Development of On-Line Test Materials with a Checklist for Information Literacy Education

Ryuichi Matsuba[1,4)], Hiroshi Nakano[2,4)], Sachi Takahashi[3,4)], Toshihiro Kita[2,4)],
Shirou Kitamura[4)], Masahiro Migita[2)], Kenichi Sugitani[2)], Norio Iriguchi[2,4)], Yasuo Musashi[2)],
Kazutaka, Tsuji[2)], Takeshi Kida[2)], Masaru Shimamoto[2)], Katsuaki Suzuki[4)], Tsuyoshi Usagawa[2,4)]
[1)] Institute for e-Learning Development
[2)] Center for Multimedia and Information Technologies
[3)] Research Center for Higher Education
[4)] Instructional Systems Program, Graduate School of Social and Cultural Sciences
Kumamoto University
Japan
matsuba@kumamoto-u.ac.jp

Abstract: We report on the development of on-line test (exam) materials for information literacy education with the use of a checklist. Some points of view to improve materials have been known by prior studies and we could collect and list up checking items to assess and improve on-line test materials. We show the results of our development of materials by comparing the score distributions before and after the improvement.

Introduction

It is often said that repeated on-line tests (or exams) with the use of a learning management system (LMS) are very effective to confirm and understand materials for learning in classes (e.g., Williams & Bialac 2006; Ozkan & Morris 2006). We have confirmed that weekly on-line tests are quite effective in courses of basic information technologies offered for more than 1,800 learners. The courses consist of “Basic Course of Information Technologies A/B” in the first (A) and second (B) semesters for freshmen of Kumamoto University. We are conducting the courses in a blended style by using identical on-line contents. This is in order to guarantee that “all graduates from whichever department of Kumamoto University acquire information skills of a certain level.’’ The target of the course “A” is that learners can use office applications and the Internet safely with good manners, and the course “B” aims that learners can obtain and send information via the Internet and can understand basic protocols and structures of the Internet. The averaged number of students per class is 40-100 supported by an instructor and a few teaching assistants. Consequently, the LMS is very effective and has been used for the courses in the blended style of both on- and off-campus learning. In the courses we offer on-line learning materials, attendance registration tools and uploading tools of weekly assignments and implement the on-line tests with the LMS (Usagawa & Nakano 1995).

The on-line exams, a means to measure academic achievement, play an important role in the basic courses of information technologies. For measurement of the achievement and offering repeated chances of studying contents particularly little understood, we offer about twenty on-line exams a year. Students can take an exam any times within a given weekly term, and only the best score is recorded in the LMS. Students can view his/her own scores and read comments just after finishing an exam with the automatic scoring system. An exam offers randomly ordered questions, and a question offers randomly ordered items. This is to shuffle the order of the questions. Though the best score is recorded and shown to individual students, all attempts, processes and scores are recorded in the LMS. Such recorded logs are very useful not only for leading learners but also for instructors who want to review the test materials and find some difficulties and evaluate the quality of the test. We can obtain all results of exams and generate score distribution data of all trials from the test logs.[i)] The score distribution patterns of the first attempts vary with tests, however, the pattern become similar after several trials. Some students earn high scores by taking the test a few times while some other students earn high scores by taking the test 70 times. More than 90% students can earn scores of a high level to be qualified as an adequate academic achievement (Nakano et al. 2005).

It appears to be appreciative if we could use a checklist of on-line exam contents to be assessed. There are many reports on practice/case examples, development of contents, and on-line assessments (e.g., Hricko & Howell 2006). Thus we can readily introduce various insights and crate a checklist to evaluate on-line contents of exams.

In the present paper, we illustrate our development of on-line test materials for information literacy education. We show the checklist to evaluate our questions. Then we show the results of some statistical data of the on-line tests before and after the improvement of materials using the checklist.

A checklist for development of on-line tests

In order to develop on-line test materials, we have made a checklist to evaluate the quality of questions. Many studies have shown points of view to make good assessments/tests (e.g., Hricko & Howell 2006, Kahn 2005; Kishi). We have summarized them in order to make a checklist for our purpose. The checklist has forty-five checking items divided into nine categories; contents, difficulty levels, questioning techniques, selection choices, strictness, expressions, specificity of e-learning, feedback statements for learners, and across-the-board. All of the items are answered by selecting "yes (reasonable/valid)" or "No (unreasonable/invalid)". We show the checklist in Table 1. The checklist is an English translated version.

Table 1: A checklist for the on-line questions/tests

Categories	Checking Items
Contents	■ The question is directly related to the subject. ■ The question asks essential and important points to confirm contents of the subject. ■ The question evaluates the depth of understanding and thinking rather than the memory. ■ There is no dependence of questions on each others. ■ The question does not require time wasting. ■ The level of the question is upper or lower than that of the main contents.
Difficulty levels	■ The difficulty of the question is within the expected level of the test. ■ There is a room for improvement by adjusting the difficulty level. ■ The problem statement is somewhat decipherable.
Questioning techniques	■ Learners can easily understand the meaning of the questionnaire. ■ It employs an appropriate question form for the asking. ■ The question has pairing answers. ■ The statements of the problem and answer are written in positive sentences. ■ There is a scheme to avoid a fluke.
Selection choices	■ There is no choice immediately recognized as an incorrect answer. ■ The question may intentionally mislead learners. ■ Learners may elicit the right answer even if he/she has a fallacy or mistaken notation. ■ All of the choices stand on an equal footing. ■ Learners can answer the question with common sense. ■ The choices can be grouped into categories. ■ The statement of the question can naturally lead to the right answer. ■ Both the statements of the problem and answer are too long. ■ In a logical reasoning question, there is no reasonable sequence of the choices. ■ The question adopts an appropriate and effective form.
Strictness,	■ The question includes a tacit condition. ■ There is an ambiguity in correct (or incorrect) answers. ■ Multiple-choices are distinct among themselves. ■ There is a misquotation of a figure, table, and reference.
Expressions	■ The expression is appropriate in the statements of both question and answer. ■ The word modification is incorrect. ■ There is an ambiguous, strictly incorrect, expression. ■ Common statements are included in choices.
Specificity of e-learning	■ The question and choices can be appropriately displayed. ■ The question or choices has a trouble of a line break. ■ In the character input question, the length of entry field is appropriate. ■ In the character input question, there is a care of character encoding.
Feedback statements	■ The feedback is appropriate in eliciting the right answer. ■ The feedback explains appropriately why the error answers are wrong. ■ The needless feedback wastes the chance of re-taking the test.
Across-the-board	■ The purpose of the test is readily comprehensible for learners. ■ The number of the questions of an exam is appropriate. ■ The acceptance criterion is well-defined and valid. ■ The test has well-chosen questions to attain the target. ■ There is such care for learners as a scheme to avoid superficial error. ■ There is uniformity in wording (which is inherent in Japanese).

The score distribution of the on-line tests

We see some heterogeneous distributions by checking the distributions of scores. The typical case is shown in Fig. 1a. It has a symmetric shape of the score distribution of the first attempt, here and hereafter, we call it the initial score distribution. The peak is shifting from the center (40-60 points) to the higher range (80-100 points) with the attempt number of times increasing. Eighty six (86) percents of learners earn more than 80 marks by 10 attempts or less which percentage is almost the same as the final result (87 percents). It should be noted that if the peak of the initial score distribution stays in the higher points than that in Fig. 1a, the trend of the peak shifting with the increasing attempt number would have a similar trend to that of Fig. 1a; most of learners earn more than 80 marks by 10 attempts or less.

In Fig. 1b, we plot the more difficult exam case which does not have a symmetric curve shape of the initial score distribution and the maximum percentage of learners stay in the range of 0-20 marks. The peak shifts slowly from the lower marks (0-20 points) to the higher marks (80-100 points) with the increasing number of attempt times. Fifty seven (57) percents of learners earn more than 80 points by 10 attempts or less which percentage is still lower than the final result (71 percents). It is found that as shown in the illustration of Fig. 1b, the attempt number of a difficult exam case becomes larger than twice of that of the typical case, though the attempt numbers are always to be within 30, typically 15-25. It indicates that the difficult exam cases decrease learning efficiency and waste learning time of students. As shown in Fig. 2, if we focus our attention on the initial score distributions of individual tests, we could see that the curves have different shapes, and the peaks take place in a wide range.

It should be considered that a curve with a peak near the lowest (0-20) or highest (80-100) range may not be good in measuring the comprehension of learners. We therefore need to pay attention to a test with a curve of a peak in a lower range like Fig. 1b because they would indicate that the test may include inappropriate questions which may not only weaken but also spoil motivation of learning.

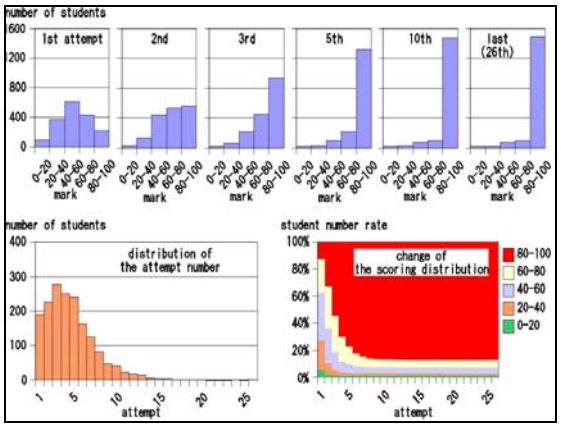

Fig. 1a: The typical score distributions, the distribution of the attempt number of times, and the change of scoring distribution.

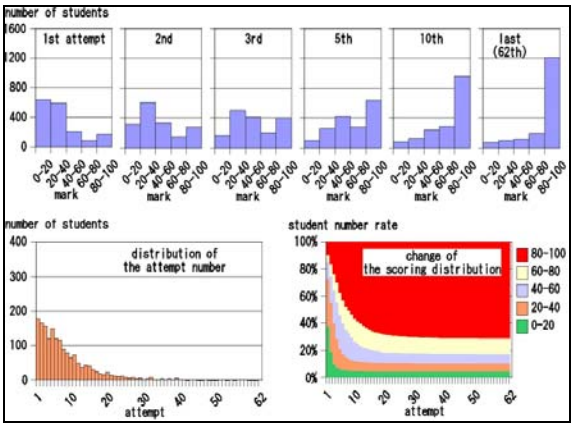

Fig. 1b: Same as in Fig.1a, but the more difficult exam case.

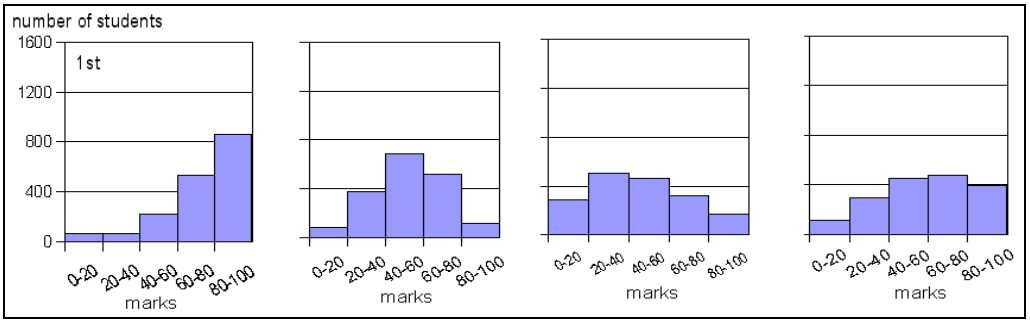

Fig. 2: The score distributions of the initial attempts. The envelope curves have different shapes, and the peaks are in a wide range.

Improvement of the on-line tests

We are offering about three hundreds on-line questions to be accessed as the exams of the Basic Course of Information Technologies A/B in the LMS. All of the questions have first been created by instructors of the course. Before offering those exams, quality of the questions are assessed and improved very carefully so as to be qualified with efforts described in the previous section. All of the questions are monitored, assessed, evaluated, and improved by other staffs who have not created the questions. It has been found that the questions used in the on-line exams of the information literacy course are well qualified, i.e., yes, reasonable, or valid, to most of the checking items. Almost all of the questions have been fully qualified for the items in the categories of contents of learning, expressions, and strictness.

It has also been found that the number of result “no” for the items in the categories of the difficulty levels and the selection choices is much more than those of terms in other categories. We provide checking items with a reasonable evaluation level in Table 2. It is shown that the worse results to the items listed in Table 2 correlate with each other. When we carefully review the questions with the worse answer for those four items, we see some similarity of the questions: there is just a little difficult expression or a paradoxical statement in the questions. In addition, other questioning form can be employed, such as changing from a multi-choice form to an input form. There are a few worse results to terms in other categories, but we suppose that those are substantively attributed to the effects of the worse results in the two categories in Table 2.

We have carried out improvement of questions for the Basic Course of Information Technologies A. We have paid attention to eliminate questions that may mislead learners and give a fallacy. Also, we have taken two approaches to improve questions requiring adjustments based on the difficulty levels. One is to change statements of the questions, and the other is to divide the questions into several new questions in order to verify understanding of learners about essential points of the subject.

In Fig. 3a and 3b, we illustrate the initial score distributions before and after the improvement. Those are representative examples of changes of the initial score curves. In both figures, the left histograms are the score distributions before the improvement and the right histograms are those after the improvement. The former (Fig. 3a) and the latter (Fig. 3b) reflect the statistical data of the Basic Course of Information Technologies A of the year of 2006 and 2007 respectively. The figures demonstrate that the scores of learners have evidently shifted to higher range after the improvement, and consequently we have become able to manage the improvement effectively. It may also be obvious that as shown in Fig. 3a, the improvement is more effective in the case that initially the points are distributed in a lower score range.

Table 2: Checking items with low evaluation.

Categories	Checking Items
Difficulty levels	■ There is a room for improvement by adjusting the difficulty level. ■ The problem statement is somewhat decipherable.
Selection choices	■ The question may intentionally mislead learners. ■ Learners may elicit the right answer even if he/she has a fallacy or mistaken notation.

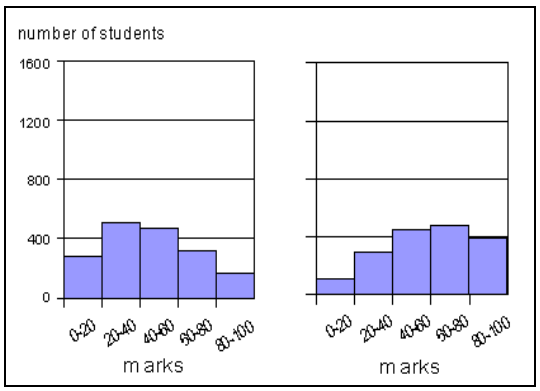

Fig. 3a: The initial score distribution before (left) and after (right) improvement with the checklist.

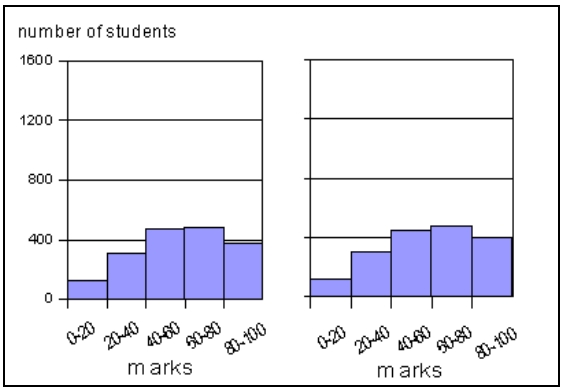

Fig. 3b: Same as Fig. 3a, but a different exam case.

Concluding Remarks

We have shown the development of the on-line tests for information literacy education with the use of a checklist. Since many prior studies indicated various concepts and views to make good assessments, we could summarize them and create the checklist with 45 checking items, and we have used it for improvement of our on-line tests. It has been found that the questions used in the on-line tests of our information literacy education have been qualified with better results to most of the checking items. Particularly, almost all the questions have been qualified with full marks for the items in the categories of contents of learning, expressions, and strictness. We have specifically focused on the improvement of the tests by checking the initial score distributions with peaks in a lower score range. Then we have compared the initial score distributions of the last year (2006) in the basic Courses of Information Technologies A with that of this year (2007). It has been found that most of learners of this year of 2007 could earn higher scores with the improved on-line tests. Thus we could conclude that the improvement works substantially.

We consider that the concept of our checklist presented here could be used in checking on-line tests in other fields than information literacy education. Meanwhile, it has been considered that the checklist may have too many items because we have tried to list up items as many as possible. Further, items in some categories could always lead to full marks. It may indicate that questions created by any instructors are intrinsically satisfied with those items, and all items may not necessarily be required to be checked thoroughly. Thus, to develop better on-line test materials, even the checklist shown here could be carefully revised and the items could be reduced down in an appropriate number.

References

Williams, G. & Bialac, R., (2006). Using Online Self-Assessment to Promote Independent Learning., Society for Information Technology and Teacher Education International Conference (SITE) 2006:1

Ozkan, B. C. & Morris. B. J., (2006). Student assessments as a tool to improve online teaching. World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (ELEARN) 2006:1

Usagawa, T. & Nakano, H., (2005). Campus-wide IT literacy education for all students by effective e-Learning environments, Proceedings of the 4th International Conference on Emerging e-learning Technologies and Applications (ICETA 2005).

Nakano, H., Iriguchi N., Sugitani, K., et. al. (2005). Proceedings of the 6th International Conference on Information Technology Based Higher Education and Training (ITHET 2005) F4B, 7-11.

Hricko, M. & Howell, S. L., (2006). *ONLINE Assessment and Measurement Foundations and Challenges*, Information Science Publishing, Hershey USA.

Kahn, B. H., (2005). *E-Learning Quick Checklist*, Information Science Publishing, Hershey USA.

Kishi, M., http://www.u-gakugei.ac.jp/~kishilab/TESTindex.htm

---

[i] The detailed statistical data are gathered directly from the logs using a Perl script we have developed.