

An Online Testing and Analysis System for Creative Problem-Solving Ability in Sciences

^aChun-Chieh Huang, ^bHao-Chuan Wang, ^aTsai-Yen Li and ^cChun-Yen Chang

^aNational Chengchi University, Taipei, Taiwan

^bCarnegie Mellon University, Pittsburgh, PA, USA

^cNational Taiwan Normal University, Taipei, Taiwan

Email: g9415@cs.nccu.edu.tw, haochuan@cs.cmu.edu,
li@nccu.edu.tw, changcy@ntnu.edu.tw

Abstract: *This paper proposes an online testing and analysis system for studying students' Creative Problem-Solving (CPS) ability in sciences. Using an open-ended essay-question-type test, students are asked to express their idea and imagine how to solve problems better. Based on previous works, we utilize an automated scorer for evaluating students' CPS ability. This system serves as a real-time (self-)assessment for online learners and a useful research tool to gather data for CPS studies. We have re-examined the inter-rater reliability of our automated scorer with new samples, and conducted a questionnaire survey for usability inspection. The preliminary results show the system is reliable for automated scoring and satisfactory for system usability. A roadmap for future development is also proposed.*

Keywords: Online Testing, Creative Problem Solving, Automated Scoring, Web-based Learning

1 Introduction

The assessment of Creative Problem-Solving (CPS) ability is taken as a useful index to investigate students' creativity and problem solving abilities in psychological and educational research (Chang & Weng, 2002). We generally believe that using assessment of an open-ended essay-question-type format is a useful approach to evaluate students' CPS ability. Students are required to express their self-generated thoughts in natural languages (e.g., Chinese) in such a test, and teachers or researchers may analyze students' essays, one of the most comprehensive and authentic representations of students' thinking and reasoning, to study their comprehension, competence or creativity regarding the domain subjects. In the context of web-based education (WBE), it is considered essential and beneficial to incorporate this type of assessments as a valuable component of online learning environments for better instructional decision-making and more advanced educational studies.

Nevertheless, although online courses have been developed and deployed extensively more than ever, online testing, especially open-ended essay-question-type tests, is still underemphasized. One possible factor is that technologies required for online testing were not systematically developed and incorporated by practitioners of WBE. Though the development of certain types of testing, such as dichotomous-, single- and multiple-choices questions, may be intuitive and mature, the situation is very different from the case of open-ended essay-question-type testing, in which technologies for language processing and learner modeling can be more complicated and were less familiar to the

WBE community. In sum, we believe that it is crucial to conduct systematic investigation on the topic of online testing for open-ended problem solving tasks in order to further the effect of online learning.

Grading is a great burden to teachers, especially for open-ended testing. With limited educational resources and given a large number of students, a prompt and valid way to evaluate students would be very useful to teachers. An approach to solve this problem is by adopting an automated scorer, made possible by Artificial Intelligence technologies, to grade a large volume of open-ended testing.

In this research, we aim to develop an online testing system suitable for science education to assess students' CPS abilities through open-ended essay-question-type testing. The system is designed to support three major functions: (1) students' online answering to the test questions, (2) reliably automated processing and grading upon students' responses, and (3) on-line inspection of statistical results with graphical visualization tools for teachers and researchers. More specifically, this work consists of three parts: (1) developing a fully integrated system that supports the aforementioned functionalities, (2) conducting a usability study of the user interface and interaction design of this system, and (3) re-testing the reliability of our automated scoring module, which was preliminarily shown to have a satisfactory performance for a small sample of students (Wang et al., 2005a). In this work, we have conducted a new empirical exploration with 70 students involved in the studies of system usability and grading reliability. In addition, we will also introduce some techniques to implement such a system, such as controlling students' online resource access and collecting statistics about students' scores.

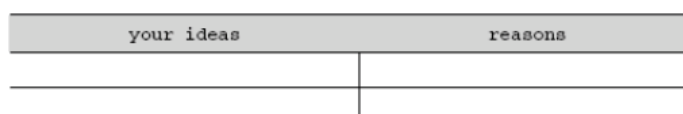
2 Previous Works

The system described in this paper is based on our previous works in this series, including theoretical and empirical investigation of students' CPS in sciences (Chang & Weng, 2002), user modeling for CPS (Wang et al., 2005a), and automated grading technologies designed for CPS studies (Wang et al., 2005b). The work pertaining to the research in this direction is described as follows.

2.1 CPS: Creative Problem-Solving

Problem-solving ability is viewed as an ability to think critically, reason analytically, and create productively, which all involves quantitative, communication and critical-response skills (American Association for the Advancement of Science, 1993). The system we have built is based on the theories of several measurements of creativity proposed in the literature (Hocevar, 1981; Yu et al., 2005).

In our previous work on designing the paper-based assessment for CPS (Wu & Chang, 2002), we have taken *Problem-Finding* and *Problem-Solving* as different testing phases, and each phase is further separated as two parts: ideation (i.e., divergent thinking) and explanation (i.e., convergent thinking). The separation of such reasoning phases was originated from some theories of CPS (Basadur, 1995; Osborn, 1963). The paper-based answer sheet was designed with a two-column table for students to express their ideas and explanations, as showed in Figure 1.



your ideas	reasons

Figure 1. A snapshot of the answer sheet showing the pair-wise relation between ideas and reasons.

In each row of the table, students may fill in two consecutive text fields with an idea and the reason(s) explaining that idea for a specific problem-solving scenario described in the testing question. More details and discussions of the design for this assessment can be found in Wu & Chang (2002) and Wang et al. (2005a, 2005b).

2.2 UPSAM: User Problem-Solving Ability Modeler

Our prior work has also proposed a user modeling system, User Problem-Solving Ability Modeler (UPSAM), which aims at abstracting and modeling students' responses in scientific problem-solving tasks (Wang, 2005a). By representing experts' and students' thoughts of problem-solving, including both ideas and reasons, in the formalism of bipartite graph, ideation and explanation are modeled as two disjoint sets connected with undirected edges. UPSAM enables us to compare any two bipartite-graph-based user models to address research questions such as *how well are this student's answers by using the expert's answers as the criteria* (i.e., essay grading) and *what concepts are known/unknown by the student* (i.e., learning diagnosis).

An important feature of the UPSAM is that it maintains an *expert model* representing the domain experts' knowledge and vocabulary against the problem-solving tasks, and it serves as a useful module to process students' open-ended answers and perform automated grading.

2.3 Automated Scorer

Based on UPSAM, we have developed an *automated scorer* for grading students' responses toward the CPS test. The core concept is to use Information Retrieval (IR) technologies to match students' responses against experts' standard answers and then represent each student as a subset of the bipartite-graph-based expert model (Wang et al., 2005b). We have collected many experts' ideations and explanations from different aspects, and organize them into an expert model, which is aimed to be objective and equitable. Because of the structured answers format from UPSAM, the automated scorer can grasp students' key concepts in order to evaluate their CPS ability based on each text entry. In addition, we also invite human graders to grade manually for comparisons.

In this prior work, we have conducted an evaluation to corroborate the viability of the automated scoring scheme empirically. We found the Pearson's correlation statistics r , which serves as an index of inter-rater reliability between human graders and our automated scorer, range from .67 to .82, which are positive, high, and statistically significant. The preliminary experiments showed that automated scoring is reliable to be used here, and it is also suggested that an integrated online system should be developed for testing, grading and further analyses (Wang, 2005a).

3 Online Testing System: WOTAS

The main objective of this paper in this line of research is to build an online system integrating our previously developed components with the newly designed modules (e.g., UI for testing and statistics reporting.) We have designed and implemented the Web-based Online Testing and Analysis System (WOTAS) such that a great number of students can access the system online concurrently with great stability and usability and teachers or researchers can easily diagnose the learning condition from those learners. The architecture of WOTAS is described below.

3.1 Design Considerations

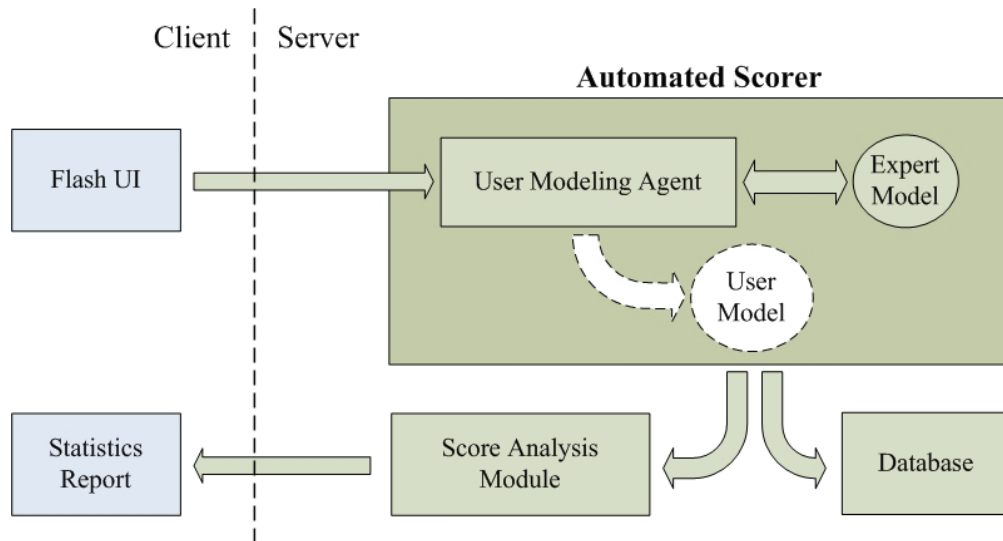


Figure 2. System architecture for WOTAS.

The system that we have design, as depicted in Figure 2, is divided into the client side and server side for different purposes. On the client side, the Flash-based User Interface (UI) is designed as a learning environment to collect students' answer texts. As described earlier in Section 2.1, the paper-based version of the CPS test comprises two testing phases. The first phase is for the students to write down their answers of *Problem-Finding*, and the second phase is for them to write down their answers of *Problem-Solving*. In these phases, students are asked to produce their ideas and point out the reasons of each idea, and then fill them in divided answer cells. Comparing to typical essays which are less structured, our CPS test was designed to be more structured with explicit identification of divergent and convergent thinking such that the answers are easier to be processed by computers. Our Flash-based UI for testing or collecting students' answers is currently designed to reflect the features of the original paper-based test, in which two testing phases and the ideation-explanation structure were realized through appropriate layout on separated sheets.

When a student's answers were posted to the server, the User Modeling Agent on the server side will first process them in order to generate the corresponding user model for further processing in the Automated Scorer Agent. In the students' user models, the relations between the ideations and explanations were represented as a subset of the prescribed bipartite-graph-based Expert Model. Based on the expert model, the Automated Scorer agent will then assign numeric scores to each student's answers, which are then used to compute each user's CPS ability score. All of the raw or processed data are stored in the database module depicted in Figure 2, and these data can be reported in various forms according to the designated purpose.

As for teachers or researchers, the system uses the Score Analysis module to show students' scoring results and overall statistics tables. Teachers may consider it useful to find those students who need learning helps and discover new ideas or concepts from students' answers. For researchers, it is an effective analysis tool for their evaluation of the experiments.

3.2 Implementation



Figure 3. Flash interactive User Interface in the front side of the online testing system.

In the client side of the testing system, we use the popular interactive technology on the web, Macromedia Flash (<http://www.adobe.com/products/flash/>), to build a rich user interface aiming to bring a vivid environment for them to deeply get involved. The environment is also designed to be non-preemptive such that the users are not allowed to access other web resources when the testing is in session. It ensures that all users can only access the same information even though the computers are connected to the network or internet.

In the testing process, the users will first be presented an introduction of this test. After signing in the test, the system starts a countdown of 25 minutes, and users will enter a simulated problem-based situation. The users will be prompted with a detail description of the scenario for questions as showed on the left of Figure 3, and given an answer sheet for them to write down their answers in the slots shown on the right of Figure 3. Students are allowed to go back and forward between the question and the answer sheets. No matter the users have finished or not, after 25 minutes, the program will send the answers that it has collected to the server. Finally, there is a questionnaire to acquire their opinions about the usability of this online testing system.

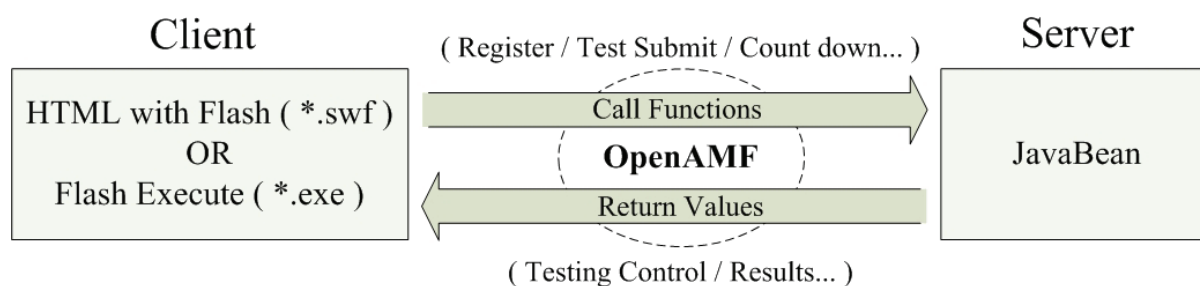


Figure 4. The Flash client connects to a Java server by using OpenAMF remoting

In order to send the collected data from the client to the server, we need a dynamic web-page technology that facilitates the data transfer. Here, we use an open-source toolkit called OpenAMF (<http://www.openamf.org/>), as sketched in Figure 4 for transferring data from a flash program to a Java-based server program. An advantage of using this technology is that the developers do not have to handle lower-level implementation details between heterogeneous programming environments,

such as Flash and Java. With the component-based object-oriented designs brought by the Flash and Java technologies, developers do not have to design a series of web pages connected with hyperlinks in order to fulfill a specific application goal. It is unlike other dynamic web-page technologies, such as JSP/PHP/ASP, which typically operate information “page by page”. A component-oriented technology on the client side like Flash and an object-oriented technology on the server side like Java jointly provide an ideal environment for modularizing the system. Such modularization is considered important in developing a sophisticated online testing system. We believe that the technologies that we have used should be noteworthy and informative to other works of this kind in the WBE community.



Figure 5. Statistics report module for students' scores and survey results

A report module is designed to present the complete detail answers that students wrote and the statistics tables of students' CPS scores graded by automated scorer and human graders, as shown in Figure 5. Dynamic statistical visualization, which designed and produced by ColdFusion (<http://www.adobe.com/products/coldfusion/>), is incorporated for teachers or researchers to analyze the distribution of students' scores and to diagnose students' misconceptions for future instructional decision-making.

4 Empirical Explorations

The purposes of our experiments on empirical explorations are to re-test the reliability of the automated scoring mechanism and to survey the system's usability. The study was conducted in a public high school in middle Taiwan, with 70 students involved in the evaluation.

4.1 Re-testing the reliability of automated scoring

In the pervious pilot study of empirical evaluation with 20 students, we have found that the correlation statistics r for human-to-computers inter-rater reliability range from .67 to .82, which are high and statistically significant. In this research, with a sample of 30 students' data from the pool of the 70 students involved in this study, we intend to re-test the machine scorer's reliability compared with the grading results derived by human graders. The result derived in this study by comparing the results by automated scorers and two human graders reveals that the Pearson's correlation r achieves .87 and .92, which indicates a high and positive correlation. The result is consistent with our previous finding of the positive reliability of the automated scoring modules as described in Wang et al. (2005b).

4.2 Usability Study

After finishing the CPS online test, students were asked to complete a questionnaire for us to investigate the usability of the user interface and the test flow. Those questions and the results are shown as follows in Table 1.

Table 1. The statistics of online questionnaire after using the online testing system ($n = 70$).

	On a rating scale of 0 to 7	Mean	STD
Q1	I can easily use this Online Testing System, and there is no any problem when I answer to every question.	3.52	1.66
Q2	I am used to this interactive browsing mode, and know the testing flow very well.	4.10	1.50
Q3	I clearly know the difference between “Ideas” and “Reasons” when I answer those questions.	3.55	1.61
Q4	I consider the creativity was produced by me myself, so I do not need other facilities to guide me.	3.68	1.72
Q5	At the online testing, if I can interact with others, such as MSN or Chat Room, I think I will be more creative.	4.68	1.69
Q6	If the system could return my testing score, I would like to use it as an evaluation tool for self-study.	4.47	1.80

Table 1 gives the mean rating, the standard deviation for six aspects of this system. Every student gives each question a score between 0 and 7. We find Q1 and Q3’s average ratings are close to 3.5, which means “No suggestion” on this scale (i.e. $(0+7)/2 = 3.5$). The mean of Q1 reveals that some students are not used to using the online testing system. Therefore, we may have to consider how to enhance students’ familiarity to the system in the future. The mean of Q3 may indicate that some students have some problems to tell the difference between ideation and explanation. We think the possible cause may include that the introductory explanation of the test is not informative enough for the students to comprehend.

The means of Q5 and Q6 are close to 4.5, which mean “I agree” in this scale. This could suggest that students have the demands of further interactive learning tools to improve their creativity learning. Therefore, we think it is reasonable for us to have the supposition that online interactive learning system could be useful for students to learn creativity although it needs to be verified in the future.

5 Future Development Roadmap

Our future work includes four parts: 1) enhancing the usability 2) designing an automated feedback faculty 3) enhancing the grading performance by a self-learning mechanism, and 4) improving the analysis module, as showed in Figure 6.

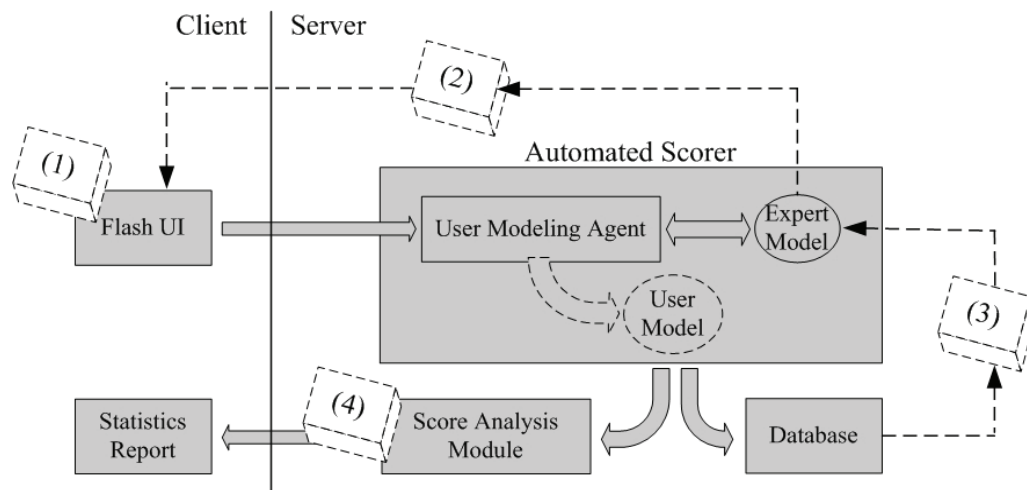


Figure 6. The future roadmap of this Web-based Online Testing and Analysis System.

1) Enhancing usability: We believe that a good using experience will lead to better testing and learning in CPS. According to our usability study, the users' feedbacks reveal that there is still room for improvement in several aspects of online testing, such as clear introduction, user interface design, interactive operations and so on. A good way for improvement is to have a focus group discussion with the students who had given lower scores on this issue. Another possible way is to record the user's interaction with the system in the course of the test process and then analyze these acquired data to propose improvement on the interface design.

2) Designing an automated feedback faculty: Several recent works on Intelligent Tutoring System (ITS) pay much attention to natural language-based tutorial system with conversational dialogues (Grasser et al., 2001). We are also planning to design a tutoring agent who can interact with students by providing timely hints in order to help the students to think critically. Furthermore, we need to conduct more studies on the nature of creativity in order to design a system that can help the students produce their ideas more creatively.

3) Enhancing grading performance by self-learning ability: We consider the self-learning ability very important for the automated scorer to improve grading quality because human's grading standards are revised according to their experience as the time goes by. In our system design, we will realize the self-learning ability with a feedback mechanism in the automated scorer. For example, a practical method would be to add a feedback inspector between the expert model and the automated scorer agent. The job of the inspector is to find better answers or more creative solutions in order to extend the expert model. We hope that with the improving expert model, the system can enhance its grading quality incrementally as the user basis expands.

4) Improving analysis module: For teachers or researchers, a desirable feature of the test system would be to find those students who need learning helps. Through the analysis of score comparison and qualitative text answers, we can easily grasp the students' learning situation. Therefore, we would like to develop a more advanced analysis module that can support the common analysis methods expected by the teachers.

6 Conclusions

In this paper, we have described the design and implementation of an online testing and analysis system for open-ended CPS tests. The system is designed based on prior works investigating theoretical and empirical foundations of CPS as well as technologies for modeling and measuring the CPS ability. In the implemented system, we have used the web technologies to integrate our previously developed components for user modeling and automated scoring and a newly designed interactive user interface for students to do problem-solving tasks. A web-based tool has also been designed to help teachers and researchers analyze the test results. A replicated correlation study containing 30 sampled students' data shows that our automated scorer is competent compared to human graders. In addition, we have conducted a usability study via a survey in order to understand usability issues and suggest future improvements. Based on this framework, we have also proposed the roadmap of future developments for improving various aspects of the system. These ongoing works will be reported in more details soon.

Acknowledgement

This research was funded in part by the National Science Council (NSC) of Taiwan under contract no. NSC 94-2524-S-003-014. We also thank the Chinese Knowledge and Information Processing (CKIP) group at Academic Sinica for providing the service of Chinese word segmentation.

References

- American Association for the Advancement of Science* (1993). Benchmarks for science literacy. New York: Oxford University Press.
- Basadur, M. (1995). Optimal Ideation-Evaluation Ratios. *Creativity Research Journal*, 8(1), pp.63-75.
- Chang, C-Y., Weng, Y-H. (2002). An Exploratory Study on Students' Problem-Solving Ability in Earth Sciences. *International Journal of Science Education*, 24(5), pp. 441-451.
- Grasser, A. C., VanLehn, K., Rose, C. P., Jordan, P. W., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. *AI Magazine*, 22(4), 39-51.
- Hocevar, D. (1981). Measurement of creativity: review and critique. *Journal of Personality Assessment*, 45. pp.450-464.
- Osborn, A. (1963). *Applied Imagination: Principles and Procedures of Creative Problem Solving*. New York: Charles Scribner's Sons.
- Wang, H-C., Li, T-Y., & Chang, C-Y. (2005a). A User Modeling Framework for Exploring Creative Problem-Solving Ability. *Proceedings of AIED Conference*, Amsterdam, The Netherlands.
- Wang, H-C., Chang, C-Y., & Li, T-Y. (2005b). Automated Scoring for creative problem-solving ability with ideation-explanation modeling. *Proceedings of 13th International Conference on Computers in Education (ICCE 2005)*, pp.522-529.
- Wu, C-L., Chang, C-Y. (2002). Exploring the Interrelationship Between Tenth-Graders' Problem-Solving Abilities and Their Prior Knowledge and Reasoning Skills in Earth Science. *Chinese Journal of Science Education*, Vol. 10, No. 2, pp. 135-156.
- Yu, Z-J., Lin X-T., & Yuan, X-M. (2005). The Framework of Web-based Versatile Brainstorming System. *Global Chinese Conference of Computers in Education (GCCCE 2005)*, pp.502-510