ICCE 2011

The 19th International Conference on Computers in Education

28 November - 2 December • Chiang Mai, Thailand

Workshop Proceedings: Supplimentary Proceedings of the 19th International Conference on Computers in Education: ICCE 2011

Editors

Ahmad Fauzi Mohd Ayub

Ben Chang

Krittaya Leelawong

Fu-Yun Yu

Tsukasa Hirashima

Gautam Biswas

APSCE

NECTEC

a member of NSTDA

TCU

THAILAND CYBER UNIVERSITY

MAEJO UNIVERSITY

Chapter 7

Modeling, Management and Generation of Problems/Questions in Adaptive Learning Environment

Preface

Solving problems/questions is one of the most indispensable and important components in the teaching and learning process. Problems/questions with adequate quality in various testing conditions are believed to enable teachers to assess individual students' capability and readiness of transfer in specific domain knowledge.

Despite this, there are still many areas in need of systematic investigation to promote knowledge and skills on problems/questions-centered learning approach, including learning by problem solving and/or generation. For instance: what criteria constitute as adequate test item quality (in addition to frequently cited psychometric index like item difficulty, discrimination index); how to best assess learner's capability with appropriate quality level within constrains (e.g., an optimal number of items, time limitation, etc.); any feasible metadata heuristics and/or techniques for problems/questions selection; any promising alternative strategies for compiling a sufficient amount of number of problems/questions; any scaffolding techniques for question/problem-generation implementation and instructional diffusion and so on.

This is the 5th workshop focusing on the same topic. This continuous workshop will provide a good and timely opportunity to present and share the results and issues about "problems/questions" in ICCE community.

Executive PC members
Kazuaki Kojima, *Waseda University, Japan*

Organizers
Tsukasa Hirashima, *Hiroshima University, Japan*
Tomoko Kojiri, *Kansai University, Japan*
Tanja Mitrovic, *University of Canterbury, New Zealand*
Fu-Yun Yu, *National Cheng-Kung University, Taiwan*

Mohd Ayub A. F. et al. (Eds.) (2011). Workshop Proceedings of the 19th International Conference on Computers in Education. ChiangMai, Thailand: Asia-Pacific Society for Computers in Education.

Predictive Effects of Online Peer-Assessment on Student Question-Generation

Fu-Yun YU[a]* & Chun-Ping WU[b]
[a]*Institute of Education, National Cheng Kung University, Taiwan*
[b] *Department of Educational Technology, TamKang University, Taiwan*
*fuyun.ncku@gmail.com

Abstract: The study examined the predictive effects of online peer-assessment on student question-generation. Specifically, the individual and collective predictive effects of two types of feedback (i.e., quantitative ratings and descriptive comments) available in peer-assessment learning systems on student question-generation performance were investigated. A total of 233 students participated in the study for six weeks. An online learning system that allows students to contribute to and benefit from the process of question-generation and peer-assessment was adopted. The regression result found that quantitative ratings and descriptive comments individually and collectively significantly predicted question-generation performance. Suggestions for learning system development are provided.

Keywords: online learning system, peer-assessment, student question-generation

Introduction

Both theoretical and empirical foundations of student question-generation support its beneficial effect on learning [1-6]. Recently, in view of the numerous advantageous features of network technology, a number of online learning systems with student question-generation as the focus have been developed. Most existing systems enable students to generate questions of different types and to incorporate media formats as part of the question. Also frequently included in these systems is an element of peer-assessment [6-11].

The benefits of including peer-assessment within the student question-generation context can be understood and appreciated in light of cognitive conflict theory, social constructivism and social learning [12-14]. Nevertheless, there is a lack of empirical evidence supporting the coupling effects of online peer-assessment with student question-generation. An investigation into such issues as "if and how feedback students receive during online peer-assessment affect student question-generation performance" will warrant its inclusion in online student question-generation systems. Since feedback can be expressed in quantitative and descriptive forms, its individual and collective predicative effects on student question-generation are examined. Three research hypotheses are proposed in the study:

1. The averaged quantitative ratings received from assessors on the composed questions will significantly predict student question-generation performance.
2. The quality of descriptive comments received from assessors on the composed questions will significantly predict student question-generation performance.
3. The averaged quantitative ratings and the quality of descriptive comments received from assessors on the composed questions will collectively significantly predict student question-generation performance.

In consideration of the fact that a considerable proportion of students do not experience question-generation during their formal schooling [15-16] and have viewed student question-generation as difficult or very difficult [11], answers to the above questions will help provide some directions for better online question-generation activity design and implementation.

1. Method

1.1 Online Learning System

A learning environment that allows students to contribute and benefit from the process of constructing question items and receiving feedback from their peers about the composed questions was used. Essentially, the question-generation sub-system enables multimedia files to be included as parts of the question and texts of different fonts, size and styles can be used (see Figure 1).

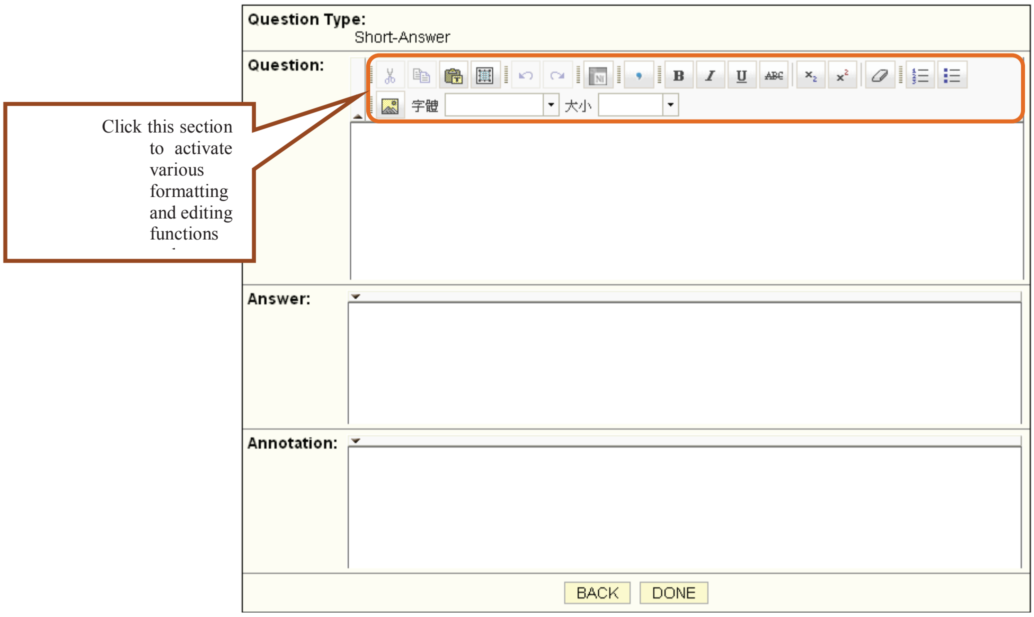
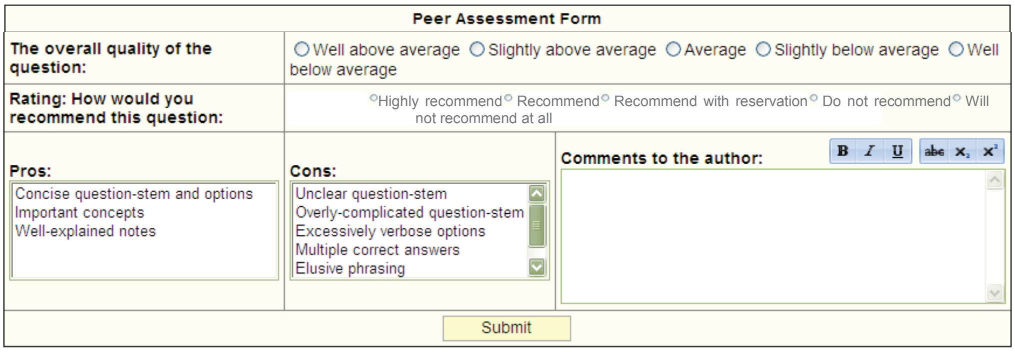


Figure 1 A screenshot of short-answer question-generation

The peer-assessment sub-system, on the other hand, enables assessors to give their evaluative feedback using an online assessment form. On the form, assessors can assess the overall quality of the generated question on a five-point rating scale (from "well below average" to "well above average") and to rate their recommendation of the question to be included in the drill-and-practice item bank (from "Will not recommend at all," to "highly recommend"). Also, assessors can give descriptive comments with regards to the question being examined in a designated feedback space by referring to a set of built-in criteria (see Figure 2).



Peer Assessment Form

The overall quality of the question: ○ Well above average ○ Slightly above average ○ Average ○ Slightly below average ○ Well below average

Rating: How would you recommend this question: ○ Highly recommend ○ Recommend ○ Recommend with reservation ○ Do not recommend ○ Will not recommend at all

Pros:
Concise question-stem and options
Important concepts
Well-explained notes

Cons:
Unclear question-stem
Overly-complicated question-stem
Excessively verbose options
Multiple correct answers
Elusive phrasing

Comments to the author:

Submit

Figure 2. Assessment form for assessors to provide feedback to question-authors

1.2 Participants and Implementation Procedures

Two hundred and thirty-three 5th graders from eight classes participated in the study for six consecutive weeks. Participants were informed that the introduced online question-generation and peer-assessment activity was intended to augment their science learning.

Each week for the duration of the study, students headed to a computer laboratory to participate in a 40-minute learning activity after attending three instructional sessions allocated for science. To ensure that participants possessed the fundamental skills of the introduced activity, a training session on generating chosen question types and the coupled online peer-assessment with hands-on activity was arranged at the commencement of the study. Considering that true/false and multiple-choice questions are among the most frequently encountered question types in primary schools in Taiwan, these two types of question-generation options were chosen. Each week students were directed to individually generate at least one question for each of the two chosen question types in accordance with the instructional content covered that week and assess at least two questions from a pool of peer-generated questions for each chosen question type.

1.3 Variables

The quantitative ratings received from assessors consisted of two parts: the overall quality of the question and recommendation for inclusion in follow-up drill-and-practice sessions. The overall quality and recommendation received from assessors per question per week were averaged throughout the activity.

The quality of descriptive comments received from assessors on the composed questions and student performance in question-generation was defined against a set of criteria. For peer-assessment, all comments question-authors received with regards to a specific question item were analyzed against a pre-defined scheme and were averaged. The averaged scores per question per week were then summed up. Specifically, the quality of descriptive comments was evaluated in terms of four discrete levels: general comments, specific comments where strengths and weakness are identified, identification for improvement and explicit suggestion for further refinement of questions.

To assess students' performances in question-generation, in reference to the Torrance creativity index [17], King's question cognitive levels [18] and questions generated by students, the following criteria were adopted: fluency, flexibility, elaboration, originality, cognitive level and importance. Each of the indexes was further operationally defined to ensure objective assessment.

2. Results

2.1 Descriptive statistics of examined variables

The means and standard deviations of the quality of feedback received on the composed questions (including quantitative peer-ratings and descriptive comments) and students' performance in question-generation are listed in Table 1.

Table 1 Descriptive statistics and correlations between variables (N=233)

Variable	Quantitative ratings	Descriptive comments	Question-generation
Mean (SD)	3.45 (0.68)	6.70 (3.60)	32.11 (13.56)

Note: * p <0.05, ** p <0.01

2.2 The predictive effect of quantitative ratings on question-generation performance

The regression result presented in Table 2 supports that the quantitative ratings significantly predict question-generation performance, ($\beta = 0.28$, $p < 0.01$).

Table2 Regression analysis for quantitative ratings predicting question-generation performance

	B	SEB	β
Model			
Constant	12.72	4.46	
Quantitative ratings	5.60	1.27	0.28**
R-square		0.08	
F		19.59**	

Note: *a.* Predictor:(Constant), Quantitative ratings

b. Dependent variable: Question-generation performance

*c.** p <0.05, ** p <0.01

2.3 The predictive effect of the quality of descriptive comments on question-generation performance

The regression result presented in Table 3 supports that the quality of descriptive comments significantly predicts question-generation performance, ($\beta = 0.37$, $p < 0.01$).

Table 3 Regression analyses for quality of descriptive comments predicting question-generatic performance

	B	SEB	β
Model			
Constant	22.80	1.74	
Quality of descriptive comments	1.39	0.23	0.37**
R square		0.14	

F	36.48**

Note: *a.* Predictor:(Constant), Quality of descriptive comments

b. Dependent variable: Question-generation performance

*c.** p <0.05, **p <0.01

2.4 The collective predictive effect of the quantitative ratings and the quality of descriptive comments received on question-generation performance

To avoid multicollinearity, Pearson correlations was conducted and found that quantitative ratings is not correlated with the quality of descriptive comments (r= 0.1, p =0.13); therefore, these two variables could collectively included in multiple regression analysis. The quality of descriptive comments significantly predicted a significant proportion of variance on students' question-generation performance (R^2 = 0.14, F=36.48, p < 0.01). Adding the variable of quantitative ratings significantly enhanced the R-square (R^2 change = 0.06, F=16.98, p < 0.01); therefore, the quality of descriptive comments and quantitative ratings collectively significantly predict question-generation performance (β_{qual}= 0.35, p < 0.01; β_{quan}= 0.25, p < 0.01, respectively).

Table 4 Multiple Regression analyses for Quality of feedback predicting question-generation performance

	Model 1			Model 2		
Variable	B	SE	β		SE	β
Constant	22.80	1.75		6.44	4.32	
Quality of descriptive comments	1.39	0.23	0.37**	1.30	0.22	0.35**
Quantitative ratings				4.91	1.19	0.25**
R-square		0.14			0.20	
F for change in R-square					16.98**	

3. Discussion and conclusions

Numerous online student question-generation learning systems have been developed for students to interact with the content by generating questions and to interact with their peers online for the improvement of the questions by peer-assessment. This study explored whether feedback received from peers contributed to question-generation performance.

The current study confirmed the coupling effects of online peer-assessment on student question-generation performance. Specifically, this study substantiated that the quantitative ratings and the quality of descriptive comments question-authors received from peers individually and collectively contributed to their question-generation performance. In other words, question-authors who received higher quantitative ratings tend to demonstrate better performance in composing questions. Also, the better quality of descriptive feedback received on their composed questions leads to higher performance in question-generation tasks. Furthermore, question-authors who received higher quality of descriptive feedback together with higher ratings on their questions tend to demonstrate better performance in composing questions.

The obtained findings have important empirical significance as well as implications for online system developments. First, despite that peer-assessment is coupled with question-generation in most existing online learning systems, its supportive effects on student question-generation performance has rarely been substantiated empirically. This present study, for the first time, evidenced the respective and collective effects of quantitative ratings and descriptive comments and supported the inclusion of peer-assessment in online student question-generation systems.

Based on the findings of this study, several suggestions are provided. First, instructors with students inexperienced in student question-generation and who can benefit from extra support for better question-generation performance are advised to include an element of online peer-assessment for the promotion of performance in the introduced task. Second, as this study found that the variable of descriptive feedback explained more variance of question-generation performance, the importance of providing question-authors with descriptive feedback could not be ignored. Finally, online student question-generation system with peer-assessment should consider including both quantitative ratings together with descriptive comments key-in space for maximal question-generation performance.

Acknowledgement

This paper was funded by research grants from the National Science Council, Taiwan, ROC (NSC 96-2520-S-006-002-MY3; NSC 99-2511-S-006 -015 -MY3).

References

[1] Abramovich, S., & Cho, E. K. (2006). Technology as a medium for elementary preteachers' problem-posing experience in Mathematics. *Journal of Computers in Mathematics and Science Teaching, 25*(4), 309-323.

[2] Barlow, A., & Cates, J. M. (2006). The impact of problem posing on elementary teachers' beliefs about mathematics and mathematics teaching. *School Science and Mathematics, 106*(2), 64-73.

[3] Brown, S.I., & Walter, M. I. (2005). *The Art of Problem Posing* (3rd ed). New Jersey: Lawrence Erlbaum Associates.

[4] Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research*, *66*(2), 181-221.

[5] Yu, F. Y. & Liu, Y. H. (2008). The comparative effects of student question-posing and question-answering strategies on promoting college students' academic achievement, cognitive and metacognitive strategies use. *Journal of Education and Psychology*, *31*(3), 25-52.

[6] Yu, F. Y., Liu, Y. H. & Chan, T. W. (2005). A Web-based learning system for question-posing and peer assessment. *Innovations in Education and Teaching International, 42* (4), 337-348.

[7] Barak, M. & Rafaeli, S. (2004). On-line question-posing and peer-assessment as means for web-based knowledge sharing in learning. *International Journal of Human-Computer Studies 61*(1), 84-103.

[8] Denny, P., Hamer, J., Luxton-Reilly, A. & Purchase, H. (2008). PeerWise: students sharing their multiple choice questions. In Proceeding of the Fourth international Workshop on Computing Education Research, pp. 51-58, Sydney, Australia.

[9] Fellenz, M. R. (2004). Using assessment to support higher level learning: the multiple choice item development assignment. *Assessment & Education in Higher Education, 29*(6), 703-719.

[10] Wilson, E. V. (2004). ExamNet asynchronous learning network: augmenting face-to-face courses with student-developed exam questions. *Computers & Education*, *42*(1), 87-107.

[11] Yu, F. Y. (2009). Scaffolding student-generated questions: Design and development of a customizable online learning system. *Computers in Human Behavior, 25* (5), 1129-1138.

[12] Falchikov, N. & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, *70* (3), 287-322.

[13] Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, 249-276.

[14] Van Gennip, N. A. E., Segers, M. & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction, 20* (4), 280-290.

[15] Moses, B.M., Bjork, E. & Goldenberg, E.P. (1993). Beyond problem solving: problem posing. In S.I. Brown & M.I. Walter (Eds.), *Problem posing: reflections and applications* (pp. 178-188). NJ, Hillsdale: Lawrence Erlbaum Associates.

[16] Vreman-de Olde, C. & de Jong, T. (2004). Student-generated assignments about electrical circuits in a computer simulation. *International Journal of Science Education* 26 (7), 859-873.
[17] Torrance, E.P. (1974). *Torrance Tests of Creative Thinking*. Scholastic Testing Service, Inc.
[18] King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist*, *27*(1), 111-126.

Mohd Ayub A. F. et al. (Eds.) (2011). Workshop Proceedings of the 19th International Conference on Computers in Education. ChiangMai, Thailand: Asia-Pacific Society for Computers in Education.

Toward better Collaborative Problem-solving in Programming Learning: Use of Pair Programming and Its Observation

Yuki HIRAI[a*] & Tomoo INOUE[a]**
[a]*Graduate School of Library, Information and Media Studies, University of Tsukuba, Japan*
*s0930522@u.tsukuba.ac.jp **inoue@slis.tsukuba.ac.jp

Abstract: To realize better programming learning, we have adopted “pair-programming” method to the introductory programming course of a university. Pair-programming is a programming method that two persons get involved with a single programming task using a single computer terminal, where only one person types the keyboard. Though there have been several researches on pair-programming learning practice and been reported its usefulness, only impressions were reported. Through the actual pair-programming practice, we could observe both successful case and failed case in solving the problem that arose in the course of completing the assigned task, and found that there seemed to be difference in utterance patterns between the successful case and the failed case.

Keywords: Pair-programming, Programming learning, Problem-solving, Computer-supported collaborative work (CSCW)

Introduction

The ability to understand the grammar of a programming language, to write a program, and to assemble an algorithm, is required in programming education. When a learner actually creates a program, some problems typically occur, even if the grammar and a (relatively easy) example of the program language are understood [12]. In programming education, numerous practices, including the support of problem-solving, have been developed to date. Education and study methods have also received considerable attention [10].
The programming method called ‘pair-programming’ originated in industry as a key component of the Extreme Programming (XP) development methodology [1]. As the name suggests, two programmers work together at the same machine while developing code. One programmer (the driver) operates the keyboard and focuses on entering code, while the other programmer (the navigator) observes the work of the driver and offers suggestions in the code. The programmers regularly exchange roles. Creating a program by pair-programming is collaborative work, and offers further benefits in respect of sharing and enhancing programming expertise and refining collaborative technique [16]. In the Computer-supported cooperative work (CSCW) context, interruptions of software teams have been investigated [4], and studies have explored interruption patterns among software developers who program in pairs versus those who program solo.
In some programming education, pair-programming has been conducted as one of the programming learning methods. Especially in introductory programming courses, for example, it has been reported that pair-programming is better than solo-programming in respect of improving the quality of programming [6,7,8,11,15]. However, there were numerous instances in which the pair-programming had faced the problem which

problem-solving did not go well. Of course, the effect of pair-programming varies with the actual composition of the pairs, but failure on the part of one of the pair, in the introductory stage, can easily spill over into later, more involved tasks. Moreover, if problem-solving does not go well, a decrease in motivation to study will typically occur. In this case, we must seek to support the pair, with a view to improving their pair-programming learning.

In this study, pair-programming was conducted in an introductory programming course. Success and failure cases in pair-programming were compared. In the comparison, we focused on the conversation between the pair in pair-programming. In the failure cases, it was found that speech length tended to be long, and there might be a great deal of continuous speech.

Our research objective in broader sense is to support programming learning. Pair programming has been focused as one of the promising techniques of programming learning. We do not intend to just using pair programming. We intend to expand pair programming to computer-supported pair programming (CSPP). This means that a computerized environment (not the computer used for programming basically) senses the learning status of the pair, and once the environment senses something wrong with the pair it intervenes in the learning. This could be a mixture of ICAI (Intelligent Computer-Aided Instruction) and CSCL (Computer Supported Collaborative Learning) under the ubiquitous computing technology. To realize such CSPP, we thought we need some symptoms to indicate the status of pair programming. This led to the study in this paper.

1. Related Works

1.1 Pair-programming in an Introductory Programming Course

Previous research suggested that pair-programming was better than solo-programming in numerous respects. For example, it was better in respect of the quality of program code [6,7], the success rate in programming courses [7,8], results of mid-term or final examinations [8], and/or submission rate of assignments [15]. Rountree et al. reported that understanding and/or ability to create program code were improved after pair-programming was conducted [11].

The aforementioned research reported the positive effects of pair-programming, but did not analyze the process of pair-programming or the pairs whose problem-solving did not go well. In this research, the conversations of some pairs in pair-programming were analyzed, and specifically, pairs that failed in problem-solving were studied.

1.2 Communication Analysis in Pair-Programming

In previous research (which did not focus on introductory programming courses), conversations of the pairs in pair-programming was analyzed. Chen et al. recorded the utterance of pairs and described the context of pair-programming. They suggested that there was a mental distance between the driver and the navigator, and communication supports such as visualizing the rules of the pair were necessary [3]. Chong et al. also recorded the utterance of pairs and described the context of pair-programming. They suggested that the distribution of expertise among the members of a pair had a strong influence on the tenor of pair-programming interaction, and keyboard control had an effect on decision-making within the pair [5]. Bryant et al. investigated the distribution of utterance categories in pair-programming, and suggested that there was no significant difference in the distribution between the driver and navigator, and both driver and navigator work at similar levels of abstraction [2].

These studies analyzed the conversation of pairs, but did not compare success and failure cases in pair-programming. In this study, interactions between the driver and navigator have been observed, communications in pair-programming have been analyzed, success and failure cases have been compared, and the characteristics of failure cases have been studied.

1.3 Roles of Conversation in Pair-Programming

Wray [16] described the roles and effects of conversation in pair-programming from his own experience. He mentioned that the roles of conversation were sharing expertise among pairs and getting on the track for problem solving. He predicted that programmers who chat about their programs more should be more productive and that those who pose deep questions for each other should be most productive of all.

His description suggests that problems occurring in pair-programming might be solved through conversation among pairs, and that conversation may be a significant indicator in comparisons between success and failure cases in pair-programming. In the present study, differences in conversation between success and failure cases in pair-programming were explored.

2. Pair-Programming Practice

2.1 Practice Setting

In this study, pair-programming was conducted in an introductory programming course, “Programming I”, which targeted freshmen in the university’s department of information. The goals of the course were as follows:

- Learners understand the description and composition of software and the mechanism of programming.
- Learners can compile and execute a program written in C language.
- Learners understand the basis of C language, such as variables, control of flow, functions, arrays, character and string handling, and file I/O.

The course involved ten weekly 75-minute lectures, from September 2010. Pair-programming was conducted in six 30-minute practice sessions as the part of the lecture.

As preparation for pair-programming practice, the training session was conducted. The training was conducted in the same setting as the following pair-programming practice, because of the possibility that some learners had not experienced pair-programming.

In each pair-programming practice session, a program-creation assignment, involving contents hitherto studied, was given to the participants. An example of the assignment is shown in Table 1. The following six instructions were given to the learners:

- Only the driver can operate the keyboard and mouse. The navigator must not touch them, but may point to the display. The navigator must observe and support the work of the driver.
- The assignment ends when the program is executed and a correct answer to the assignment is obtained. Please end the assignment as soon as possible.
- The driver and navigator may refer to the textbook [14]. You must not refer to any web pages.
- The teacher and teaching assistants (TA) do not accept any questions concerning the assignment while practicing. Please call on them only in the event of equipment trouble.
- Please create the program easy to understand by adding pertinent comments.
- You have 30 minutes to success the assignment. Please submit your code even if failure, when the time limit is reached.

A total of 62 learners participated in the practice session (52 freshmen and 10 upper-years). Pair combinations were decided by one of the authors. The participants did not exchange roles (of driver and navigator) in each practice session because the practice time was short. Figure 1 shows a screenshot of the practice session. Figure 2 shows a scene from the practice session. Three cameras were used for recording communication.

Table 1. An example of the exercises in the pair-programming class.

Assignment 1:
Create a program for permutation and combination according to the following specification.

<u>Specification</u>
* Input: n, r (integer)
* Output: "nPr = ?, nCr = ?"(? is calculated value)

<u>Example</u>
When 8 is input to n and 3 is input to r, the calculated result is displayed as follows:
8P3 = 336, 8C3 = 56

<u>Hint</u>
As for permutation and combination, the general formulas are given as follows:

$$n\mathrm{Pr} = \frac{n!}{(n-r)!} \quad (n \geq r > 0), \quad nCr = \frac{n!}{r!(n-r)!} = \frac{n\mathrm{Pr}}{r!} \quad (n \geq r > 0), \quad n! = \begin{cases} n \times (n-1)! & (n \geq 2) \\ 1 \quad (n = 0,1) \end{cases}$$

Figure 1. Setup of the cameras for data collection.

Figure 2. Scene from the practice session.

2.2 Definition

In this study, "Success" , "Failure" and "Problem" are defined as follows:

- Success: "Success" is the identifier of showing the problem was solved. It does not relate to the learner's "success" of learning.
- Failure: "Failure" is the identifier of showing the problem was not solved within the given limited time. It does not relate to the learner's "failure" of learning.
- Problem: "A problem" is a compilation error that occurs when learners compile their program, or a runtime error that occurs runtime including whose result does not meet the learners' expectation.

Although we know those concepts of Vygotsky's Zone of Proximal Development and Lave and Wenger's Legitimate Peripheral Participation, and "failure" is not just failure there [9], we do not deal with that "failure" in this paper. There failure can be resource for learning. Here the term "failure" is used as an identifier of unsuccessful result of solving the error that occurred during programming. In other words, the term "failure" and "success" in this paper do not imply any notion known in learning sciences. They are simple and clear identifiers of the result of solving the errors.

2.3 Problems Occurring in the Practice Session

Table 2 shows the problems which occurred among the pairs whose communication was recorded. These problems occurred in pairs of first-year students. Table 2 shows six success cases and three failure cases. Some pairs attempted to solve two or more problems in a given practice session. Problem-solving went well in the success cases. The problems listed in Table 2 were causes of the error that the pair finally identified. In Failure Case A and B, problems which the authors recognized by observing the video are listed, because the respective pair did not recognize the cause of error. There were only three failure cases in this practice session. This is because the assignments given to the participants were easy. Most of the pairs completed the assignment within the time limit.

Table 2. Problems occurring in the practice session.

Case	**Pair**	**Problems**
Success A	Pair A	Compilation error Semicolon was not written at the end of a line.
Success B	Pair B	Compilation error The string "enum" was a reserved word.
Success C	Pair B	Compilation error The source file was not preserved in the superscription.
Success D	Pair B	Compilation error, Segmentation error The "scan" sentence was written like 'scanf("%d", a);'. – "&" was missing.
Success E	Pair C	Run-time error Beginning of a block did not correspond to the end. There were some spelling mistakes.
Success F	Pair C	Run-time error The return value of a function was not correctly returned.
Failure A	Pair D	Run-time error The case divided by 0 was included in the "for" sentence.
Failure B	Pair A	Run-time error The value of a variable was not correctly substituted by the global variable declaration.
Failure C	Pair E	Compilation error Neither the main file nor the header file was correctly linked.

3. Difference between Success and Failure Cases

Success and Failure cases in problem-solving were analyzed and compared in term of pairs' conversation. The utterances of the pairs and the context of pair-programming were recorded with iCorpusStudio [13], which is a video-analysis support tool. With the tool, we can simultaneously view the recorded data as multiple video, audio, and motion, while annotating the interpretations of the interactions as labels.

3.1 Examples of Success and Failure Cases

We show two example sequences including utterances and some descriptions; one for "Success" case and the other for "Failure" case.
Table 3 shows a conversation in Success case A. In this case, the following error message "19: error: expect ';' before 'return'" was output. The learners solved this problem in 100 seconds. Speech length marks the time from the point that the learner started his/her speech, to the point that the learner ended the speech.
Table 4 shows a part of conversation in Failure case B. In this case, there was no output though the program was executed and the driver input a value to a variable. The learners tried to move the "while" sentence to another line. The learners spent 588 seconds solving this problem, but the problem was not solved. The driver uttered 19 times in this case, while the navigator uttered 61 times.

Table 3. A conversation in Success case A.

Utter. no.	Spe-aker	Speech length (sec.)	Utterance
1	D	0.9	The 19th line.
2	D	0.9	Ah… This line.
3	N	1.5	Ah… "return 0".
4	N	4.1	Line numbers are shown when a setting is changed.
5	D	0.9	Really?
6	D	1.8	I do not compile this program.
7	N	2.7	Did you save this program? Ah, you did.
8	D	1.9	I try to delete unnecessary lines.
9	D	1.1	(I think) the way is not good.
10	N	1.6	return 0…
11	D	0.7	This point
12	N	1.4	Ah… after the "printf" sentence.
13	N	0.7	Um…
14	N	1.5	functional…
15	N	1.2	The 19th line
16	N	1.3	No changes are appeared.
17	D	1.5	This program consists of 17 lines.
18	D	1.7	Ah…, 19, the last line…
19	N	3.1	Parentheses… Let's make sure the position of parentheses
20	N	1.7	The number of braces is wrong? …
21	D	0.7	Ok. (the problem was solved)

* Speaker - D: Driver, N: Navigator
* Speech length - The length more than 2 seconds is highlighted.
* Utterance - Description in the parentheses is the supplement by the authors.

Table 4. Part of a conversation in Failure case B.

Utter. no.	Spe-aker	Speech length (sec.)	Utterance
14	N	1.2	The “while” sentence...
15	D	0.4	Umm.
16	N	1.6	Let’s move outside of the ”main” function.
17	D	1.1	”Main”?
18	N	1.1	Please move above the function.
19	D	0.7	Umm.
20	N	1.7	From this line to this line... Ok.
21	D	1.1	Umm.
22	N	2.6	Please cut the selected lines.
23	N	0.7	Next...
24	N	1.2	Let me see...
25	N	1.2	”While” sentence... (The driver operates.)
26	N	2.9	Not ”while” sentence. Sorry, please undo.
27	N	2.1	Sorry, it became strange.
28	N	4.4	You may move this function outside. (The driver operates.)
29	N	3.5	From this line to this line... (The driver operates.)
30	N	5.6	Because this function was moved outside, the declaration of the variable might be wrong.
31	N	4.0	”jyun” (= a variable) is ok. ”ans” (= a variable) is ok. ”n” (=a variable) is ...
32	N	1.3	”n” is...
33	N	1.9	Is it correct to declare this variable outside the function?
34	N	3.0	Global...?
35	N	0.7	Index...
36	N	4.5	Global... global variable.
37	N	1.6	Ok. It is possible to declare this variable outside the function.

* Speaker - D: Driver, N: Navigator
* Speech length - The length more than 2 seconds is highlighted.
* Utterance - Description in the parentheses is the supplement by the authors.

3.2 Findings obtained from the Examples

As for the speaker, in the failure case, the driver and navigator spoke alternately from utterance 14 to 22. From utterance 23, however, the navigator spoke continuously; that is, the driver did not talk. The navigator spoke more continuously in the failure case than in the success case. As for the speech length, there were 9 utterances that are more than two seconds in length in the failure case. Especially, from utterance 26 to 31, the navigator spoke continuously and all of his succeeding utterances were more than two seconds in length.
The investigation of these example dialogues suggests that there may be a relationship between speech length and/or speech continuity and success/failure of problem-solving.

Discussion of the relation may require further investigation; for example, through observing more cases in the practice sessions.

4. Conclusion

We have adopted pair-programming method in software engineering to programming learning. Naturally there occurred both successful case and failed case in solving the problem when the problem arose in the course of completing the task. We observed a few such cases and found that there seemed to be difference in utterance patterns between successful case and failed case. We will analyze the learners' conversation and behavior more in detail to obtain clearer symptoms to indicate the status of pair programming. Then we will develop a computer-supported pair programming system that uses the symptoms.

Acknowledgments

The authors would like to thank Naiwen Tei for her contribution to the data collection. This research was partially supported by the Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for JSPS Fellows 23.2956, the JSPS Grant-in-Aid for scientific research 22500104, and the Research Projects of Graduate School of Library, Information and Media Studies, University of Tsukuba.

References

[1] Beck, K. (1999). Extreme Programming Explained: Embrace Change, Reading, PA: Addison-Wesley.
[2] Bryant, S., Romero, P., and Boulay, B. (2008). Pair Programming and the Mysterious Role of the Navigator, *International Journal of Human-Computing Study*, *66*(7), Academic Press, 519-529.
[3] Chen, W. and Nordbo, M. (2007). Understanding Pair-Programming from a Socio-cultural Perspective, *Proc. Computer-supported Collaborative Learning (CSCL)*, International Society of the Learning Sciences (ISLS) Press, 138-140.
[4] Chong, J., and Siino, R. (2006). Interruptions on Software Teams: A Comparison of Paired and Solo Programmers, *Proc. Computer-supported Cooperative Work (CSCW)*, ACM Press, 29-38.
[5] Chong, J., and Hurlbutt, T. (2007). The Social Dynamics of Pair Programming, *Proc. International Conference on Software Engineering (ICSE)*, IEEE Press, 354-363.
[6] Hanks, B., McDewell, C., Draper, D., and Krnjajic, M. (2004). Program Quality with Pair Programming in CS1, *Proc. Innovation and Technology in Computer Science Education (ITiCSE)*, ACM Press, 176-180.
[7] McDowell, C., Werner, L., Bullock, H., and Fernald J. (2002). The Effects of Pair-Programming on Performance in an Introductory Programming Course, *Proc. ACM SIGCSE*, ACM Press, 38-42.
[8] Nagappan, N., Williams, L., Ferzli, M., Wieve, E., Yang, K., Miller, C., and Balik, S. (2003). Improving the CS1 Experience with Pair Programming, *Proc. ACM SIGCSE*, ACM Press, 359-362.
[9] National Research Council. (1999). How People Learn, National Academies Press.
[10] Onishi, K. (2010). How to Teach What in a Course of Programming?: Foreword. *Information Processing, 51*(10), Information Processing Society of Japan (IPSJ) Press, 1341. (in Japanese)
[11] Rountree, J., Rountree, N., Robins, A., and Hannah, R. (2005). Observations of Student Competency in a CS1 Course, *Proc. Australasian Computing Education Conference (ACE)*, Australian Computer Society Press, 145-149.
[12] Shinkai, J., and Miyaji, I. (2009). Effects of C Programming Education Which Makes a Point of Process with Evaluation Activity. *Journal of Japanese Society for Information and Systems in Education (JSiSE)*, *26*(1), 16-25. (in Japanese)
[13] Sumi, Y., Yano, M., and Nishida, T. (2010). Analysis Environment of Conversational Structure with Nonverbal Multimodal data, *Proc. International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI)*, ACM Press.
[14] Takahashi, M. (2007). Yasashii C (in Japanese Title), Softbank Creative Press, ISBN-10: 9784797343663.
[15] Urness, T. (2009). Assessment Using Peer Evaluations, Random Pair Assignment, and Collaborative Programming in CS1, *Journal of Computing Sciences in Colleges*, *25*(1), 87-93.
[16] Wray, S. (2010). How Pair Programming Really Works, IEEE Software, Jan./Feb., IEEE Press, 50-55.

Mohd Ayub A. F. et al. (Eds.) (2011). Workshop Proceedings of the 19th International Conference on Computers in Education. ChiangMai, Thailand: Asia-Pacific Society for Computers in Education.

Experimental Study on Failures in Composing Solution Structures in Mathematical Problem Problems

Kazuaki Kojima[a*], Kazuhisa Miwa[b] & Tatsunori Matsui[a]
[a]*Faculty of Human Sciences, Waseda University, Japan*
[b]*Graduate School of Information Science, Nagoya University*
*koj@aoni.waseda.jp

Abstract: Problem posing is identified as an important activity in mathematics education and a critical skill to be acquired. Several studies implemented support systems for learning of problem posing which aid novice learners in successfully posing appropriate problems. However, such learners may not necessarily success in posing appropriate problems without the support. Toward further support for novice learners in acquiring problem posing as a mathematical skill, we have to understand failures occurring in problem posing by novice learners. This study experimentally investigated problem posing by novices and empirically described their failures. In our investigation, participants were engaged in a learning task to study an example by reproducing it and a novel generation task to pose their own problems, with the results indicating that some participants composed problems whose texts and solutions were inconsistent in the learning task.

Keywords: Mathematical learning, problem posing, learning from examples

Introduction

Problem posing is identified as an important activity in mathematics education, as well as problem solving is [12, 13]. Although problem posing is rarely adopted in general education due to certain constraints in practical classrooms, it is as critical a skill as problem solving. One of the reasons why problem posing is unadopted may be that problem posing is extremely difficult for novice learners. Because problem posing is a production task that requires idea generation and synthesis of structures, it imposes heavy cognitive load on learners.

Several studies have addressed support for problem posing by learners. For example, some e-learning systems adopt problem posing as a learning task and aid it through the peer-assessment of learner problems [1, 4, 14, 16]. Hirashima and his colleagues implemented several systems that can evaluate problems posed by learners [3, 15]. Their environments offer computer-supported learning exercises to generate problems solved by specified solutions and to alter instance problems into new ones. These studies have also reported that learning with the systems improved learner understanding of domain knowledge or solution methods embedded in problems. Our previous studies proposed a support system that facilitates diverse problem posing through learning from examples [5, 6], and experimentally confirmed that our system could improve problem posing by learners to some extent.

Although problem posing is difficult, learners can successfully pose problems with support by the systems mentioned above. However, they may not necessarily success in posing appropriate problems without the support. Toward further support for novice learners in acquiring problem posing as a mathematical skill, we have to understand failures occurring in problem posing by novice learners. Leung and Silver [10] studied problem posing by prospective elementary school teachers and empirically obtained a certain number of non-mathematical or unsolvable problems, even though, their focus was not on analysis of such inappropriate problems.

This study experimentally investigated problem posing by novices and empirically described their failures. In the investigation, we used a learning task of problem posing proposed in the previous study [6].

1. Experimental Method

In our investigation, participants were engaged in two problem posing tasks. One of them was a learning task to reproduce a problem given as an example, and the other was a task to generate novel problems.

1.1 Tasks

In each of the experimental tasks, participants were required to generate one or more problems in the domain of a problem initially given as a base. In the first learning task, they were provided with a base and an example problem as a good response in the task, which was generated by altering the base. They were then asked to reproduce the example. When reproducing, the example itself was hidden and information indicating how to generate the example from the base was shown. The generation process information of the example was automatically generated by our support system implemented in the previous studies, which included sufficient information to reproduce the example. This activity in the task had been designed to provide novice learners with ideas feasible in composing novel problems through *imitation* of varied examples. We empirically confirmed that learners could successfully transfer what they learned from an example through imitation with the system into novel problem posing by the learners. Figure 1 indicates the basic framework of the learning task (For more detail on the support system, see [6, 9]).

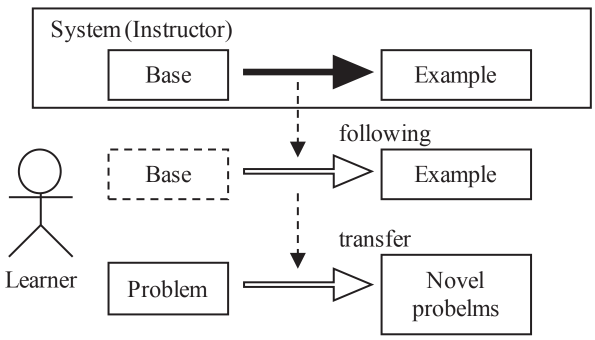


Figure 1. Basic framework of learning task

The learning task was followed by the novel generation task where the participants were asked to pose problems as many, varied and unique as possible from another base problem. In this task, participants' problems were evaluated based on four categories shown in Figure 2. These categories indicate similarities in *situations* and *solutions* between their problems and the base. Situations of problems denote surface features of contextual settings in problem texts (e.g., purchase of goods or transfer by vehicle), and solutions mathematical structures of the problems. Therefore, Category I / I indicates problems almost the same as the base, D / I indicates those generated by altering a situation of the base, I / D indicates those generated by altering a solution, and D / D indicates those generated by combining both alterations. It is desirable for mathematical learners to pose diverse problems across these categories controlling features of situations and solutions. However, previous studies revealed that leaner problems tend to lack diversity [2, 11] and have simple or inappropriate structures in their solutions [8].

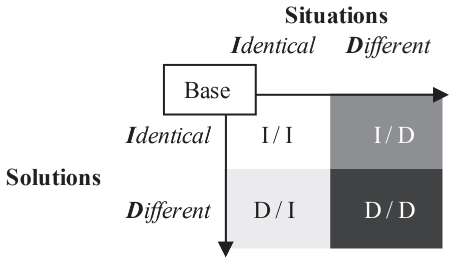


Figure 2. Categories for evaluating the variety of problems

1.2 Procedures

Undergraduates were participated in the investigation conducted in a lecture class of cognitive science. They were fist engaged in the learning task without the support system. The participants were told that the aim of the task was to learn what is problems posing and how it is done through an example before their own problem posing task. Prior to start of the learning task, the following problem was presented as a base.

> *Base*) I bought some 60-yen oranges and 120-yen apples for 1020 yen. The total number of oranges and apples was 12. How many oranges and apples did I buy?
> Solution.
> Let x denote the number of oranges and y denote the number of apples.
> $x+y=12$
> $60x+120y=1020$
> According to the equations above, $x=7$, $y=5$.

The base was printed in sheets of paper provided the participants. The participants were also presented the following problem as an example on a big screen of the classroom.

> *Example*) Last year I bought some 40-yen pencils and 110-yen pens. The total number was 13. This year I bought 2 times as many pencils as last year, the same number of pens as last year, and a 300-yen pen case for 1430 yen. How many pencils and pens did I buy last year?
> Solution.
> Let x denote the number of pencils and y denote the number of pens.
> $x+y=13$
> $40*2x+110y=1430-300$
> According to the equations above, $x=10$, $y=3$.

The example has the setting of purchase of goods identical to the base, and a solution formed by adding a third object other than x and y objects and an operation to calculate a coefficient of x in the lower equation to the base. Thus, it belongs to Category I / D in Figure 2.

When starting the task, the example was removed from the screen. The participants were asked to reproduce the example based on generation process information printed in the sheets. The generation process information contained the situation, numeric parameters appearing in the text, a basic structure of the solution, mathematical operations added in the solution, and keywords in the text of the example. It explicitly indicated that the situation of the example was identical to the base and the solution was altered. The participants were also instructed that they didn't have to completely literally reproduce words in the text of the

example as long as the contextual setting and equations in the solution were appropriately reproduced.
The participants were then engaged in the novel generation task. In this task, the following problem solved with a unitary equation was presented as a base.

Base) I want to buy some boxes of cookies. If I buy some 110-yen boxes of cookies, then I have 50 yen left. If I buy some 120-yen boxes of chocolate cookies, then I need 20 yen more. How many boxes do I want?
Solution.
Let x denote the number of boxes.
$110x + 50 = 120x - 20$
According to the above equation, $x = 7$

They were told that their problems had to be necessarily solved with unitary equations and any problems in other domains were unacceptable.

1.3 Data Analysis

Participants were classified into groups based on problems they reproduced in the learning task. The groups were as follows.

Reproduced Appropriately (R-A): succeeded in composing a problem whose contextual setting and solution were identical to the example
Reproduced Sufficiently (R-S): almost succeeded in composing a problem identical to the example but partially changed its contextual setting (actually, *2 times* in the problem was not used as the number of pencils, but as the price of a pencil)
Reproduced but Modified (R-M): succeeded in composing the same solution structure but partially changed its surface parameters (numerals and their objects)
Altered solutions (A): composed a problem whose solution was different from the example
Lacked parameters (L): didn't succeed due to absence of numeric parameters, such as 300 (the price of a pencil box) or 13 (the total number).
Inconsistently composed (I): didn't succeed due to inconsistency between a text and a solution of a problem composed, although the solution was identical to the example

Problems newly composed by the participants in the novel generation task were categorized into the four categories in Figure 2. We also analyzed problems posed by altering solutions, from the aspect of structural complexity. However, this paper doesn't present more detail on the results in the novel generation task due to limitations of space. They will be reported precisely in another paper.

2. Results

One hundred and thirty-two undergraduates participated in the investigation. In the results below, eight undergraduates who didn't complete the learning task were excluded.

2.1 Problems Reproduced in Learning Task

Figure 3 indicates the proportions of participants in each group in the learning task. Although half of the problems were appropriately or sufficiently composed, the others were different from the example in some ways.

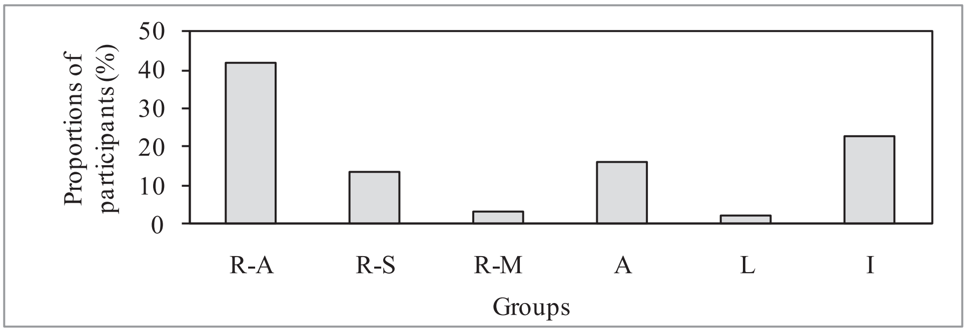


Figure 3. Proportions of reproduced problems in each group

2.2 Problems Posed in Novel Generation Task

In the following results, problems in domains different from the base and unsolvable problems were excluded. Figure 4 indicates the proportions of posed problems in each category in the novel generation task. "C" in the figure, denoting a *control* group, is the result of undergraduates who were engaged in the same novel generation task in the previous study [8] without learning of any example. This revealed that few problems in I / D were posed without supportive intervention. Although no significant differences between most of the groups and the control group were found due to the small numbers of participants, there was a significant difference between the R-A and control groups ($\chi^2(3)=15.29$, $p<.01$). Residual analysis revealed that the number of I / I problems was high in the control group and low in the R-A group ($p<.05$), and that of I / D was high in the R-A group and low in the control group ($p<.01$). Thus, appropriate reproduction of the example increased posed problems in I / D.

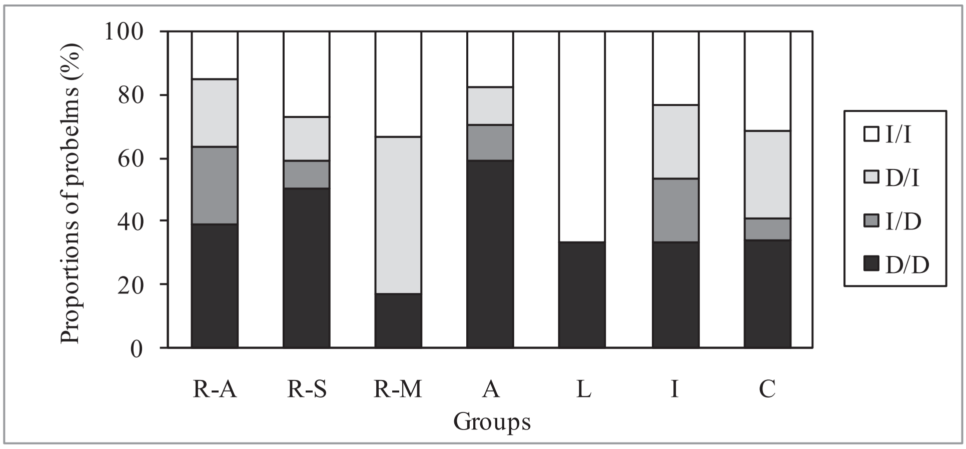


Figure 4. Proportions of posed problems in each category

Figure 5 indicates the proportions of problems posed by altering solutions whose operations in the solutions increased or decreased from the base. In the control group, half of the solution-altered problems were simpler than the base. The I group also posed many simple problems, whereas the R-A group posed many complex problems. There was also a significant differences between the R-A and control groups ($\chi^2(2)=11.36$, $p<.01$), and no difference between each of the other groups and the control group. Residual analysis revealed that the number of increase was high in the R-A group and low in the control group ($p<.05$), and that of decrease was high in the control group and low in the R-A group ($p<.01$). Thus, the appropriate reproduction also increased posed problems more complex than the base.

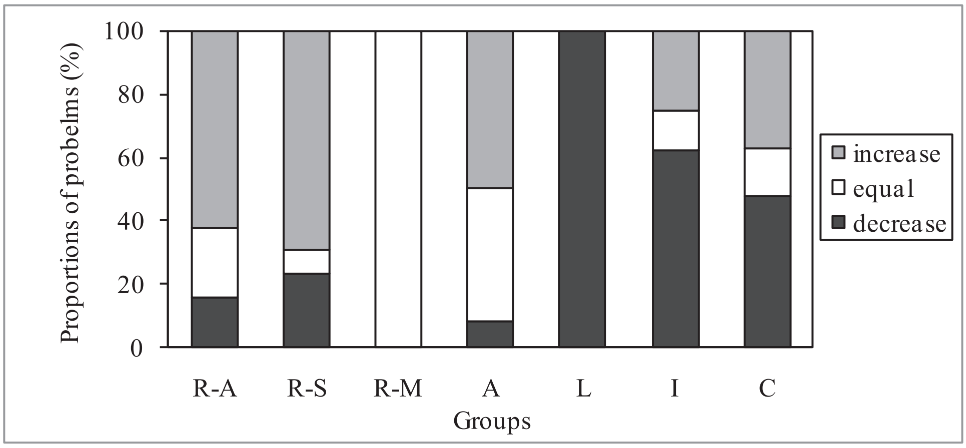


Figure 5. Proportions of solution-altered problems whose operations increased or decreased

3. Discussion

3.1 Failures in Reproduction

Despite sufficient information to reproduce the example was provided in the learning task, not more than half of the participants exactly reproduced it. The R-S and R-M groups didn't exactly do, though, their problems had solution structures identical to the example. Thus, it can be regarded that the two groups almost succeeded in the reproduction.

The A group didn't reproduce the example but composed problems different from the example. Therefore, the participants in the groups must have merely misunderstood the instruction in the learning task.

The participants in the I groups failed in reproducing the example. Although they described the same solution as the example, their problem texts were inconsistent with the solution. In the texts, some mathematical relationships were incorrectly described or inappropriate relationships were included[1] so that the solution was never formulated from the texts. Therefore, the participants didn't understand the inconsistency. Of course, none of the participants must fail in solving the example, which is a quite simple problem for undergraduates. We preliminary confirmed that undergraduates can successfully solve it [7].

The L group also failed in the reproduction. However, all of their problems could be completed by adding a description such as "the total number of the pencils and pens was 13". Thus, the participants must have carelessly forgotten to include some numerals into their problem texts.

As described in Section 1.1, the reproducing task adopted in this investigation is used in our support system [6]. No participants failed in the same reproduction task in an experimental evaluation of the system [9], although a few participants composed problems different from the example like the A group did. According to the facts, novice learners who successfully poses problem with supporting intervention can fail in appropriate problem posing without the intervention. Another important insight is that novice learners occasionally pose problems whose texts and solutions are inconsistent. To improve problem posing of novice learners, hence, further support is needed to endow the learners with a skill to appropriately compose problems.

[1] Some examples of problems in the learning task are presented in Appendix.

3.2 Novel Problem Posing after Learning

As described in Section 2.2, learning through appropriate reproduction of the example increased posed problems in I / D. It also increased problems whose solutions were more complex than the base, because the example allowed the participants to learn how to add operations. These results in the current study are consistent with experimental evaluation of the support system in the previous study [9]. On the other hand, sufficient learning effect wasn't gained through inappropriate reproduction.

According to the results in Figures 4 and 5, half of problems posed by the I group were in I / D or D / D, which fact indicates that many of the problems had solutions different from the base. The I group varied solutions in their problem posing to some extent. However, such problems in the I group were mostly simpler than the base which was quite simple and elementary. Therefore, the participants in this group didn't thoroughly learn the example, although they were examined it.

4. Conclusions

This study experimentally investigated problem posing by novices and empirically described their failures. In our investigation, participants were engaged in a learning task to reproduce an example and a novel generation task to pose their own problems, with the results indicating that some participants composed problems whose texts and solutions were inconsistent, in other words, they failed in reproduction. Our next task is, of course, to study and design a supporting method to prevent such failure in the learning activity.

Acknowledgment

This study was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B), 23700990.

References

[1] Barak, M., & Rafaeli, S. (2004). On-line Question-Posing and Peer-Assessment as Means for Web-Based Knowledge Sharing in Learning. International Journal of Human-Computer Studies, 61, 84-103.

[2] English, L. D. (1998). Children's Problem Posing within Formal and Informal Contexts. Journal for Research in Mathematics Education, 29, 83-106.

[3] Hirashima, T., Yokoyama, T., Okamoto, M., & Takeuchi, A (2007). Learning by Problem-Posing as Sentence-Integration and Experimental Use. In R. Luckin, K. R. Koedinger & J. Greer (Eds.) Proceedings of the 13th International Conference on Artificial Intelligence in Education (pp. 254-261). Amsterdam, Netherlands: IOS Press.

[4] Hirai, Y., Hazeyama, A., & Inoue, T. (2009). Assessment of Learning in Concerto III: A Collaborative Learning Support System Based on Question-posing, In V. Uskov (Ed.) IASTED International Conference on Computers and Advanced Technology in Education (pp. 36-43). Calgary, Canada: ACTA Press.

[5] Kojima, K., & Miwa, K. (2008). A System that Facilitates Diverse Thinking in Problem Posing. International Journal of Artificial Intelligence in Education, 18, 209-236.

[6] Kojima, K., Miwa, K., & Matsui, T. (2009). Study on Support of Learning from Examples in Problem Posing as a Production Task. In S. C. Kong et al. (Eds) Proceedings of 17th International Conference on Computers in Education (pp. 75-82). Jhongli, Taiwan: Asia-Pacific Society for Computers in Education.

[7] Kojima, K., Miwa, K., & Matsui, T. (2010). Experimental Study on Methods of Learning from Examples and their Effects in Problem Posing. 62th SIG on Advanced Learning Science and Technology in Japanese Society for Artificial Intelligence, SIG-ALST-B001,13-18. (In Japanese)

[8] Kojima, K., Miwa, K., & Matsui, T. (2010). An Experimental Study on Support for Leaning of Problem Posing as a Production Task. Transactions of Japanese Society for Information and Systems in Education, 27, 302-315. (In Japanese)
[9] Kojima, K., Miwa, K., & Matsui, T. (*in press*). Study on the Effects of Learning Examples through Production in Problem Posing. 19th International Conference on Computers in Education.
[10] Leung, S. S., & Silver, E. A. (1997b). The Role of Task Format, Mathematics Knowledge, and Creative Thinking on the Arithmetic Problem Posing of Prospective Elementary School Teachers. Mathematics Education Research Journal, 9, 5-24.
[11] Mestre, J. P. (2002). Probing Adults' Conceptual Understanding and Transfer of Learning via Problem Posing. Journal of Applied Developmental Psychology, 23, 9-50.
[12] Polya, G. (1945). How to Solve it. Princeton, NJ: Princeton University Press.
[13] Silver, E. A. (1994). On Mathematical Problem Posing. For the Learning of Mathematics, 14, 19-28.
[14] Takagi, M., & Teshigawara, Y. (2006). A WBT System Enabling to Create New or Similar Quizzes Collaboratively by Students. In J. R. Parker (Ed.) IASTED International Conference on Education and Technology (pp. 263-268). Calgary, Canada: ACTA Press.
[15] Yamamoto, S., Waki, H., & Hirashima, T. (2010). An Interactive Environment for Learning by Problem-Changing. In S. L. Wong, S. C. Kong & F. Yu (Eds) Proceedings of the 18th International Conference on Computers in Education (pp. 1-8). Malaysia: Universiti Putra Malaysia.
[16] Yu, F., Liu, Y., & Chan. T. (2005). A Web-based Learning System for Question-Posing and Peer Assessment. Innovations in Education and Teaching International, 42, 337-348.

Appendix: Examples of problems in Learning Task

A group

I bought some 100-yen apples and 30-yen bananas. The total number was 13. I gave a 1000-yen bill and received 190 yen as the change. How many apples and bananas did I buy?

Solution.

Let x denote the number of apples and y denote the number of bananas.

$x+y=13$

$100x+30y=1000-190$

According to the equations above, $x=6$, $y=7$.

I group

Last year I bought some 40-yen pencils and 110-yen pens. The total number was 13. <u>This year I also bought 13 pencils and pens</u>. The number of pencils this year was 2 times as many as last year. In addition to pencils and pens, I bought a 300-yen pencil box. The payment was 1430 yen. How many pencils and pens did she buy?

Solution.

Let x denote the number of pencils and y denote the number of pens.

$x+y=13$

$40*2x+110y=1430-300$

According to the equations above, $x=10$, $y=3$.

(The total number this year is wrong)

A girl bought pencils and pens. The total number was 13. <u>The number of pencils was 2 times as many as pens</u>. A pencil was 40 yen and a pen was 110 yen. She found a 300-yen lovely pencil box near the cash desk, and took it with pencils and pens. The payment was 1430 yen. How many pencils and pens did she buy?

Solution.

Let x denote the number of pencils and y denote the number of pens.

$x+y=13$

$40*2x+110y=1430-300$

According to the equations above, $x=10$, $y=3$.

(Parameters associated with the relationship "2 times" are wrong)

Mohd Ayub A. F. et al. (Eds.) (2011). Workshop Proceedings of the 19th International Conference on Computers in Education. ChiangMai, Thailand: Asia-Pacific Society for Computers in Education.

A Phone-Based Question Management System to Facilitate Questioning and Comprehension Monitoring

Po-Yao CHAO[a*], Howe-Yii LING[b] & Baw-Jhiune LIU[c]
[ab]*Department of Information Communication, Yuan Ze University, Taiwan*
[c]*Department of Computer Science & Engineering, Yuan Ze University, Taiwan*
[*]poyaochao@saturn.yzu.edu.tw

Abstract: Research showed that only very small proportion of questions in class was posed by students. Students tended to ignore the questions encountered in class. They were unlikely to actively tackle their confusion or questions. Therefore, the goal of this study is to develop a mobile phone based questioning management system. By integrating mobile phones, course materials with online discussion forum, the system can assist students to ask, trace, monitor and solve comprehension questions encountered in class or after class. From functional perspective, in class or after class when studying, students can: (a) capture contents of paper slides or online courseware, assemble the contents with questions as a whole, and post the assembled questions on a discussion forum via mobile phones; (b) monitor and track the status of post questions via mobile phones as well as organize their personal notes based on the questions during the processes. From the result of formative evaluation for the question management system, students thought the mark mechanism helped questioner enhance the clarity of the questions and thought the progress icons helped them monitor question resolving state and comprehension regarding instructional materials.

Keywords: Comprehension monitoring, questions management system, mobile learning

Introduction

Effective learners are sensitive to their knowledge deficits. They adopt self-regulatory strategies to improve their knowledge deficits [8]. Research also showed that effective learners can monitor and correct their failures in comprehension [16]. Students inevitably will confront with difficulties regarding course contents or teaching materials. In many cognitive models, questions and confusion are the foundation of the text understanding and social behavior [9], as well as the basis of problem solving [10]. However, according to research, a very small proportion of questions were asked by the students in the classroom settings [8]. Students seldom ask questions or take meaningful strategies during or after the classes. Many factors and costs affect a student's motivation to ask [13], which Graesser [8] pointed out that physical and social factors affect the way in which students deals with their confusion. Physical factors include the gap between the questioner and the answerer. Social factors include the colleague's negative feelings towards the questioner. Only when students break the social and physical barriers, the students will have the motivation to ask questions and try to solve the questions.

Many studies has been attempted to enhance student's quality of questions and the motivation to ask questions through computer supported systems. Among these systems, a discussion forum is the most common and popular components [11]. Some Web-based or mobile discussion forum systems provided students with a place to review and express their confusion regarding instructional materials [e.g., 12, 14]. Student discussed together on a forum and resolved their questions collaboratively. However, the design goal of these Web- or phone-based forums seems mainly for discussion but not for the support of question solving. Mechanisms that help learners monitor and evaluate the status of questions were not explicitly supported. Several studies also employed mobile devices to facilitate information organization [e.g., 6], collaboration [e.g., 17], or communication with messages [e.g., 5]. However, few of these tools emphasized the integration of mobile devices with paper-based materials and supported the process of question resolving.

Q&A (Question and Answering) system [e.g., 1] provides a place for students to find answers based on questions. Some of the systems are featured in learning communities and others provide FAQ lists. These systems provide well-designed knowledge structures to help learners find out the answers of questions. These structures also require learners to have certain skills or background knowledge so that they can find target answers effectively. On the other hand, class communication tools [e.g., 3] provides an interactive channel between teachers and students in class. These tools assist teachers to instantly assess students' learning status and adjust the teaching strategies accordingly. However, this kind of tools generally requires dedicated equipments or settings.

Although resolving question has become a major role in student knowledge acquisition process, most computer supported systems seldom emphasize the process of the questioning resolving. Practical classroom environment seldom guarantee personal computers for every students. This will leads to a problematic situation that student's confusion and context of their questions are difficult to be captured. With advance of mobile and computer technologies, mobile phones become popular and create possibilities of ubiquitous learning. Due to the requirements of questioning in classroom environments, the goal of this study is to develop a mobile phone based questioning management system. By integrating mobile phones, course paper-based materials with online course-based discussion forum, the system can assist students to ask, trace, monitor and solve comprehension questions encountered in class or after class. From functional perspective, in class or after class when studying, students can: (a) capture contents of paper slides or online courseware, assemble the contents with questions as a whole, and post the assembled questions on a discussion forum via mobile phones; (b) monitor and track the status of post questions via mobile phones as well as organize their personal notes based on the questions during the processes.

1. Stages for questioning management

When students confront with difficulties and attempt to solve them, search for answers, self-regulation and decision making are involved. To ask a question, Graesser [8] suggested three components of question asking: anomaly detection, question articulation, and social editing. In this paper, we based on Graesser's model and extend it with self-regulatory components for question asking and management. As shown in Figure 1, regarding a question, suggested stages are confusion discovery, question formulation, question announcement, response evaluation, and Q&A organization. For the confusion discovery stage, according to Piaget's cognitive-developmental theory, the mismatch between external information and a learner's internal knowledge structures causes cognitive disequilibrium [15]. This mismatch initiated by new information generates confusion. This stage corresponds to anomaly detection component of Graesser's questioning model.

If the confusion discovered cannot be transformed into a form of questions, it will be difficult to assess and manage. Therefore, question formulation not only encourages students to retrieve their prior knowledge but also enables devices for students to evaluate or construct knowledge base on the questions. This stage corresponds to question articulation component of Graesser's model and requires support to help students formulate questions with clarity. For the question announcement stage, students uncover their questions to a learning community. According to Vygotsky's zone of proximal development, student's cognitive development requires help from more capable peers or teachers [15]. The display of questions may enhance the interaction, discussion, or collaboration among classmates or teachers. This stage corresponds to social editing component of Graesser's, which suggests that a questioner evaluates the benefit and cost of questioning in order to decide whether to raise questions for help in the public. This stage requires the support to lower the cost and increase the benefit of questioning.

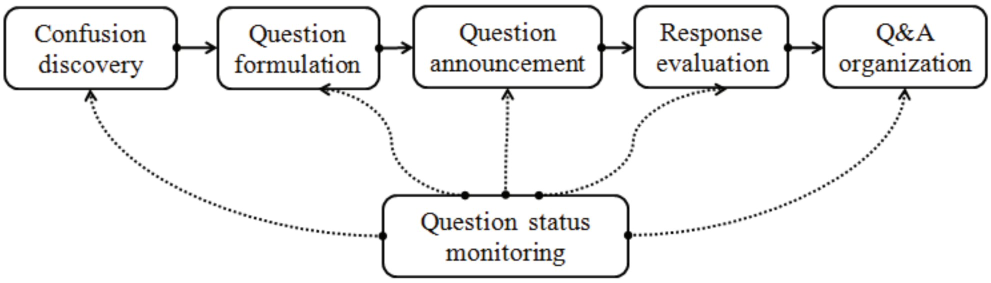


Figure 1. Process of question monitoring

Feedback for questions from peers and teachers is crucial for questioners to evaluate whether they have resolved the questions. During the evaluation, the questioners themselves have to critically assess the contribution of replies, which benefits the construction of knowledge [2]. The response evaluation stage also requires questioners synthesize from difference sources and filter out irrelevant information. Finally, the Q&A organization stage provides students with support that enables connections between questions and corresponding answers. Based on the results of question evaluation stage, students draw conclusions from relevant feedback. Research found that making notes and summary help students establish their own knowledge architecture [4], and create effective knowledge structures. The proposed questioning stages make question-resolving explicit and systematic. Students can follow the suggested steps and expect the arriving tasks for the planning and management of their learning resources. As showed in Figure 1, five-stage question resolving framework is used to develop the phone-based question management system.

2. System design

Figure 2 shows the system architecture of the question management system. The system incorporates mobile phones, paper-based slides, and online discussion forum to facilitate students to trace, monitor and to solve their comprehension questions. The paper-based slides are embedded with unique barcodes so that students can use a camera-based phone to scan the codes and formulate their questions regarding specific confusing areas of slides. The questions together with confuse-marked slides can be post on an online discussion forum. Trough discussion by other students, feedback and answers are gathered and stored in a learning portfolio database. Students can monitor the current status and know the latest feedback and questions updates. Students can evaluate the feedback of questions which are most relevant and mark the feedback as useful messages. Finally, students organize questions and answers for future reference.

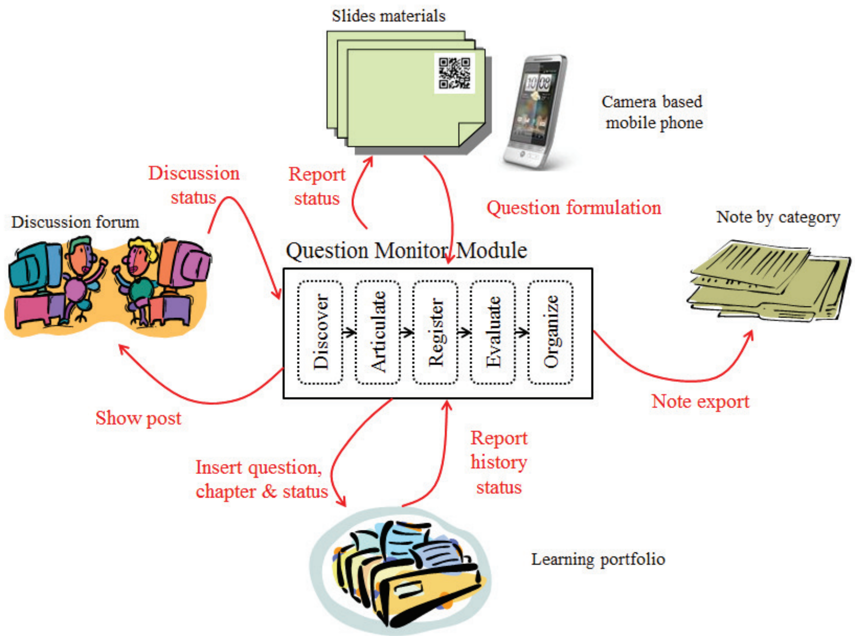


Figure 2. System architecture

2.1 Support for question asking

On a tradition discussion forum, a questioner often has to descript the question clearly. But most of the person who asked is only a beginner and the knowledge is limited. Some student thinks that asking directly with a book is much faster and easier because they can point out the question and describe using symbol and draft. We take this scenario into a classroom, we need to ensure that when students face difficulties they can capture the difficulties immediately and transform them into questions which can be discuss with others. For this point some students may only need examples to understand or they totally don't understand with the content of lecture slides. As shown in Figure 3(a), a lecture slide evenly divided into 9 regions, each region is selectable by touching the regions. The marked regions are highlighted with green color. Students can mark the areas where the students feel confused. They can either find related question on this slides or create a new question based on the marked areas.

With these marks students can intuitively know where they are confused and it is helpful for other students who would like to answer the questions. When the student click the "Ask a question" button, another form for question formulation appears (Figure 3b). A question bank is prepared for students to choose suitable questions. This saves time for students as mobile phones may slow down text input process. After choose the most suitable question the student can add-on additional description for questions at the box provided. At the left side of the screen the marked slide is sown for reference. The students have the choice to keep the question to their self or post out to public area for discussion. After posting questions, students can see their questions are shown on a discussion forum from a question register list (Figure 3c). The system automatically sets the title as a combination of question description, the details of the questioner and the marked slides. Students can freely switch the display of questions between traced and non-traced questions.

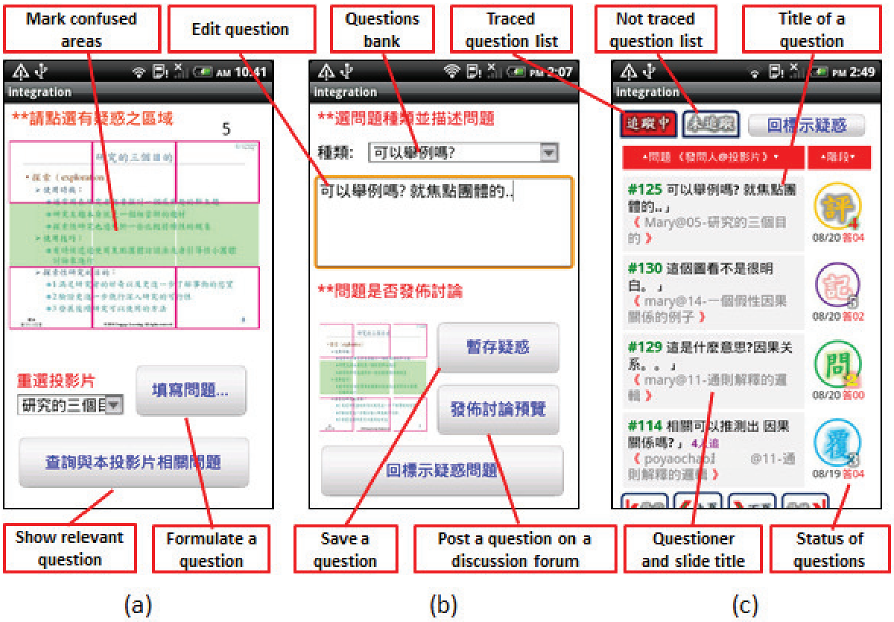


Figure 3. (a) Mark confused areas of a slide; (b) formulate questions based on marked regions; and (c) a list of questions with status icons

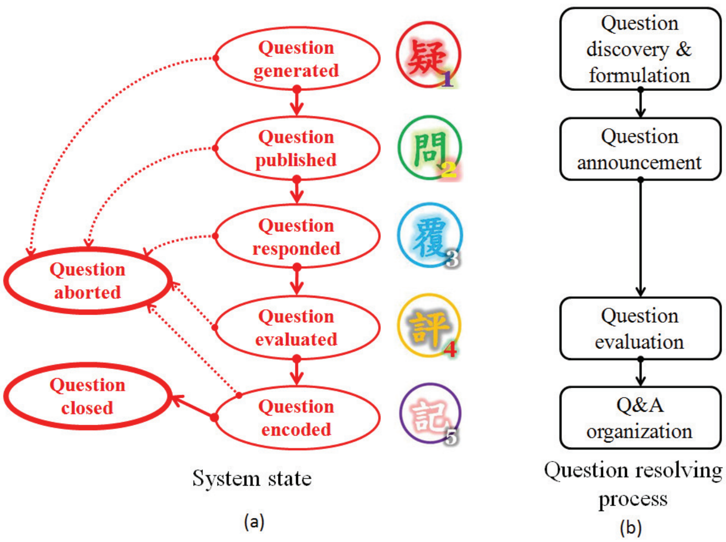


Figure 4. System state and question resolving process

2.2 Support for question resolving and monitoring

Students can monitor the progress of their question as shown in Figure 3(c). Question status represented in the system are categorized into 5 states (Figure 4a). Except the question responded state, each stage in question resolving process is initiated by questioners and corresponds to one state represented in the system. For example, the question generated state stands for the time when students mark down where they are confused and formulate corresponding questions. Therefore, from student's discovery of confusion to organization of Q&A notes, the system has corresponding state in every stage (Figure 4b).

Figure 5(a) shows a student respond to a question. Students name and time of responses will be recorded when a student makes a post. Students can choose to reply the post or evaluate the post when enough post is available. Student who asked the question can delete the question if he/she found out that the questions is not suitable. Students who trace a question can stop tracing the question if no longer of interested. Figure 5(b) shows a post being selected as useful. When the questioner decides to evaluate the answer, he/she will choose appropriate answers from different students and mark it as useful. These answers then will be transfer to Figure 5(c) which shows the editing of the selected post and being transformed to a note.

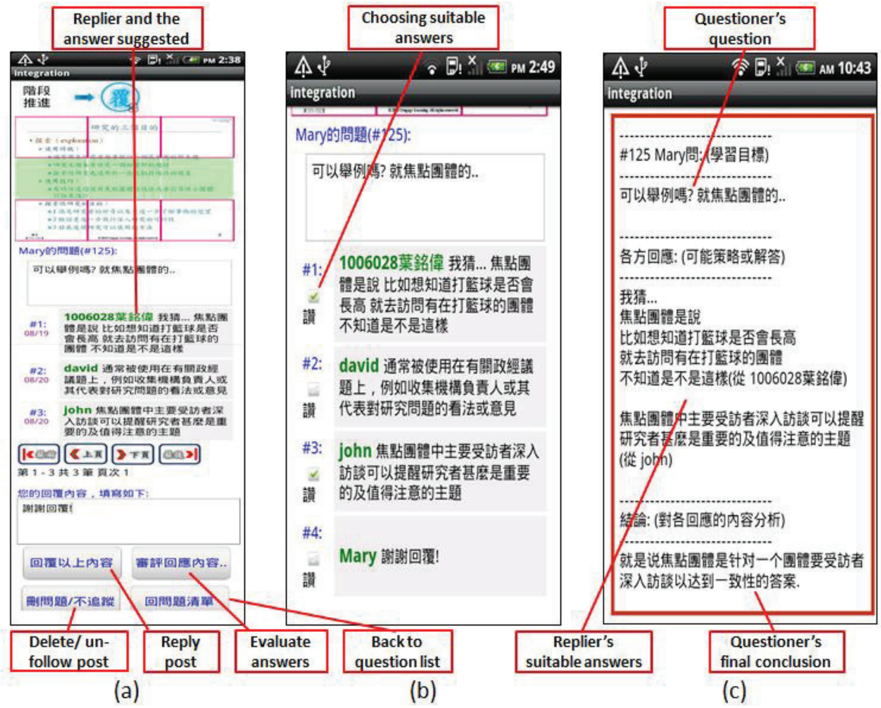


Figure 5. (a) Respond to questions; (b) evaluate replies from different sources; and (c) organize questions and useful replies

3. Evaluation

A formative evaluation was conducted to examine the system's usability and it's implication on question resolving. Flagg [7] stated that formative evaluation is one of the most critical steps in the development of learning materials. The goal of the formative evaluation is to help system designers during its early development stages improve system's quality.

3.1 Participants and procedure

We adopt user-based evaluation that involves users completing tasks in an appropriate environment. In this study, five graduate students aged around 23 (4 male and 1 female) majoring in information communication in a university participated in the formative evaluation. Each participant was given a camera-enabled mobile phone. These mobile phones can read barcodes and connect with the Internet through wireless networking functions. Paper-based slides included 50 A4 printed pages and each page is embedded with two dimensional barcodes. Participants were asked to study the paper-based slides during one week evaluation. They are also asked to use the question management system for comprehension question resolving.

Participants were asked to write down operational difficulties and opinions in diary. At the end of the evaluation, participants are asked to perform specific tasks and interviewed. Therefore, observation, interview, and computer logs were collected for analysis.

3.2 Question formulation

There are 43 questions posted by participants using the question management system during the evaluation (9 in stage 2, 13 in stage 3, 18 in stage 5, and 2 questions been deleted). Overall there are 70 replies message during the one-week evaluation and a total of 33 notes has been generated (15 duplicated notes which copied by other students).

From question formulation aspect, three out of 5 students used to generate questions without posting it on discussion forum first. They thought that the clarity of the questions was important before announcement of questions. They would make sure that other people could understand what they were asking about. For example, one of the student states "I usually generate a question whenever I think the confusion exist without posting it out first. After finishing the chapter, I will then refine the question…" The register of questions helps them find answers from rest of the chapter and then refine the questions.

We also found that all participants thought the mechanism of marking confused regions would benefit them. The mechanism helps them shorten the questions when asking. Since there is a marked slides aided, the question would be expressed in a more clear and specific way. Moreover, with the help of marked slides, students can better understand other students' questions and provide feedback. This mechanism is especially beneficial when asking questions regarding figured-based slides.

3.3 Question resolving and monitoring

Representing questioning stages by icons provides students with guidance in question resolving and helps them develop a processing priority. Students finished question resolving stage by stage. These progress icons assisted students to predict incoming stages and the course toward their goals. For example one student said "I would go from post stage to evaluation stage and then create my notes…" The progress icons also helped students develop personal processing priority and adopt different strategies accordingly. For instance, one student said "I would first read questions which have reached the final stage because the questions usually have specific conclusions. Then I will read questions that most students have responded to them…"

Students also used the progress icons to assess the degree of comprehension to the lecture slides. The overall state of the questions may also influence students' confidence to the mastery of lecture slides. Most students reported that the more questions in organized stages, the more confidence to a test they can have. On the other hand, if most of a student's questions still stay in question published stage and no followers trace the student's questions. The student will come to a conclusion that he/she may post questions with low quality or importance.

All participants' thought the questions organized as notes was beneficial to questioner and the students tracing them. Four out of 5 participants reported that they would copy their notes in the system and annotate the notes on their textbooks or slides. They all confirmed that these notes are useful when preparing for a test.

4. Conclusion

The goal of this study is to develop a mobile phone based questioning management system that assist students to ask, trace, monitor and solve comprehension questions encountered in class or after class. Students use mobile phones to capture contents of paper slides, assemble the contents with questions as a whole, and post the assembled questions on a discussion forum. The system tracks questions stages and provides students with progress icons to help monitor and trace the status of post questions as well as organize their personal notes based on the questions during the processes.

From the result of formative evaluation, students thought the mark mechanism could help questioner enhance the clarity of the questions. This is especially useful for novice students with limited background knowledge. Students also thought the progress icons could guide them to the final stage of question resolving and help them assess the comprehension regarding instructional materials. Some students provided useful suggestion for the system. First, a summative status report was suggested. Students could know the distribution of question stages and tackle with different strategies accordingly. Second, a function of anonymous reply seemed useful for students to provide advice or feedback. This is especially beneficial for students who would like to contribute their opinions but fear to make mistakes.

Acknowledgements

We appreciate the work of our anonymous reviewers in suggesting revisions for this paper. This research was partially funded by the National Science Foundation under NSC 99-2511-S-155-003-, NSC 100-2511-S-155-007-, NSC 100-2632-S-155-001-, NSC 100-2631-S-008-004-, and NSC 97-2511-S-008-006-MY3.

References

[1] Ackerman, M.S. and McDonald, D.W. (1996). Answer Garden 2: Merging organizational memory with collaborative help. In *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work* (pp. 97-105). Boston, Massachusetts, United States.

[2] Anderson, L.W. and Krathwohl, D.R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objects*. New York: Longman.

[3] Brotherton, J.A. and Abowd, G.D. (2004). Lessons learned from eClass: Assessing automated capture and access in the classroom. *ACM Transactions on Computer-Human Interaction, 11*(2), 121-155.

[4] Caverly, D.C., Orlando, V.P., and Mullen, J.-A.L. (2000). Textbook study reading. In R.F. Flippo and D.C. Caverly (Eds.), *Handbook of College Reading and Study Strategy Research* (pp. 105-148). London: Lawrence Erlbaum Associates.

[5] Chen, G.D., Chang, C.K., and Wang, C.Y. (2008). Ubiquitous learning website: Scaffold learners by mobile devices with information-aware techniques *Computers & Education, 50*(1), 77-90.

[6] Corlett, D., Sharples, M., Bull, S., and Chan, T. (2005). Evaluation of a mobile learning organiser for university students. *Journal of Computer Assisted Learning, 21*, 162-170.

[7] Flagg, B.N. (1990). *Formative evaluation for educational technologies*. New Jersey: Lawrence Erlbaum associates.

[8] Graesser, A.C. and McMahen, C.L. (1993). Anomalous information triggers questions when adults solve quantitative problems and comprehend stories. *Journal of Educational Psychology, 85*(1), 136-151.

[9] Hilton, D.J. (1990). Conversational processes and causal explanation. *Psychological Bulletin, 107*(65-81).

[10] Reisbeck, C.K. (1988). Are questions just function calls? *Question Exchange, 2*(17-24).

[11] Salter, G. (2000). Making use of online discussion groups. *Australian Educational Computing, 15*(2), 5-10.

[12] Scardamalia, M. and Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of The Learning Sciences, 3*(3), 265-283.

[13] van der Meij, H. (1988). Constraints on question asking in classrooms. *Journal of Educational Psychology, 2*(232-238).

[14] Wei, F.-H. and Chen, G.-D. (2006). Collaborative mentor support in a learning context using a ubiquitous discussion forum to facilitate knowledge sharing for lifelong learning. *British Journal of Educational Technology, 37*(6), 917-935.

[15] Woolfolk, A.E. (1995). *Educational Psychology* (7th ed.). Boston: Allyn and Bacon.

[16] Zimmerman, B.J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology, 81*, 329-339.

[17] Zurita, G. and Nussbaum, M. (2004). A constructivist mobile learning environment supported by a wireless handheld network. *Journal of Computer Assisted Learning, 20*, 235-243.

Mohd Ayub A. F. et al. (Eds.) (2011). Workshop Proceedings of the 19th International Conference on Computers in Education. ChiangMai, Thailand: Asia-Pacific Society for Computers in Education.

Stepwise Selection of English Multiple-choice Cloze Questions Based on Difficulty-based Features for Keeping Motivation

Tomoko KOJIRI[a*], Yuki WATANABE[b] & Toyohide WATANABE[b]
[a]*Faculty of Engineering Science, Kansai University*
[b]*Graduate School of Information Science, Nagoya University*
* kojiri@kansai-u.ac.jp

Abstract: We had constructed the automatic generation system of English multiple-choice cloze questions. By using the system, plenty number of questions can be generated automatically, but learners become difficult to find appropriate questions. Therefore, objective of this research is to develop the method for providing questions that fit for learners. In the self-learning, questions that increase motivation for learners are effective. This research determines features that affect to difficulties of questions (difficulty-based features) and proposes the method for selecting questions according to the difficulty-based features for the stepwise learning. In order to manage the relations among questions, a question network is introduced in which questions are structured based on differences of each difficulty-based feature. Questions are selected by following appropriate links according the learners' answers..

Introduction

Multiple-choice cloze questions are often used in English learning. Such type of question is effective for checking the knowledge of grammar and lexicon. In addition, by tackling these questions repeatedly, the knowledge of English grammar and lexicon is able to be acquired. Only limited number of knowledge is included in one question, so many questions are needed to be solved for the purpose of acquiring whole grammar and lexicon knowledge. We have constructed the automatic generation system of English multiple-choice cloze questions; MAGIC [1]. By using the system, multiple questions can be generated automatically. However, to fit questions to learners' understanding situation is not focused. If difficult questions are posed to learners repeatedly, they do not feel like studying with the system for a long time.

Since questions contain plenty knowledge of grammar and lexicon, it is sometimes difficult to determine acquired/ in-acquired knowledge. In addition, learners' motivation is affected by their feelings whether they think "difficult" or "easy" for the questions. Such feelings may arise from the superficial features of questions. If the features of questions indicate that the question is too difficult, learners do not feel like tackling the questions. If the question is too easy, learners think questions are meaningless for them.

Traditional Intelligent Tutoring System or computer-adaptive testing tends to provide learning con-tents/test items that are appropriate for learners' understanding knowledge [2-4]. These systems analyze learners' acquired/in-acquired knowledge from their learning activities, such as answers of exercises. However, questions that are selected based on such knowledge-based approach do not always keep learners' motivation. Therefore, the objective of this research is to develop the method that provides questions based on the features that affect to learners' motivation (difficulty-based features). Difficulty-based features consist of more than one feature, and learners' feelings for these features may be different for each learner. So, the basis for selecting questions should be dynamically changed according to the learners.

Currently, target questions are questions that are generated automatically by MAGIC. So, the difficulty-based features need to be acquired systematically from the generated questions. Target learners are non-native speakers who do not understand basic grammatical knowledge.

1. Difficulty-based Features of English Multiple-choice Cloze Question

Figure 1 is an example of English multiple-choice cloze questions. A question consists of *sentence*, *blank part*, and *choices*. Choices include one correct choice and three distracters. Learners select one from choices for filling in the blank part.

There are various definitions or findings about difficulty features of English questions. Kunichika et al. defined difficulty features of English passage reading questions for non-native speakers as difficulties of *understanding of original texts*, *understanding of question sentences*, and *understanding of answer sentences* [5]. In English multiple-choice cloze questions, both original text and questions sentence correspond to question sentence, and answer sentences correspond to distracters. Therefore, following difficulty-based features are defined.

1) Difficulty of sentence --- Readability is one of the features that prevent learners of understanding the meaning easily. Researches about readability of English sentences insisted that lengths of sentences or difficulties of words affect to the readability [6]. Based on this result, *lengths of sentence* and *difficulties of words* are defined as one of the difficulty-based features of a sentence.

2) Difficulty of distracters --- There are various relations between distracters and a correct choice. In some questions, all distracter types are the same. The number of the distracter types affects to the difficulty of questions. If all distracter types are the same, it is easier to find the correct choice. On the other hand, questions become more difficult if similar types of distracters exist in it. Therefore, *the number of distracter types* in choices is defined as a difficulty-based feature. As the distracter types, 12 types defined in MAGIC are applied.

2. Question Selection Method Based on Question Network

Learners are motivated to learn repeatedly if difficulties of questions become gradually increasing. If questions seem easy, learners feel that they cannot acquire new knowledge from it. Appropriate questions for learners should contain a little difficult difficulty-based feature than those in solved questions.

In order to represent the stepwise relations among questions, a question network is introduced that organizes all questions based on difficulty levels for each difficulty-based feature. In the question network, questions in the same levels for all difficulty-based features form one node, and nodes whose levels are next to each other are connected by links. By following this question network, learners are able to tackle questions from easier one to more difficult one according to their understanding levels. Figure 2 illustrates the conceptual framework of the question network. Nodes without incoming links correspond to the easiest questions. Nodes without outgoing links have the most difficult questions.

The levels of each difficulty-based feature are defined as follows.

- Length of sentence---The number of words is regarded as one of the viewpoints of defining the length of sentence. Based on the analysis of 1500 questions in the database of our laboratory, it is revealed that sentences consist of 4 to 32 words. Thus, we

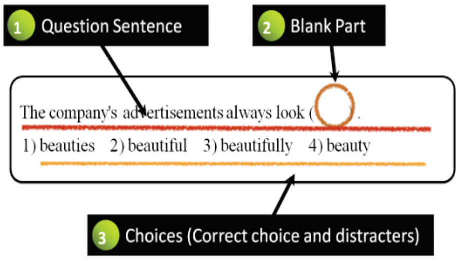


Figure 1: Example of English multiple-choice cloze question

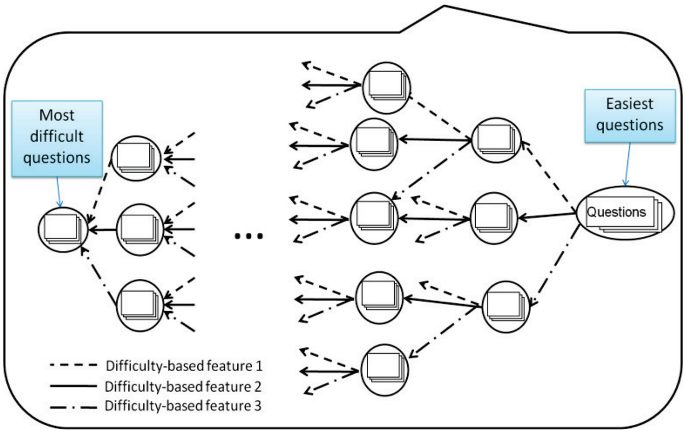


Figure 2: Conceptual framework of question network

categorize the length of sentence into four levels according the number of words. Table 1 shows the levels of the length of sentence.

- Difficulty of words---In this research, the difficulty of words followed SVL12000[7], which is the list of word difficulties defined by ALC. In SVL12000, 12000 words that are useful for Japanese are classified into five levels of difficulty. The level of a question is defined as the highest level in all words including the sentence and choices.
- The number of distracter type---Distracter types correspond to generation rules to generate distracters in MAGIC. Since choices of the same distracter types may be more difficult than that of the different one, the difficulty based on the number of distracter type is set as Table 2

Table 1: Levels of length of sentence

Level	1	2	3	4
# of words	less than 11	12 to 18	19 to 25	more than 26

Table 2: Levels based on the number of distracter type

Level	1	2	3
# of distracter types	3	2	1

Using the question network, learners' next questions are selected based on the answers of former questions. Figure 3 shows the process of selecting questions from question network. Currently, we assume that the set of questions is given at one learning.

STEP1: Based on the answers for questions in the last learning, learners' levels for each difficulty-based feature are determined. Levels for each difficulty-based feature *i* for time *t* are calculated as Equation 1. The average differences of solved questions from current level are added to the current level. In the first learning, *Level (i, t-1)* is zero.

$$Level\ (i,t) = Level\ (i,t-1) + Av.distance\ of\ solved\ questions\ from\ current\ node \quad \ldots (1)$$

STEP2: The number of solvable nodes becomes large if the learner solved questions in farther node, while it becomes small if the learner only could solve the questions in the nearer nodes. The range of solvable nodes at time *t* is calculated by Equation 2.

$$Range(t) = Range(t-1) + \left(\begin{array}{l}Ave.distance\ to\ solved\ questions - \\ Ave.distance\ to\ incorrectly\ answered\ questions\end{array}\right) \quad \ldots (2)$$

STEP3: Questions are selected from several solvable nodes. More questions should be selected from nodes that are nearer to the learner's current node. The probabilities for selecting questions for each node *i* is calculated by Equation 3. The ratio of questions from node *i* in all questions is determined by following the normal distribution based on the distance from the current node.

$$Probability(i) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{Distance\ of\ node\ i\ from\ current\ node)^2}{2}\right) \quad \ldots (3)$$

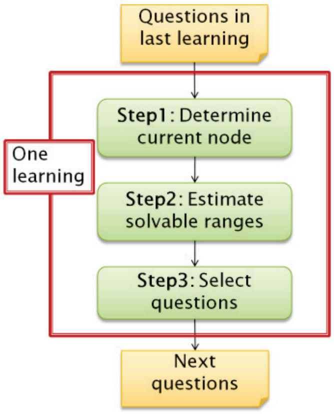


Figure 3: Process of selecting questions

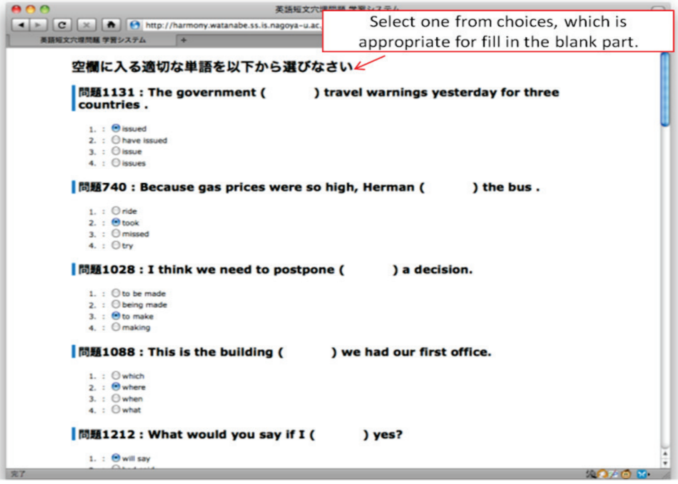


Figure 4: Interface of Prototype System

3. Evaluation

Experiments were conducted using the prototype system. The prototype system is implemented as a web-based system. When the learning starts, the selected questions are shown in the web page as shown in Figure 4. Currently, 10 questions are selected in one learning phase. Learners answer the questions by selecting the radio buttons of the correct choice. After learners select all answers and push the send button, their answers are evaluated, and the result and explanation are displayed.

12 members in our laboratory became examinees of the experiment. First, examinees were asked to solve a pretest which consists of 20 questions and examinees' initial levels were calculated based on the result of the pretest. The questions were carefully prepared by authors to include all levels of difficulty-based features as the equal ratio. In the learning phase, they were asked to answer 10 questions for 10 times. All 10 questions are different.

As the counter methods, we have prepared following two methods:

- **Random link selection method (RLSM)** which selects links randomly in selecting nodes in the question network,
- **Random question posing method (RQPM)** which selects questions randomly from the database.

In RLSM, the movement of the node occurs when the learner can solve 70 percent of the questions in the node. 4 examinees were assigned for each method. Averave understanding levels of examinees who assigned for each method were almost the same.

The correct questions in each learning were evaluated. Table 3 is the average number of correct questions and its variance for each learning. The average numbers are almost the same for all three methods. However, the variance of our method is the smallest of the three. This indicates that the number of correctly answered questions is almost the same for every learning. This result shows that our method could provide questions whose levels are similar to the learners, even if the understanding levels of learners change during the 10 learning.

Table 3: Result of learning phase

	Average # of correct questions	Variance of # of correct questions
Proposed method	5.725	1.585
RLSM	5.850	2.057
RQPM	5.825	2.665

The questionnaire result for acquiring the consciousness of examinees for the proposed questions is shown in Table 4. In each questionnaire item, 5 is the best and 1 is the worst.

Items 1 and 2 got high values. Based on the result of item 1, examinees felt questions become difficult as the learning proceeded. Based on the item 2, they also felt that words were getting more difficult. Table 5 shows the number of links that examinees who use the prototype system with proposed method followed during the learning. All examinees follow links of *difficulty of words* more than 2 times. The worst result of item 4 may be caused by the small number of following links based on *the number of distracter type*. Based on the result, if links are followed, learner can feel the difficulties of questions. Therefore, questions are arranged appropriately by its difficulties in the question network.

Table 4: Questionnaire result

	Contents	Average value
1	Did the questions become difficult?	4.00
2	Did the words in questions become difficult?	4.00
3	Did the question sentences become difficult?	3.50
4	Did the distracters become difficult?	2.75

Table 5: # of links that examinees followed

	Difficulty of words	Length of sentence	The number of distracter type
Examinee 1	3	1	1
Examinee 2	2	2	0
Examinee 3	2	3	1
Examinee 4	3	1	0

4. Conclusion

In this paper, the method for posing questions based on the subjective difficulty-based features was proposed. Based on the experimental result, defined features are intuitive and match to learners' consciousness. In addition, using the question network which arranges questions according to the levels of difficulty-based features, questions that fit for learners' levels were able to be selected in spite of change of learner's situation during the learning. In our future, we need farther experiments with students of various understanding levels for confirming the effect of our method.

Currently, three difficulty-based features have been prepared. However, there are still several other features in questions, such as grammatical structure. Thus, for our future work, to investigate other features of questions is necessary if they become difficulty-based features or not.

References

[1] Goto, T., Kojiri, T., Watanabe, T., Iwata, T., and Yamada, T.: "Automatic Generation System of Multiple-choice Cloze Questions and its Evaluation", An International Journal of Knowledge Management and E-Learning, Vol.2, No.3, pp.210-224 (2010).

[2] Suganuma, A., Mine, T., and Shoudai, T.: "Automatic Generating Appropriate Exercises Based on Dynamic Evaluating both Students' and Questions' Levels", Proc. of ED-MEDIA, CD-ROM, (2002).

[3] Chen, C.M., Lee, H.M. and Chen, Y.H.: "Personalized E-learning System Using Item Response Theory", Computers & Education, Vol. 44, No. 3, pp.237–255 (2005).

[4] Songmuang, P., Ando, M., Nagamori, M., Ueno, M., Okamoto, T.: "E-testing Construction Support System with some Prediction Tools", Proc. of ICALT, pp.437-438 (2007).

[5] Kunichika, H., Urushima, M., Hirashima, T., and Takeuchi, A.: "A Computational Method of Complexity of Questions on Contents of English Sentences and its Evaluation", Proc. of ICCE 2002, pp.97-101 (2002).

[6] Kate, R. J., Luo, X., Patwardhan, S., Franz, M., Florian, R., Mooney, R. J., Roukos, S., and Welty, C.: "Learning to Predict Readability Using Diverse Linguistic Features", Proc. of ICCL 2010, pp.546-554 (2010).

[7] "Standard Vocabulary List 12000: SPACE ALC", http://www.alc.co.jp/eng/vocab/svl/index.html [in Japanese].

Mohd Ayub A. F. et al. (Eds.) (2011). Workshop Proceedings of the 19th International Conference on Computers in Education. ChiangMai, Thailand: Asia-Pacific Society for Computers in Education.

Incorporating Framing into SQL-Tutor

Moffat MATHEWS & Antonija MITROVIC
Intelligent Computer Tutoring Group, University of Canterbury, New Zealand
moffat.mathews@canterbury.ac.nz
tanja.mitrovic@canterbury.ac.nz

Abstract: In this paper, we present our work on framing with the view of implementing it in an Intelligent Tutoring System (ITS). The process of framing a learning activity, in our case problem solving, consists of having the activity in between a pre-action (or priming) phase and a post-action (or reflective) phase. In previous work, we found that simulated framing, in which the priming and reflection phases were led by a human teacher while the learning activity itself was performed in an ITS, significantly reduces learning time and requires less effort for similar gains. This paper presents the next stage of the project, in which the priming phase is implemented in the ITS. We performed a pilot study using the extended system, which resulted in the same trends as simulated framing.

Keywords: Intelligent tutoring systems, framing, teaching strategies

Introduction

In previous work [1] we have presented the initial results on the framing teaching strategy. Framing is a pedagogical strategy that we have distilled from educators' practices in tertiary institutions and high schools. The strategy consists of three sequential phases whereby the learning activity (action phase) is preceded by a pre-action (or priming) phase and followed by a post-action (or reflection) phase. All three phases together form a learning session. In the classroom, students generally participate in the pre- and post-action phases as a group, while the action phase is done either individually or as a group.

The purpose of the pre-action phase is to prepare the student for the learning activity by helping them focus on the concepts that will be used in the learning activity. The aim of this phase is not to teach them the declarative knowledge required but to "set the scene" for the learning activity. Teachers could lead the short, interactive session by (re-) introducing the target concepts, linking them to previously learned concepts, working through examples, discussing common misconceptions, and setting the "boundaries" for the session.

The learning activity phase immediately follows the pre-action phase. Here, the student takes part in some activity that helps them interact with material relating to the target concepts. For example, students might solve problems, engage in discussion, conduct exploratory research, or run experiments. This phase is self-directed, enabling the student to put into practice what they have learned, and the teacher might provide feedback.

Once the learning activity is complete, the teacher leads the students in the reflection phase. The purpose of this phase is to encourage students to reflect on what they have learned in the previous two phases. Students are encouraged to analyze their errors (including the source of these errors) thereby uncovering misconceptions.

There are several theories that make framing a plausible teaching strategy. Cognitive Load Theory [3] suggests that problem solving for novices generates heavy working memory loads, which could be detrimental to learning. To balance these loads, teachers should provide guided instruction, help narrow the problem search space by creating "boundaries" to each session, and alleviate the working memory restriction by making sure that only items relevant to the task are loaded into the working memory. This is exactly what happens during the priming phase.

Meta communication about the presentation of the subject is important for learning [4]. Prior to the learning activity, the student needs to know the boundaries of the lesson segment (exactly what the lesson will contain), which should be well defined by the teacher. The student needs to know the content of the session, differentiating between the old and the new material. They also need to know the links between the new knowledge and previously learned knowledge [4].

Many learning models view learning as a cyclic process, around which knowledge acquisition, knowledge application and reflection occur. Andreasen and Wu [5] discuss a few of the commonly used experiential models. Framing is a simplified (and thus possibly easier-to-implement) version of many of the models.

Reflection promotes deep learning [6, 7, 8]. Critically analysing the learning experience helps challenge the student's underlying perception of the domain, identify and correct misconceptions, and integrate new knowledge with existing knowledge [9]. This also allows the student to transfer the newly acquired knowledge to other types of problems or scenarios. Self-explaining one's actions [10] and monitoring one's progress via open student models [11, 12] have been shown to be useful reflective tools that benefit learning.

Our project consists of three main stages:

1. The learning activity is implemented within SQL-Tutor [2], while the pre- and post-action phases are facilitated by a human tutor. The purpose of this stage was to investigate the potential of framing to improve learning before actually implementing it in the ITS. The results of this stage were presented in [1] are reviewed in Section 1.
2. The pre-action and learning activity phases are implemented in SQL-Tutor. This is the current stage of our project.
3. The reflection phase is also implemented in the ITS.

We chose a suite of metrics early on to validate whether the way in which we implemented Framing helps to achieve the intent. These metrics include Learning Efficiency [13], help-usage metrics e.g. High-Level Help, Requests for Help [14, 15], meta-data about problems solved and problems attempted (including difficulty levels), learning curves, and pre and post-tests. The pre and post-tests were designed by a teacher to measure students' knowledge. The same pre and post-tests are to be used in all stages.

1. Stage 1: Simulating the Framing Strategy

As stated earlier, the purpose of this stage was to simulate the Framing strategy in the manner in which we planned to implement it in the ITS. This helped us gather some information about framing with regards to learning and test out our decisions prior to implementation. We selected a set of target SQL concepts, namely the concepts covered by queries using the *Group By* and *Having* clauses, which students generally find difficult to learn. SQL-Tutor was restricted to only present problems relating to these target concepts. The study was held immediately after the relevant concepts had been covered in lectures. The learning activity was problem solving in SQL-Tutor. The pre- and post-action phase were interactive, whiteboard, group sessions, led by a human teacher. The pre-action and post-action phases were limited to 10 minutes each, while the whole session lasted 100 minutes. In the pre-action phase, the teacher briefly reminded students of the target concepts (taught in lectures previously) and, eliciting student participation, worked through a few examples of varying difficulty. The teacher also discussed typical misconceptions. After interacting with SQL-Tutor, the students were prompted to reflect on their learning experience by commenting on some of their own mistakes. The teacher also showed them the most common mistakes that are usually made during the problem-solving phase. Students were asked to find the errors (in terms of concepts and methods) in those incorrect solutions before collectively working through to reach a correct solution.

Thirty-eight students from a second year database course participated in the evaluation for no monetary reward. We divided participants randomly into two groups: experimental and control. The idea was to perform the evaluation in a setting that was as close to the normal learning environment faced by students. As such, the experimental and control sessions were held during regular course lab sessions. Students in both groups completed a pre-test and a post-test, which were of comparable difficulty and contained three questions relating to the target concepts with the maximum mark of 12. After the pre-test, the experimental group went through priming, followed by problem-solving and reflection phases, which were run as described. In contrast, the control group entered the problem-solving phase immediately after the pre-test. The pre-test, post-test, and problem-solving phases for both groups were identical.

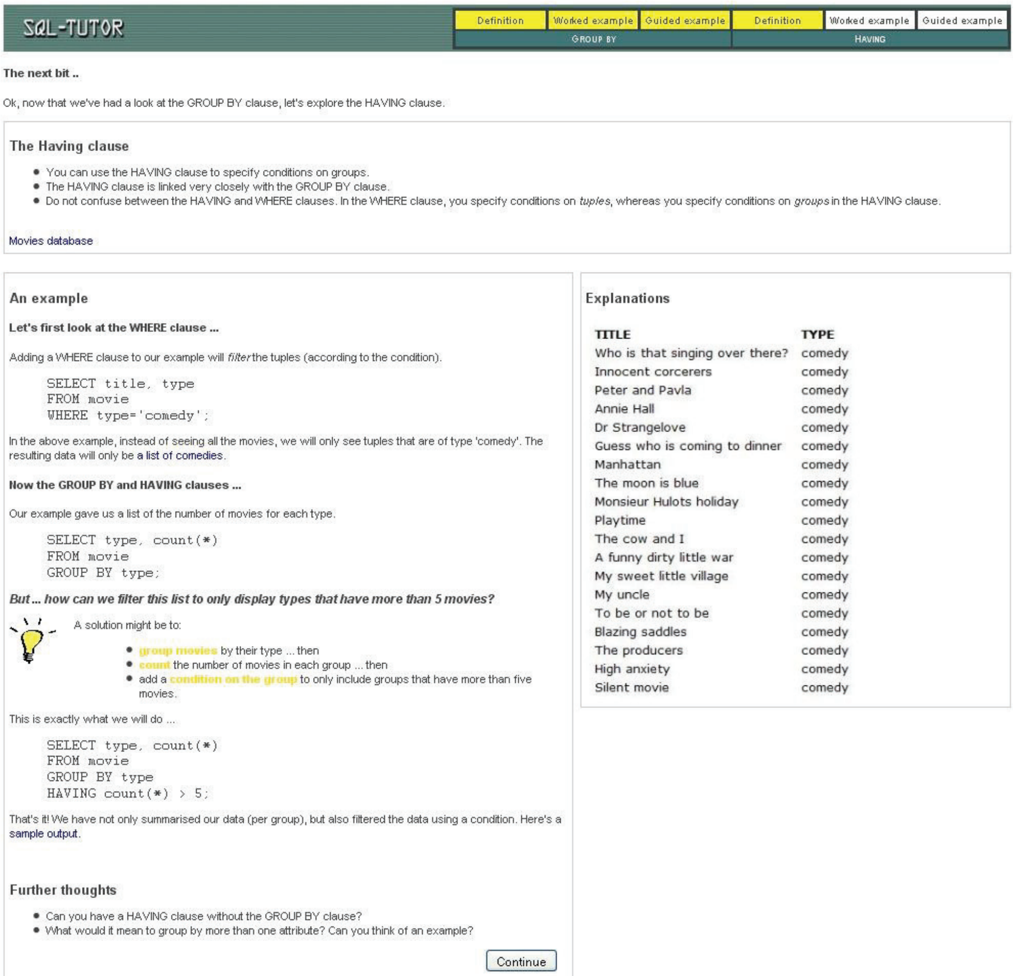


Figure 1: Information about the Having clause

The results of this preliminary study [1] showed that the experimental group had a higher problem-solving speed even though they attempted and solved problems of similar difficulty while using similar levels of help. Furthermore, the experimental group was significantly more efficient in their problem-solving phase than the control group. In other words, while they did not learn more than the control group, they expended significantly less effort and therefore were more efficient.

2. Implementing Priming

We implemented the priming (pre-action) phase within the ITS using the lessons learned from stage 1. The design of this stage was similar, except that we excluded the post-action phase. The pre-action phase contained three steps for each of the target clauses (i.e. three for the Group By clause followed by three for Having clause). Each of the three steps increased from passive to more active in terms of student interaction. The first step contained the declarative knowledge about the clause followed by an example (see Figure 1). The example provided detailed explanation on how to solve the problem, and a possible solution. Students could also click on a link to display the result of the query. A "Further thoughts" section at the end gave more information to extend their knowledge of the clause. Once the student read the information on this page, they proceeded to the next step.

The second (and fifth) step contained a worked example. Students could hover over parts of the solution to get a detailed explanation for that part of the solution. Figure 2 shows the worked example and the explanation for the condition in the Having clause. Hovering over each part of the solution also highlighted the relevant part of the problem statement, allowing students to link the problem to the solution. Students could also click on a link to view the intended output of the query, or view the database schema.

The third (and sixth) step contained a strictly guided example. Similar to step 2, this step contained a problem statement and an empty solution statement (with blanks that the student had to fill in). When the student clicked on one of the blanks, the explanation for solving that part of the problem was displayed. Figure 3 shows the situation when the student asked for the explanation for the blank in the Having clause. The student could click the "Check" button to check their solution. If the student made an error on one of the parts of the solution, the explanation also contained a bottom out hint telling the student what to enter.

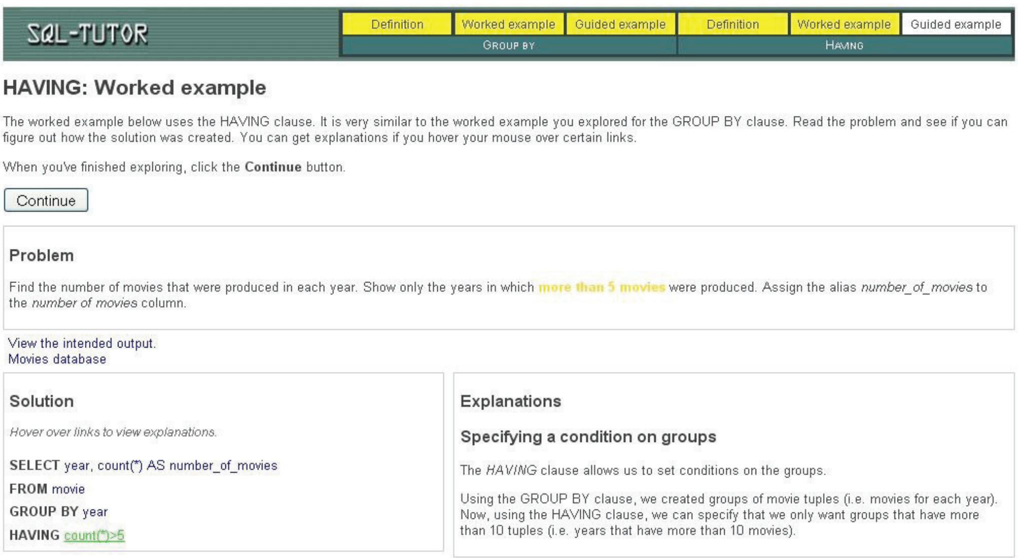


Figure 2: Worked example

The learning activity (problem-solving) immediately followed the six steps of the pre-action phase. This phase was identical to stage 1, where students worked on problems in SQL-Tutor. The problem set was restricted to problems using the target concepts.

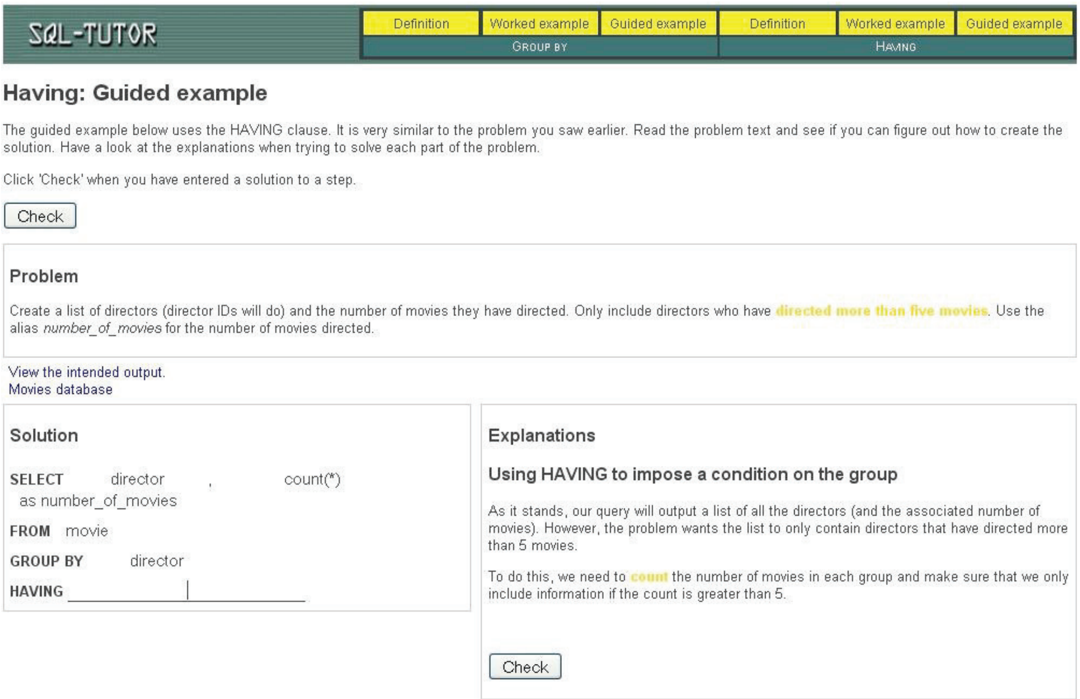


Figure 3: Guided example

3. Results

Thirty students participated in the evaluation for no monetary reward. We divided them randomly into two groups: experimental and control. Two sessions were held during regular course lab sessions (100 minutes long) on 13 and 14 May 2009 respectively. The students participated in the study during the lab session they normally attended throughout the course. Students in both groups completed the pre- and post-test (the same ones used in stage 1 of the project). Following the pre-test, the experimental group went through the pre-action phase in the newly added component, while the control group went directly onto the problem-solving phase.

Table 1: Matched means and standard deviations for test scores (%) and gains

	Pre-test	Post-test	Gain
Experimental group (*n*=5)	57.14% (*s.d* = 39.7)	75% (*s.d* = 16.6)	33.3% (*s.d* = 39.5)
Control group (*n*=12)	52.9% (*s.d* = 26.3)	88.3% (*s.d* = 11.2)	36.6% (*s.d* = 24.7)

The data we collected and analyzed included the pre/post-test results and just over 29 hours (total) of SQL-Tutor student models and logs in which 30 students collectively made 1,769 submissions to the system. There were 17 students in the control group and 13 in the experimental. However, only 17 students sat both tests, and we give the matched results in Table 1. There were no significant differences in the performances of the two groups on the pre-tests, post-tests or between gains (the gain is the difference between post- and pre-test score). There was a significant difference between the pre- and post-test performance of each group, indicating that students improved their domain knowledge during the session.

These results provide us with "trends" even with the low number of students from the experimental group that sat both tests (n=5).

The rest of the analyses were carried out on all the thirty students and the results are reported in Table 2. The trends in this stage were very similar to those found in stage 1. The experimental group spent less time solving problems; this was marginally significant (t(25)=1.3, *p*=.09). Students in both groups solved a similar number of problems. This means that the experimental group solved problems at a slightly faster rate, which was also marginally significant (*t*(19)=0.46, *p*=.09).

We analyzed the problem difficulty levels for both groups. Did students in one group attempt or solve problems that were significantly more difficult than the other that might account for the differing speed of problem solving? Each problem in SQL-Tutor is assigned a difficulty level by the SQL expert who authored the problems. Difficulty levels range from 1 (easy) to 9 (difficult) with non-trivial differences in difficulty between levels. SQL experts have checked problem difficulty levels such that problems with the same difficulty level are of similar difficulty. The problems attempted and solved were also of similar difficulty between groups. This was also confirmed for the highest and lowest difficultly level of problems attempted and solved in both groups i.e. students solved similar types of problems. On average, the experimental group made 49 (26.6) attempts while solving problems, while the control group made 68 (49.3) attempts; the difference was not significant showing that the both groups got similar amounts of feedback from the system. However, to check that students from one group did not receive higher levels of feedback (e.g. they used full solution much more than the other group), we calculated the high-level help used for both groups. *High-level help* (HLH) [14, 15] is defined as the type of help given by a system that provides (part or all of) the correct solution to the student rather than having the student to *solve* the problem; e.g. full solution is a type of HLH. Another important characteristic of HLH in SQL-Tutor is that the HLH levels have to be manually requested by the student whereas the ITS might automatically provide other types of feedback (Low-level help). The *HLH ratio* is the number of HLH attempts divided by the total number of attempts. This shows us the proportion of HLH use, from 0 (no HLH use) to 1 (the student used HLH on every attempt). Students from both groups used similar amounts of high-level help during this phase; 0.46 (0.26) for the experimental group and 0.43 (0.34) for the control group.

Table 2: Results for experimental and control groups

	Experimental	**Control**
Difficulty of problems attempted	5.02 (0.59)	4.95 (0.53)
Difficulty of problems solved	5.01 (0.55)	4.91 (0.56)
Lowest difficulty of problems attempted	3.53 (0.51)	3.64 (0.49)
Highest difficulty of problems attempted	7.0 (1.35)	6.82 (1.7)
Lowest difficulty of problems solved	3.61 (0.50)	3.64 (0.49)
Highest difficulty of problems solved	6.92 (1.32)	6.70 (1.82)
Number of problems solved	10.15 (5.03)	10.5 (6.4)
Time spent on problem solving (min)	52.46 (18.09)	65.17 (33.9)
High-level Help (HLH) ratio	0.46 (0.26)	0.43 (0.34)
Request for Help (RFH) attempts	1.84 (0.68)	1.88 (1.40)
Relative learning efficiency (E)	0.11	-0.12

The relative *learning efficiency* (E) is defined as the performance gained in one condition (the experimental condition) over the effort expended in relation to another condition (the control condition). A condition is more efficient if "1) their performance is higher than expected on the basis of their effort and/or 2) their invested effort is lower than might be expected on the basis of their performance" [13]. To calculate the efficiency of problem-solving for each group, we used "time" as the effort spent and "test gains" as the performance measure. The relative efficiency is found by first converting each of the raw

scores to a *z* score by subtracting the grand mean from the raw score and dividing by the standard deviation. E scores then are found by calculating the perpendicular distance between each *z* score and the E=0 line when plotted on a Cartesian graph. As with stage 1, the efficiency of the experimental group (E = 0.11) was higher than that of the control group (E = -0.12). This was marginally significant ($t(28)=1.11$, $p=0.1$, one-tailed, assuming unequal variances).
We also plotted learning curves for both groups (4). Although the differences were not significant, the trend lines indicate that the experimental group learned at a higher rate than the control group.

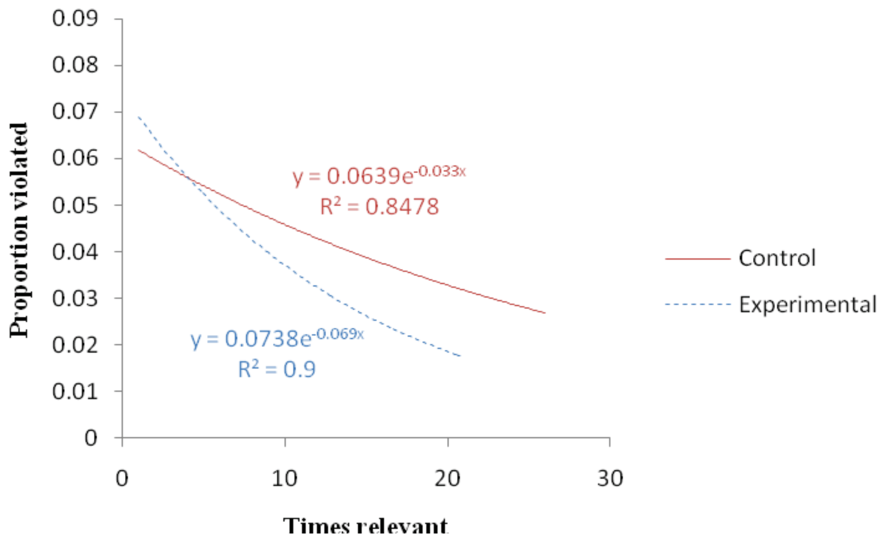


Figure 4: Case study 2: Learning curves for experimental and control groups

4. Discussion and Conclusions

This paper presented the second stage of our project, aiming to implement the framing teaching strategy in an ITS. In previous work, we performed a preliminary study with simulated framing, in which the pre-action and post-action stages were led by a human teachers instead of being implemented in the ITS. The aim of stage 1 was to see whether framing is an effective strategy for an ITS before actually implementing it and therefore committing significant recourses. The results of Stage 1 show that Framing results in significantly faster and more efficient learning.
In this current (second) stage of the project, we implemented the priming phase in SQL-Tutor. This is the first time framing has been implemented in ITSs. The trends gathered from the evaluation of this stage suggest that this implementation worked in a similar manner to that in stage 1. Note that this does not mean that we have achieved an ideal implementation. In fact, although the trends were similar to stage 1, the results gained were not as significant. We have pinpointed at least four possible reasons. First, we had a small number of participants, and therefore cannot make solid conclusions. Secondly, even though the pre-action phase in stage 1 was non-adaptive to the individual, the human teacher adapted to the group as a whole, especially during the worked examples step (when the teacher interacted with the group). This might have increased the effectiveness of the pre-action phase in stage 1. Thirdly, we decided to omit the “common misconceptions” step and only concentrate on correct knowledge. One reason was to keep the pre-action phase reasonably short (to stop it encroaching on the problem-solving). Another reason for the

omission was that presenting correct knowledge followed by incorrect knowledge (common misconceptions) did not seem intuitive using our method of presentation. Finally, the method of presentation differed in both stages. While we had a human teacher (animated, expressive, utilizing the student's visual and auditory senses) presenting in the first stage, we had a series of web pages with limited interaction in the second stage.

The results from this stage, added to that of the previous stage, increase our knowledge and give us a more detailed picture about various decisions we made and aspects of this strategy. Due to the evidence gathered in these stages, it is possible to implement the post-action phase in stage 3 and thus have a system that fully employs the Framing strategy in SQL-Tutor. However, information gathered from this stage suggests that we also could split the development path into a spike that evaluates some of the reasons given in the discussion above and tries to improve on the pre-action phase (say, stage 2B) while continuing development on stage 3. As we have gathered baseline information in stage 2 regarding the pre-action phase, the two stages (stage 2B and stage 3) can be undertaken concurrently. If the spike in stage 2B is successful, the improved pre-action phase can be added with confidence to the system at a later date.

References

[1] Mathews, M., Mitrovic, A. Does framing a problem-solving scenario influence learning? In: Kong, S.C., et al. (Eds.) Proc. 17th Int. Conf. Computers in Education ICCE 2009, Hong Kong: Asia-Pacific Society for Computers in Education, pp. 27-34, 2009.

[2] Mitrovic, A. (2003). An Intelligent SQL Tutor on the Web. *Artificial Intelligence in Education, 13*(2-4), 173-197.

[3] Kirschner, P. A., Clark, R. E., & Sweller, J. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist, 41*(2), 75-86.

[4] Leinhardt, G., & Ohlsson, S. (1990). Tutorials on the structure of tutoring from teachers. *Artificial Intelligence in Education, 2*(1), 21-46.

[5] Andreasen, R. J., & Wu, C.-H. (1999). Study Abroad Program as an Experiential, Capstone Course: a Proposed Model. 15th Annual Meeting of the Association for International Agricultural and Extension Education, Trinidad & Tobago.

[6] Katz, S., Allbritton, D., & Connelly, J. (2003). Going Beyond the Problem Given: How Human Tutors Use Post-Solution Discussions to Support Transfer. *International Journal of Artificial Intelligence in Education, 13*(1), 79-116.

[7] Katz, S., Connelly, J., & Wilson, C. (2007). Out of the Lab and into the Classroom: An Evaluation of Reflective Dialogue in Andes. In R. Luckin, K. R. Koedinger & J. E. Greer (Eds.), *Artificial Intelligence in Education: Frontiers in Artificial Intelligence and Applications* (Vol. 158, pp. 45-432): IOS Press.

[8] Lee, A. Y., & Hutchison, L. (1998). Improving Learning from Examples through Reflection. *Journal of Experimental Psychology: Applied, 4*(3), 187-210.

[9] Atkins, S., & Murphy, K. (1993). Reflection: A Review of the Literature. *Journal of Advanced Nursing, 18*(8), 1188-1192.

[10] Chi, M. T. H. (2000). Self-Explaining Expository Texts: The Dual Processes of Generating Inferences and Repairing Mental Models. In R. Glasser (Ed.), *Advances in Instructional Psychology: Educational Design and Cognitive Science* (Vol. 5, pp. 161 - 238). Mahwah, NJ: Lawrence Erlbaum Associates.

[11] Kay, J. (1997). *Learner Know Thyself: Student Models to give Learner Control and Responsibility.* Paper presented at the International Conference on Computers in Education, Sarawak, Malaysia.

[12] Mitrovic, A., & Martin, B. (2007). Evaluating the Effect of Open Student Models on Self-Assessment. *Artfiicial Intelligence in Education, 17*(2), 121-144.

[13] Paas, F. G. W. C., & Merrinboer, J. J. G. V. (1993). The Efficiency of Instructional Conditions: An Approach to Combine Mental Effort and Performance Measures. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 35*, 737-743.

[14] Mathews, M., & Mitrovic, A. (2008). *How does students' help-seeking behaviour affect learning?* Paper presented at the 9th International Conference on Intelligent Tutoring Systems, Montreal, Canada.

[15] Mathews, M., Mitrovic, A. & Thomson, D. (2008). *How does students' help-seeking behaviour affect learning?* Analyzing high-level help seeking behaviour in ITSs. W. Nejdl et al. (Eds.): AH 2008, LNCS 5149, pp. 312-315, 2008.

Mohd Ayub A. F. et al. (Eds.) (2011). Workshop Proceedings of the 19th International Conference on Computers in Education. ChiangMai, Thailand: Asia-Pacific Society for Computers in Education.

Preliminary Evaluation of an Intelligent Authoring System for 'Graph of Microworlds'

Tomoya Horiguchi[a*] & Tsukasa Hirashima[b]
[a]*Graduate School of Maritime Sciences, Kobe University, Japan*
[b]*Department of Information Engineering, Hiroshima University, Japan*
* horiguti@maritime.kobe-u.ac.jp

Abstract: *Graph of Microworlds* (GMW) is a framework for indexing a set of microworlds for computer-supported adaptive and progressive learning with microworlds. It is difficult to describe a GMW because an author must make a set of microworlds and organize them with model-based indices. Therefore, we proposed a method for semi-automating the GMW-authoring and evaluated it by hand simulation. In elementary mechanics, a GMW of practical size could be described with the method and each microworld was judged to be effective as a learning material. Additionally, by 7 subjects, the explanations generated by the method were judged to be useful in describing a GMW.

Introduction

In physics education, it is important for a student to acquire the ability to make appropriate models of various phenomena in the domain. For this purpose, a set of problems are provided in which he/she must think about some physical systems and their behaviors. In each problem, the range of systems and their behaviors are usually limited from some educational viewpoint in order for him/her to be able to understand the laws/principles behind the phenomena. This is called a *microworld*. For the systematic understanding of the domain theory, therefore, it is necessary to sequence a set of microworlds of various complexity (from relatively simple systems/phenomena to more complicated ones) adaptively to the context of learning.

In designing ITSs (Intelligent Tutoring Systems) with such a function, it is essential to appropriately index a set of microworlds. Especially, it is important to explain why, in the situation given by a microworld, the laws/principles are applicable and why the model is valid. It is also important to explain why/how the model changes if the situation is changed. In order to make such explanations, it is necessary to index a set of microworlds based on their models and the process of modeling.

Therefore, we proposed the *Graph of Microworlds* (GMW), which is a framework for indexing the microworlds and the relations between them based on their models and the process of modeling [4]. We also showed, by using GMW, it becomes possible to design a function for adaptively selecting the microworld which a student should learn next, and a function for assisting a student in transferring between microworlds.

However, it isn't easy to describe a GMW because an author must make a lot of indices in a model-based way. He/she must have the expertise in the process of modeling. Therefore, we also proposed a method for semi-automating the description of GMW by introducing an automatic modeling mechanism [5] (i.e., compositional modeling [3, 6]). Though the authoring system which implements this method is currently under construction, we described the domain knowledge for it which covers elementary mechanics and successfully simulated its behavior by hand. In this paper, we report the result of a preliminary experiment which was conducted by hand-simulation and validated the usefulness of our method.

1. Graph of Microworlds and its Authoring

An example of GMW for elementary mechanics is shown in Fig. 1. Each microworld is indexed with the situation it deals with, the model of the situation and the process of modeling. A student can learn the physical law(s)/principle(s) necessary for the modeling and the skill(s) for the model-based problem solving in each microworld (they are called a learning item). Two microworlds which deal with similar situations but different models (i.e., different law(s)/principle(s) is(are) necessary) are linked to each other with an edge. Parameter-change rules [1] are attached to such an edge which relate the difference between the situations of two microworlds to the difference between the behaviors of their models. This means one model is the necessary evolution of the other (with the perturbation of situation). Such a relation between two microworlds is called an educationally meaningful relation. In order to make a student learn the domain theory progressively [2], a GMW should include as many such relations as possible.

Fig. 2 shows the framework for authoring GMW. An author describes a GMW as follows: Suppose a learning item network is given which consists of a set of concepts and their partial ordering to be learned. First, he/she finds a situation for learning an item. Then, he/she perturbs the situation to make a new situation for learning another item adjacent to the former. For each situation, the system generates its model and indexes it with its modeling assumptions automatically by *compositional modeling* [3, 6]. By repeating such perturbation, he/she finally gets a GMW which covers the learning item network. In this process, the system generates explanations about the differences between situations to help an author judge whether they have 'educationally meaningful' relation or not (the detail for generating explanations is described in [5]).

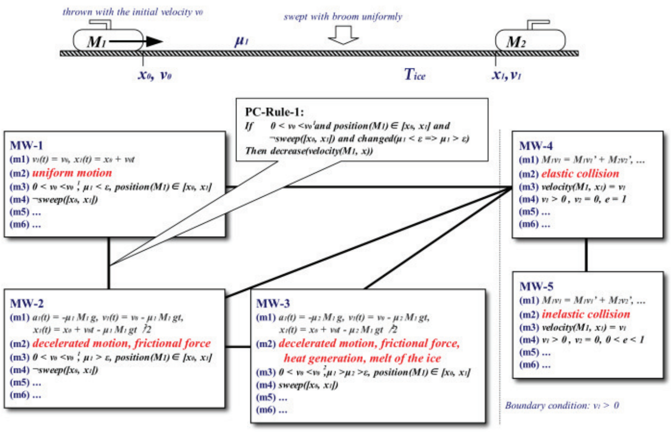


Fig. 1. An example of Graph of Microworlds.

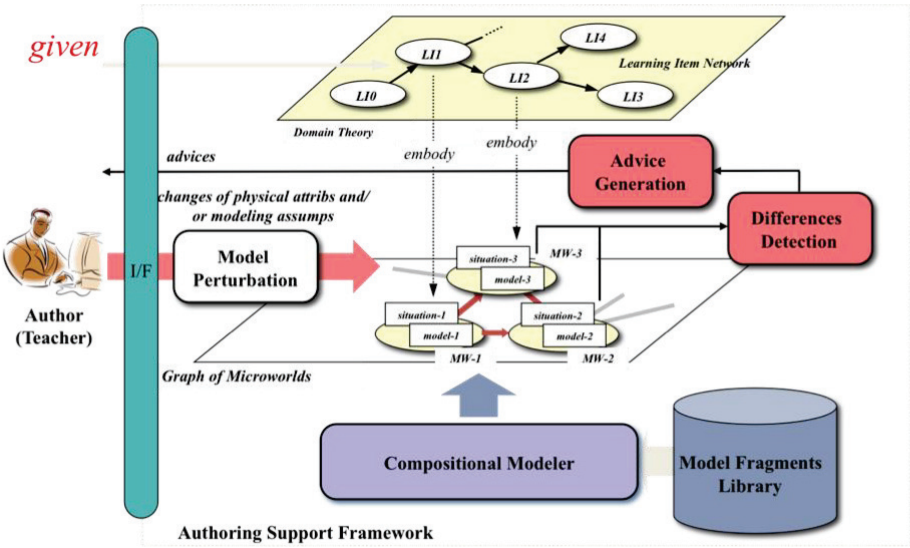


Fig. 2. Framework for authoring GMW

2. Preliminary Experiment

2.1 Design of the Experiment

First, we examined whether one can describe a GMW covering a given learning item network of practical size with our method. We made a learning item network of elementary mechanics (which includes 38 learning items and is partially shown in Fig. 3) by analyzing 2 textbooks in physics for senior high schools. We then tried to describe a GMW which covers it with the method.

Second, we examined whether each microworld made in the above process can be effective in learning the corresponding learning item. A microworld was judged to be effective if a problem which deals with the same item and the same situation as it was found in text/exercise books[1].

Third, we evaluated the ability of the method to generate explanations about the differences between microworlds. In text/exercise books, there aren't always 'educationally meaningful' relations between two situations of the problems which deal with adjacent learning items. Therefore, after selecting 6 such pairs of problems from text/exercise books, we made the 'bridging' microworlds with our method (a 'bridging' microworld has 'educationally meaningful' relations with both problems in a pair). We then asked 7 subjects (who were under/graduates majoring in engineering) to judge whether these microworlds were effective in facilitating progressive learning and whether the explanations generated with the method about their differences were useful.

2.2 Results

First, a GMW covering the learning item network shown in Fig. 3 could be described with the method. It, besides the microworlds corresponding to the given items, has 4 extra microworlds each of which was inserted to bridge the gap between two microworlds (where one couldn't make the adjacent microworlds corresponding to adjacent items by perturbing the situation). Though the GMW was described by the authors, we think it reasonable

[1] We assumed the situations of the problems in text/exercise books were guaranteed to be effective in dealing with the corresponding learning items, and tried to reproduce as many such situations as possible when describing the GMW.

because the purpose of this experiment is to evaluate the ability of the method to make models by perturbing situations (not the usability for end users).

Second, every microworld except one[2] in the GMW (partially shown in Fig. 3) could be made the same as the situation of the problem in text/exercise books (we referred 5 ones). Therefore, the microworlds made with the method can be effective in learning the corresponding learning items.

Third, the evaluation result by the subjects about the effectiveness of the 'bridging' microworlds is shown in table 1. It reveals that the inserted microworlds between those of too different situations were effective in complementing the gaps with 'educationally meaningful' relations, and that the explanations generated with the method were useful in understanding the differences between microworlds (in case-1 and 4). Even though the effectiveness of the microworlds was negatively evaluated since they were inserted between those which weren't judged to be too different, the usefulness of the explanations by the method were positively evaluated (in case-2, 5 and 6). That is, in all the cases, the method could generate useful explanations for understanding the differences between microworlds.

In this experiment, because we tried to reproduce as many situations of the problems in text/exercise books as possible when describing the GMW, some edges became ineffective in facilitating progressive learning (i.e., in case-2, 5 and 6). This matter, however, could be improved when such constraint is removed.

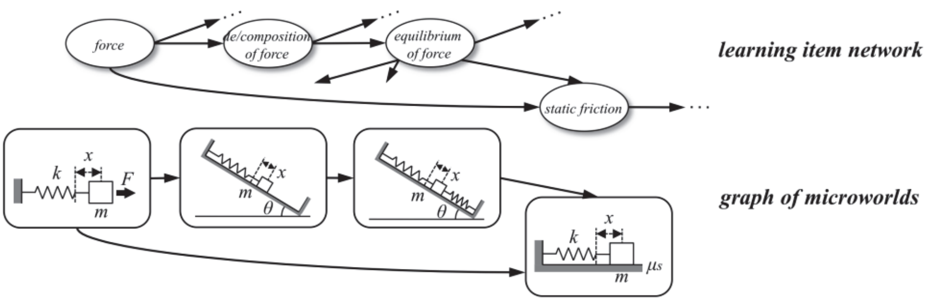


Fig. 3. Learning item network and a GMW covering it

Table 1. Evaluation result about bridging microworlds

	Case-1	Case-2	Case-3	Case-4	Case-5	Case-6
Q1	0.86	2.43	2.00	1.43	3.29	3.57
Q2	3.57	1.43	2.14	3.29	2.43	1.29
Q3	3.43	2.86	3.00	3.71	3.00	2.57

Q1: Are the situations of two microworlds close enough to facilitate progressive learning?
Q2: Is the microworld inserted between two microworlds effective in facilitating progressive learning?
Q3: Is the explanation by the system useful to understand the difference between situations?
*Each score is the average of five degree ratings by 7 subjects (0: the most negative - 4: the most positive).

3. Conclusion and Future Work

These results suggest our method is useful in describing a GMW covering a learning item network of practical size which effectively facilitates progressive learning. One of our important future work is to complete the prototype by adding a GUI-based interface and examine what GMWs could be described by end-user authors.

[2] Thirty-one microworlds (out of thirty-eight learning items) were examined because seven microworlds dealing with 'work' and/or 'energy' became the same situation as the others. Though the only exception was a microworld dealing with 'de/composition of forces,' its situation wasn't ineffective because it was a part of the situation dealing with the following learning item 'angled projection.'

References

[1] Addanki, S., Cremonini, R. and Penberthy, J.S.: Graphs of models. *Artifi. Intelli.*, 51: 145-177 (1991)

[2] Burton, R.R., Brown, J.S. & Fischer, G.: Skiing as a model of instruction, In Rogoff, B. & Lave, J. (Eds.), *Everyday Cognition: its development in social context*, Harvard Univ. Press (1984)

[3] Falkenhainer, B. and Forbus, K.D.: Compositional Modeling: Finding the Right Model for the Job, *Artifi. Intelli.*, 51, pp.95-143 (1991)

[4] Horiguchi, T. & Hirashima, T.: Graph of Microworlds: A Framework for Assisting Progressive Knowledge Acquisition in Simulation-based Learning Environments, *Proc. of AIED2005*, pp.670-677 (2005)

[5] Horiguchi, T. & Hirashima, T.: Intelligent Authoring of 'Graph of Microworlds' for Adaptive Learning with Microworlds based on Compositional Modeling, *Proc. of AIED2009*, pp.207-214 (2009)

[6] Levy, A.Y., Iwasaki, Y. & Fikes, R.: Automated model selection for simulation based on relevance reasoning, *Artifi. Intelli.*, 96, pp.351-394 (1997)

[7] White, B. & Frederiksen, J.: ThinkerTools: Causal models, conceptual change, and science education, *Cognition and Instruction*, 10, pp.1-100 (1993)

[8] White, B. & Frederiksen, J.: Causal model progressions as a foundation for intelligent learning environments, *Artifi. Intelli.*, 42, pp.99-157 (1990)

ICCE 2011

The 19th International Conference on Computers in Education

28 November - 2 December • Chiang Mai, Thailand

Proceedings of the 19th International Conference on Computers in Education ICCE 2011

ISBN 978-616-12-0186-9

THAILAND convention & exhibition bureau

Bangkok Bank ธนาคารกรุงเทพ

สวทช NSTDA

สสวท