. Pearson analysis result.

	X_1 (Score of homework)	X_2 (Reading time)	X_3 (Times of logging in)	X_4 (Time of interaction)
Y (Learning effect)	.628**	.488**	.398**	.294*

Note: *p < .05, **p < .01

values were related to the dependent variables. Therefore, it was appropriate to consider all these variables in the stepwise regression analysis. The stepwise regression analysis results are described in the following subsections.

4.1 F Test

The F test examines the overall regression model and is also called Analysis of Variance (ANOVA). It considers all the independent variables as a whole and analyzes whether there is a significant linear relationship between them and the dependent variable. A null hypothesis H_0 representing no linear regression relationship and an alternative hypothesis H_1 (meaning that there is at least one independent variable with a linear relationship with the dependent variable) are formed as follows:

H_0: $b_1 = b_2 = \ldots = b_p = 0$;
H_1: there exists at least one $b_i \neq 0$, $i = 1, 2, \ldots, p$.

If the p-Value of the F test is less than a designated level of significance (usually 0.05), then the null hypothesis is rejected. Thus, there is a linear relationship between the independent variables and the dependent variable. On the other hand, if the p-Value is greater than or equal to 0.05, then H_0 cannot be rejected, and the linear regression relationship is supported. The analysis result of the F test is shown in Table 2. The p-Value is less than 0.005. Therefore, the linear relationships of our models are well established.

Table 2. Anova result.

Source	Sum of square	Degrees of freedom	Mean square	F	p-value
Regression	1898.139	4	474.535	8.655	0.000*
Error	2302.840	42	54.830		
Total	4200.979	46			

Note: *p-value < 0.005

4.2 Coefficient of Multiple Determinations

The coefficient of multiple determinations measures the proportion that independent variables are able to explicate the dependent variable. Assuming that the number of independent variables is p, the formal definition of coefficient of multiple determinations is

$$R^2 = \frac{SSR(X1, \ldots, Xp)}{SST} = 1 - \frac{SSE}{SST}.$$

SSR is the part of the value of the dependent variable (i.e., *SST*) that these *p* independent variables can account for, and *SSE* represents the remaining part that cannot be accounted for by the independent variables.

Another related measure is the adjusted coefficient of determination, defined as follows:

$$\overline{R}^2 = 1 - \frac{SSE(X1, \ldots, Xp)/(n-p-1)}{SST/(n-1)}.$$

n and *p* represents the number of samples and the number of independent variables, respectively. Since the degree of freedom is considered in this measure, it is considered more representative than the coefficient of determination. Applying these measures, the coefficient of multiple determination and adjusted coefficient of determination for the model of the academy and literature category are 0.452 and 0.400, respectively. This indicates that the independent variables in the models can account for around 40 % of the dependent variable. We consider it satisfactory since some learning activities in the classroom that contribute to the learning effect are not considered in this model.

4.3 t Test

The purpose of the t test is to examine whether there is a significant linear relationship between each independent variable and the dependent variable. If there is no significant linear relationship between an independent variable *X*i and the dependent variable *Y*, then the coefficient of the variable b_i should be set to zero. Formally, the following two hypotheses are adopted:

H_0: $b_i = 0$, $i = 1, 2, \ldots, p$;
H_1: $b_i \neq 0$, $i = 1, 2, \ldots, p$.

H_0 is the null hypothesis, and H_1 is the alternative hypothesis. Let the level of significance be α (usually 0.05); if the probability value (p-Value) of the test is less than α, then the null hypothesis H_0 is rejected, meaning that there is a significant linear relationship between the independent variable and the dependent variable; otherwise, if p-Value $\geq \alpha$, then H_0 is accepted, implying that there is not sufficient evidence to support the existence of such a linear relationship. The t test was applied to all the independent variables, and the results for the two categories are shown in Table 3.

According to Table 3, the order of variables is “homework score” (X_1), “reading time” (X_2), “login times” (X_3) and “interaction times” (X_4). The standardized coefficient of variables indicates a positive relationship. The resulting linear regression model is shown as follows:

Y (Learning Effect) = $0.482X_1 + 0.203X_2 + 0.096X_3 + 0.041X_4$.

Table 3. Multiple regression analysis results (Model-1).

Model	Un-Standardized coefficient		Standardized coefficient	t	Significance
	Beta Estimated value	Standard error	Beta distribution		
Score (X_1)	.262	.073	.482	3.573	.001**
Reading time (X_2)	2.210E-02	.015	.203	1.486	.145
Times of logging in (X_3)	1.700E-02	.026	.096	.644	.523
Times of Interaction (X_4)	.217	.750	.041	.289	.774

Note: **$p < .01$

One of the basic assumptions of multiple regression analysis is that the error e_i of independent variables should not be self-related. The Durbin-Watson (DW) test is employed to check for this possibility. If the DW value is close to 2, then we can conclude that the error is not self-related. From our analysis, the DW value of the models was 1.704. This revealed that there was no significant self-relationship.

Since the two variables X_3 and X_4 were insignificant, compared with the other variables, we removed them from the original model (model-1) and construct a second model (model-2) that considered only two variables, X_1 and X_2. Regression analysis was performed again to obtain the results in Table 4. As Table 4 shows, although the significance of X_2 increased, it still did not reach a significant level of 0.05. The resulting linear regression model is shown as follows:

$$Y = 0.513X_1 + 0.242X_2.$$

Table 4. Multiple regression analysis (Model-2).

Model	Un-Standardized coefficient		Standardized coefficient	t	Significance
	Beta Estimated value	Standard error	Beta distribution		
Score (X_1)	.279	.070	.513	3.989	.000**
Reading time (X_2)	2.634E-02	.014	.242	1.885	.066

**$p < .01$

The two different models are compared in Fig. 3.

4.4 Discussion

As the regression model shows, the variable X_1 (Score of homework) is the highest coefficient in the learning effect. This implies that teachers can gain insight into the

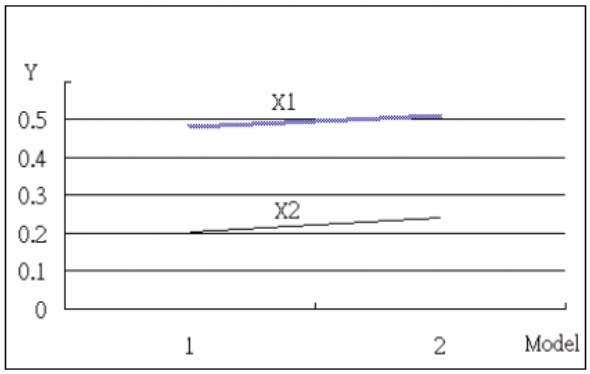

Fig. 3. Comparison of Model-1 and Model-2.

learning effect of a student from his/her performance in homework assignments. This is particularly the case for courses that require a certain amount of practice for the students to actually understand the relevant concepts and knowledge.

Knowledge that involves understanding how to perform cognitive activities is classified as *procedural* knowledge [15]. Program design is certainly a kind of procedural knowledge since designing a program encompasses analyzing the problem, designing data structures and algorithms, and writing suitable code. The other kind of knowledge is *declarative* knowledge, which refers to facts, beliefs, opinions, theories, and so on. The two types of knowledge can be distinguished in terms of facts vs. skills [13]. Our results may be applicable to the learning of procedure knowledge, since a skill is best learned through practice. And in the case of program design, practice is often offered in the form of homework. Whether this result applies to declarative knowledge is more dubious, and further investigation is needed.

However, the other three variables, reading time (X_2), online times (X_3) and interaction times (X_4) are not highly correlated to the learning effect in a blended e-learning setting. This is different from the research results obtained in a 'pure' e-learning setting where the reading time was highly correlated to the learning effect [6]. A major reason for this difference is the fact that the learning materials in blended e-learning can be presented in both a classroom and in cyberspace, thus reducing the importance of online browsing. Although about 30% of the materials were offered exclusively online in our experiment, the impact was obviously not very significant. Likewise, the importance login the number of times and interactions may be moderated by the hours spent each week in a traditional classroom, where learners can interact with teachers and classmates.

5. CONCLUSIONS

As e-learning technology matures, more and more conventional classes may adopt e-learning to complement classroom learning. Some 'pure' e-learning classes also require a certain number of hours of face-to-face meeting to reduce the dropout rates commonly encountered in distance education. Therefore, the blended e-learning model may become more and more popular.

In this study, we investigated the relationship between learning records and the learning effect in a blended e-learning setting. The learning records considered included the grade for online assignments, the reading time, the total number of login times, and total number of online discussions. The learning effect was defined as the total of grade for two monthly exams and one final exam. To collect learning record data, an e-learning system was designed that integrates the data collection functionality of learning activities with a teaching material managing module so that the learning records of all the learners can be recorded automatically. With this system, an experiment was conducted on a program design course in a local high school. A regression model was established after analyzing the experimental data, and the following findings were obtained:

1. The online homework performance was the only item that significantly accounted for the learning effect. Program design is a kind of procedural knowledge. Our result may be applicable to the learning of procedure knowledge, since a skill is best learned through practice. Whether this result applies to declarative knowledge is more dubious, and further investigation is required.
2. The reading time, login times, and interaction times were not highly correlated with the learning effect in our blended e-learning setting. This differs from research results obtained in a "pure" e-learning setting. A major reason for this difference may be the fact that the learning materials in blended e-learning can be presented in both a classroom and in cyberspace, thus reducing the importance of these online behaviors.

REFERENCES

1. A. Bandura, "Self-efficacy: toward a unifying theory of behavioral change," *Psychological Review*, Vol. 84, 1977, pp. 191-215.
2. J. S. Brown, A. Collins, and P. Duguid, "Situated cognition and the culture of learning," *Educational Researcher*, Vol. 18, 1989, pp. 32-42.
3. C. Chang, "A study on the evaluation and effectiveness analysis of web-based learning portfolio (WBLP)," *British Journal of Educational Technology*, Vol. 32, 2001, pp. 435-458.
4. S. Chang, "Constructivism and science education," *Science Education Monthly*, Vol. 202, 1999, pp. 17-25.
5. G. Chen, C. Liu, K. Ou, and M. Lin, "Web learning portfolios: a tool for supporting performance awareness," *Innovations in Education & Teaching International*, Vol. 38, 2000, pp. 19-30.
6. M. Chen, "The development and effectiveness of a web-based learning system for in-service teachers," in *Proceedings of the Working Conference of Research and Development for On-the-Job Training of Teachers*, 2000, pp. 1-12.
7. Y. J. Katz, "Attitudes affecting college students' preferences for distance learning," *Journal of Computer Assisted Learning*, Vol. 18, 2002, pp. 2-9.
8. B. H. Khan, (ed.), *Web-based Instruction*, Englewood Cliffs, New Jersey, 1999.
9. J. Lave and E. Wenger, *Situated Learning, Legitimate Peripheral Participation*, Cambridge University Press, Cambridge, 1991.
10. C. Lee, M. Wu, C. Wu, and Y. Lee, "Research on discovering learning portfolio

styles among world wide web," in *Proceedings of TANET*, 2000, pp. 448-455.
11. C. Lee, *Blended e-Learning Model and Analysis of its Learning Portfolios and Learning Effect*, Master thesis, Institute of Computer and Information Education, National Kaohsiung Normal University, 2004.
12. M. G. Moore and G. Kearsley, *Distance Education: a System View*, Belmont, Wadsworth Publishing Company, 1996.
13. S. Papert, *Mind Storms: Children, Computers and Powerful Ideas*, Basic Books, New York, 1980.
14. J. Piaget, *Structuralism*, Routledge and Kegan Paul, London, 1971.
15. D. H. Schunk, *Learning Theories: an Educational Perspective*, 2nd ed., Prentice Hall, Englewood Cliffs, New Jersey, 1996.
16. E. F. Shores and C. Grace, *The Portfolio Book: a Step-by-Step Guide for Teachers*, Gryphon House, Inc., 1998.
17. G. Trentin, "Telematics and on-line teacher training: the POLARIS project," *Journal of Computer Assisted Learning*, Vol. 13, 1997, pp. 261-270.
18. J. Tsou, "General introduction to digital learning," presentation slides from Education and Training Division at Institute for Information Industry, Taipei, 2003.
19. L. Vavrus, "Put portfolios to the test," *Instructor*, Vol. 100, 1990, pp. 48-53.
20. C. Wen and M. Wu, *Information Education Theory and Application in a New Era*, Sung-Gung Publisher, Taipei, 1999.
21. M. L. Williams, K. Paproch, and B. Covington, *Distance Learning: the Essential Guide*, SAGE Publications, Thousand Oaks, 1999.
22. D. Wood, J. S. Brunet, and G. Ross, "The role of tutoring in problem solving," *Journal of Psychology and Psychiatry*, Vol. 17, 1976, pp. 89-100.

Dowming Yeh (葉道明) is an Associate Professor in the Institute of Information and Computer Education at the National Kaohsiung Normal University. He received his Ph.D. degree in Computer Science from the University of Utah in the U.S.A. Before assuming his current post, he was an Associate Professor in the Department of Management Information Systems at National Pingtung University of Science and Technology and a manager at the Institute for Information Industry in Taiwan. His research interests include software reengineering, e-Learning, web engineering, program analysis, and human computer interaction. Dr. Yeh is a member of IEEE and ACM.

Chun-Hsiung Lee (李春雄) received his B.S. degree from National Pingtung University of Science and Technology in 1998, and his M.S. degree from National Kaohsiung Normal University in 2004. He is working toward the Ph.D. degree in the Department of Information Management at National Taiwan University of Science and Technology. His current research is focused on e-learning.

Pei-Chen Sun (孫培眞) is currently an Assistant Professor and serves as System Manager at National Kaohsiung Normal University. He received his Ph.D. degree in Management Information Systems from National Sun Yat-sen University. He is a member of the Phi-Tau-Phi Scholastic Honor Society. His main research areas are e-learning and information systems. He has participated in several international conferences, such as ICIS, HICCS, PACIS, and ACME. His research papers have been published in the Journal of Information Management, Computers and Education, and the International Journal of Innovation and Learning.