Analysis of Multiple-Choice Tests Through Erroneous Choices Using a Technique of Automatic Problem Generation

Noriyuki Matsuda[1](✉), Hisashi Ogawa[2], Tsukasa Hirashima[3], and Hirokazu Taki[1]

[1] Faculty of Systems Engineering, Wakayama University, Wakayama, Japan
{matsuda, taki}@sys.wakayama-u.ac.jp
[2] The Joint Graduate School in Science of School Education, Hyogo University of Teacher Education, Hyogo, Japan
ogawa@hyogo-u.ac.jp
[3] Department of Information Engineering, Hiroshima University, Hiroshima, Japan
tsukasa@lel.hiroshima-u.ac.jp

Abstract. In multiple-choice problems, incorrect choices that hide correct answer do not play the only significant function. A more significant function is to capture learner errors and provide learners with the opportunity to amend those errors. In this sense, explanation texts for incorrect choices might be important in allowing learners to amend their own errors correctly. In this paper, we term this "meaningful erroneous choice." We confirm the existence of meaningful erroneous choice through actual problems used in an examination. We then compare choices using a technique of automatic problem generation. The result indicates that the difference between correctness knowledge and incorrectness knowledge plays an important role in problem posing.

Keywords: Multiple-choice problems · Automatic problem generation · E-learning · Prolog

1 Introduction

Erroneous answers in multiple-choice problems not only make the correct answer more difficult to determine, but also indicate why the correct choice is suitable and the erroneous one unsuitable when compared to the correct answer. Munby [10] asserts that erroneous answers should prompt learners to make careful considerations before making their selection; in this scenario, learner instruction is affected by the selected choice, and texts explaining the erroneous answers. We distinguish between "meaningless erroneous distracters" (MDs) and "meaningful erroneous choices" (MCs). The latter (MCs) stimulate learning through a ripple effect. For instance, MCs can explain why an answer is incorrect, but MDs do not provide such an explanation, stating only "because it differs in appearance from the correct answer."

This paper attempts to explain the role of MCs as compared to MDs. Our approach followed two steps. Firstly, we have surveyed and analyzed multiple-choice tests from

S. Yamamoto (Ed.): HIMI 2015, Part II, LNCS 9173, pp. 362–369, 2015.
DOI: 10.1007/978-3-319-20618-9_36

actual learning material and examinations. Secondly, we have attempted to generate the erroneous choices surveyed using a technique to generate multiple-choice tests. This technique is one of Simulation of Erroneous Solutions (SES), which generates erroneous choices with explanations of the error through problem solving knowledge written in Prolog. We have tried to confirm the differences in generating techniques between meaningful and meaningless choices.

Through surveying erroneous choices in 80 multiple-choice tests, a university teacher and a high school teacher have estimated the design intention of the tests and analyzed the process of problem construction.

Our SES is a method of generating a solution from incorrect knowledge to incorrect answer, through a perturbation from problem-solving knowledge to knowledge of incorrect answer derivation. The perturbation is projected onto the design intention of the teacher who creates the problem, and expresses the difference between the erroneous solution and the correct solution. For instance, a 48-question and 128-explanation text was automatically generated from 17 Prolog clauses regarding knowledge about SES. The SES has the potential to generate relevant choices and erroneous solutions. In this paper, we are concerned with the difference in perturbation that generates meaningful and meaningless choices. Meaningful perturbation changes slightly correct solutions written in Prolog and the latter perturbation randomly changes solutions. We also discuss the differences between the explanation texts that were generated automatically by SES.

2 Analysis of Erroneous Choices

We analyzed actual problems randomly collected from an information technology engineers' examination in Japan. Two teachers, a university teacher and high school teacher who create problems professionally, read all problems and discussed whether incorrect choices should be categorized as MDs or MCs. They then selected MD or MC for each choice. Incorrect choices were classified into the following five categories (A to E).

- A. Problems that require selection of concepts to satisfy any given conditions (16 problems and 48 choices)
  All erroneous choices satisfy some part of the given conditions in a problem. For example, Problem 1 below consists of three erroneous choices that satisfy the given condition "protocol used on TCP/IP network."

(Problem 1) Which of the following is a TCP/IP protocol that provides a bidirectional interactive text-oriented communication facility using a virtual terminal connection?

(a) FTP
(b) HTTP
(c) SMTP
(d) TELNET

Therefore, all incorrect choices are recognized as MCs that can be generated by perturbation; such as changing some part of the knowledge. For example, incorrect choice (a) FTP can be generated by changing "text-oriented communication facility" into "computer file transfer facility."

- B. Problems that can be solved using procedural knowledge (32 problems and 96 choices)
  An answer is obtained using a procedural solution, such as a calculated problem.

(Problem 2) When the manipulations shown below are executed for the empty stack, which of the following are the appropriate remaining data in the stack?

```
push 1 -> push 2 -> pop -> push 3 -> push 4 -> pop ->
push 5 -> pop
```

(a) 1 and 3
(b) 2 and 5
(c) 1 and 5
(d) 4 and 5

Twenty-four incorrect choices are recognized as MDs because it is impossible to find an adequate reason for error. Seventy-two incorrect choices are recognized as MCs. For example, meaningful erroneous choice (d) in Problem 2 is obtained by replacing "stack" with "queue."

- C. Problems that require selection of an explanation
  (42 problems and 126 choices)
  Problems provide explanatory choices for a word given.

(Problem 3) Which of the following is the most appropriate explanation of DHCP?

(e) A protocol that automatically provides an IP address and networking parameters
(f) A protocol used in order to access a directory service
(g) A protocol that sends emails
(h) A protocol that converts a private address to a global address

All incorrect choices were recognized as MCs. Seventy-two incorrect choices explained another concept that could be distinguished with a correct answer. Twenty-four incorrect choices partially switched in explanation of another concept. Twenty-four incorrect choices partially switched in the opposite meaning word. Six incorrect choices partially switched in an explanation of similar concepts.

- D. Problems that require selection of a correct combination
  (Eight problems and 24 solutions)
  Problems provide a combination of relationships, such as causal relationship and correlations.

(Problem 4) Which of the following is the appropriate combination of a threat and countermeasure of information security? (The upper row shows a threat and the lower row shows a countermeasure.)

(a) Logical data destruction by misoperation
    Disk array
(b) Earthquake and fire
    Data duplication in a machine using a virtual machine

(c) Unauthorized access to data during transmission
CRC by HDLC
(d) Alteration of message
Digital signature by public key encryption

All incorrect choices are recognized as MCs. They are generated by switching in another related concept or simply other concept which has no relation.

- E. Problem that requires selection of a correct order
(Two problems and six choices)
Problems provide the relationship of the order or position as choices.

(Problem 5) Which of the following is an appropriate order to establish ISMS of JIS Q 27001:2006?

1. Create an application declaration
2. Select management objectives for risk operation and control measures
3. Analysis and assessment of risk

(a) 1 -> 2 -> 3
(b) 1 -> 3 -> 2
(c) 2 -> 3 -> 1
(d) 3 -> 2 -> 1

All incorrect choices are recognized as MDs. The correct answer for Problem 5 is choice (d). However, a reasonable cause of error cannot be found.

We have confirmed the existence of MCs in actual problems. Next, we consider the difference between MDs and MCs in problem posing.

3 Simulation of Erroneous Solution

To consider the differences between MDs and MCs, we use a technique of automatic problem generation. We have already proposed a technique to generate problems and explanation of incorrect choices, called Simulation of Erroneous Solution (SES) [7, 11]. Currently, much research exists, for example, in an e-Learning context. Automatic generation techniques that use text processing and statistical analysis [3, 6, 9], a corpus thesaurus of general knowledge within a particular field [2, 5, 13], and domain ontology [1, 4, 8, 12] have been suggested. While these techniques can automatically generate large numbers of problems at a single time, SES can generate multiple-choice problems and explanation texts for incorrect choices (see Fig. 1). Before this, SES used correctness knowledge to solve the problem. This knowledge is written in Prolog, so we can obtain the correct answer by running the Prolog code. To express learner errors, we defined perturbation, which slightly changes the correctness knowledge to simulate learners' incorrect solutions. This perturbation is also written in Prolog clauses. In this way, we can obtain the incorrect answer and its explanation text, which is constructed of clauses in the resolution process in Prolog with a template of explanation.

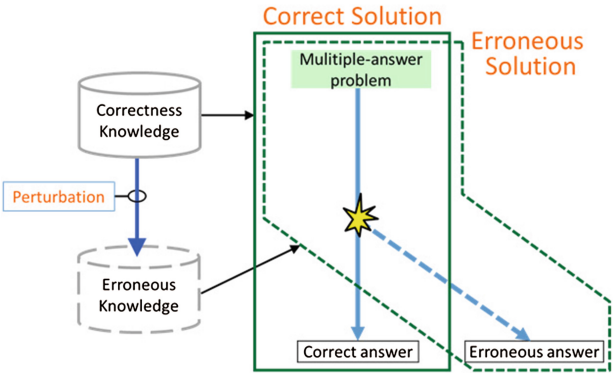

Fig. 1. Outline of SES provides erroneous choices and explanations for the errors with perturbation of correctness knowledge. Perturbation that slightly changes correctness knowledge expresses learners' erroneous solutions.

Here, we start to prepare SES to generate Problem 2. Correctness knowledge is shown in Fig. 2. It includes rules for stack and queue. The "Push" operation is the same and the "Pop" operation is different.

```
% Knowledge
%   For Stack
push(A, stack)  :- retract(stack(B)),      asserta(stack([A|B])).
pop(stack)      :- retract(stack([_|B])), assert(stack(B)).
%   For Queue
push(A, queue)  :- retract(queue(B)), asserta(queue([A|B])).
pop(queue)    :- retract(queue(A)), shift(A, B), asserta(queue(B)).

% Private Procedure
shift([_], []).
shift([A|B], [A|C]) :- shift(B, C).
```

Fig. 2. Correctness knowledge for Problem 3

The given condition in the problem is shown in Fig. 3. There, stack and the procedure of push and pop are empty.

```
stack([]).
procedure-stack :- push(1, stack), push(2, stack), pop(stack), push(3, stack),
                   push(4, stack), pop(stack), push(5, stack), pop(stack).
```

Fig. 3. Given conditions in Problem 3. When we execute this Prolog code with correctness knowledge in Fig. 2, the correct answer and explanation text for the correct solution are generated.

Next, we write down the perturbation for the learner's error, such as misunderstanding of stack and queue (see Fig. 4). When perturbation is added on top of correctness knowledge, the MC (d) and explanation text are generated.

```
pop(stack) :- retract(stack(A)), shift(A, B), asserta(stack(B)).
```

Fig. 4. Perturbation of Problem 3. It shows a misunderstanding of stack and queue. When this perturbation is added on top of correctness knowledge in Fig. 2, meaningful erroneous choice (d) is obtained.

On the other hand, MD is randomly generated from a combination of list [1–3] except for the correct answer [1, 3].

Thus, perturbation of correctness knowledge has a significant function for MCs. Perturbation might be considered to capture learner errors in order to give the learner important learning opportunities. In addition, perturbation might be considered to express a teacher's educational intention. Therefore, the teacher who created the problem naturally writes down an explanation of the incorrect choice to lead a learner to the correct solution.

Another example is shown below. A problem asks an adequate bio-indicator in river (see Fig. 5). A perturbation shows learner's typical misunderstanding about the relation between electrical conductivity (EC) and dissolved matter in the river. Correct knowledge is that EC is proportional to the amount of dissolved matter. Typical incorrect

Problem: When the river's EC is low, flow volume is also low, and upstream land is not used as forest. What is the bio-indicator?

Correct answer	Plecoptera
Erroneous answer	Corixidae
Erroneous answer	Chironomidae
Erroneous answer	Luciola

<u>Text explaining the erroneous answer</u>

Corixidae: Incorrect. If EC is low and flow volume is low, you should assume that a small amount of dissolved matter is present. Did you mistakenly think there would be a large amount of dissolved matter?

The correct understanding is that because EC and flow volume are low, a small amount of dissolved matter will be present. Since a small amount of dissolved matter is present and the upstream land is not used for forest, the water quality will be clean. As the water quality will be clean, the bio-indicator will be Plecoptera. Therefore, the correct answer is Plecoptera.

Fig. 5. An example problem which asks bio-indicator. A learner needs to know the relations that the amount of dissolved matter is determined by upstream land use, flow volume, and electrical conductivity (EC), and also that the bio-indicators are determined by water quality.

knowledge is that EC is inversely proportional to the amount of dissolved matter. Therefore perturbation is to change from direct proportion to inverse proportion. Such perturbation can generate MCs which can have some explanation to fix typical misunderstanding about the relationship between EC and dissolved matter. On the other hand, random selection for bio-indicator or living things in river can generate MDs. MDs cannot have no explanation about the error.

4 Conclusions

We have confirmed the existence of MCs through actual problems in an examination. In addition, we have compared MCs and MDs with regard to perturbation of SES. We believe that incorrect choices are not only distracters but also capture learners' errors. In future work, through automatic problem generation, perturbations for MCs should be analyzed into categories and formulated.

References

1. Alsubait, T., Parsia, B., Sattler, U.: Mining ontologies for analogy questions: a similarity-based approach. In: Klinov, P., Horridge, M. (eds.) OWLED, CEUR Workshop Proceedings, vol. 849 (2012)
2. Brown, J.C., Frishkoff, G.A., Eskenazi, M.: Automatic question generation for vocabulary assessment. In: Proceedings of HLT/EMNLP 2005, pp. 819–826 (2005)
3. Correia, R., Baptista, J., Mamede, N., Trancoso, I., Eskenazi, M.: Automatic generation of cloze question distractors. In: Proceedings of the Interspeech 2010 Satellite Workshop on Second Language Studies: Acquisition, Learning, Education and Technology (2010)
4. Holohan, E., Melia, M., McMullen, D., Pahl, C.: Adaptive e-learning content generation based on semantic web technology. In: International Workshop on Applications of Semantic Web Technologies for E-Learning (SW-EL 2005) – at the 12th International Conference on Artificial Intelligence in Education AIED 2005, pp. 29–36 (2005)
5. Sumita, E., Sugaya, F., Yamamoto, S.: Measuring non-native speakers' proficiency of English by using a test with automatically-generated fill-in-the-blank questions. In: Proceedings of ACL 2005 Workshop: Building Educational Applications Using Natural Language Processing, pp. 61–68 (2005)
6. Gotoh, T., Kojiri, T., Watanabe, T., Iwata, T., Yamada, T.: Automatic generation system of multiple-choice cloze questions and its evaluation. Knowl. Manag. E-Learn. Int. J. **2**(3), 210–224 (2010)
7. Matsuda, N., Ogawa, H., Hirashima, T., Taki, H.: A Generating Technique and Knowledge Representation of Multiple-Answer Problems for Learning with Solving Knowledge, Research and Practice in Technology Enhanced Learning (2015)
8. Mitkov, R., Ha, L., Karamanis, N.: A computer-aided environment for generating multiple-choice test items. Nat. Lang. Eng. **12**(2), 177–194 (2006)
9. Moser, J.R., Gütl, C., Liu, W.: Refined distractor generation with LSA and stylometry for automated multiple choice question generation. In: Thielscher, M., Zhang, D. (eds.) AI 2012. LNCS, vol. 7691, pp. 95–106. Springer, Heidelberg (2012)
10. Munby, J.: Read and Think. Longman, Harlow (1968)

11. Ogawa, H., Kobayashi, H., Matsuda, N., Hirashima, T., Taki, H.: Knowledge externalization based on differences of solutions for automatic generation of multiple-choice question. In: Proceedings of the 19th International Conference on Computers in Education, pp. 271–278 (2011)
12. Papasalourosa, A., Kotisb, K., Kanarisb, K.: Automatic generation of tests from domain and multimedia ontologies. Interact. Learn. Environ. **19**(1), 5–23 (2011)
13. Lin, Y.C., Sung, L.C., Chen, M.C.: An automatic multiple-choice question generation scheme for English adjective understanding. In: Workshop on Modeling, Management and Generation of Problems/Questions in eLearning, the 15th International Conference on Computers in Education, Hiroshima, Japan, pp. 137–142, November 2007