The Current Applications of Simulations in Computer-based Science Assessments

Che-Yu KUO[a*] & Hsin-Kai WU[a]
[a]*Graduate Institute of Science Education, National Taiwan Normal University, Taiwan*
*chkuo0108@gmail.com

Abstract: The purpose of this review study is to examine how science simulations are used for assessment purposes. Eleven assessments were identified from 67 reviewed computer-based science assessments. Based on these 11 assessments, we concluded that by including simulations, science assessments would measure critical competences that are authentic and meaningful. By focusing on identifying meaningful features of response units and patterns of interaction data, automated scoring of simulations could be feasible and facilitated the applications of simulation into assessment practices. The simulations are thus able to serve as curriculum embedded assessments by taking advantages of evaluated interactions to offer contingent scaffolding to students while they are engaging in learning activities.

Keywords: science simulation, automated scoring, curriculum-embedded assessment

Introduction

Simulations refer to technology-enabled applications or environments imposing dynamic representations, mimicking the behavior of a system or in a situation, and allowing manipulations of the system or active involvement in the situation. In simulations, students construct their understandings of depicted situations and conduct scientific investigations. Simulations thus have been advocated as a powerful tool to make complex cognitive competences visible so as to more closely tied to the constructs and skills of interest (National Research Council, 2001). It has also been argued that simulations entail potentials to support learning by offering interactive activities and learner feedbacks, deemed an embedded assessment (Means & Haertel, 2001). However, a recent review about science simulations suggested that only few studies took advantages of simulations for assessments (Scalise, Timms, Moorjani, Clark, Holtermann, & Irvin, 2011). One obstacle to preclude from incorporating simulations into assessment activities may lie in scoring large amount of interaction data from the simulations. Therefore, this review study aims to examine the current status of simulations in assessment practices and delineate ways in which researchers and educators in science leverage simulations to support assessment activities.

1. Scope of the Study

Our review identified 67 computer-based assessments related basic science and medical disciplines from 1985 to June, 2012 in the ERIC (Educational Resource Information Center) digital library and the SSCI (Social Sciences Citation Index) database. Four categories including text-based items (TI), graphic-based items (GI), animation items (AI), and simulation items (SI) were used to identify different degrees of interactivity and multimedia inclusion in the assessments. This study particularly focused on the assessments adopting simulations. This study examined the related constructs and skills of interest in the simulations, how the performances on the simulations were scored, and how the simulations were used for curriculum-embedded assessments.

2. Simulation Tasks in Computer-based Science Assessments

2.1 The Prevalence and Related Constructs and Skills of Interest

Among 67 assessments, simulations were adopted by 11 assessments and the least among the four categories (TI = 29.8%, GI = 26.8 %, AI = 12.0%, SI = 16.4%). In addition, 5 among the 11 assessments were related to medical disciplines. Three of these 5 presented clinical cases depicted simulated interactions between the practitioners and patients in video clips (Bersky, 1994; Fizergerald, Wolf, Davis, Barclay, Bozynski, Chamberlain, Clyman, Shope, Woolliscroft, & Zelenock, 1994; Stevens, Ikeda, Casillas, Palacio-Cayetano & Clyman, 1999). Examinees could study patient information, request consultation and ask for further tests, and were required to offer procedures, treatments and other decisions that were necessary and appropriate as the cases proceeded. Another two assessments presented visualized and dynamic models of organ systems in clinical conditions over time (Shaw, Effken, Fajen, & Garrett, 1997). Students were required to monitor the conditions and offer appropriate amount of treatment to achieve optimal conditions.

On the other hands, 6 assessments presented tasks related to basic sciences. Two assessments allowed students to access various lab tests and other resources (for non-data information) to solve the simulated domain-specific problems (Chung, Vries, Cheak, & Stevens, 2002; Vendlinski & Stevens, 2002). Students thus expressed their scientific reasoning skills without risk of test failure in the lab. Another assessment designed a simulated Web environment to evaluate students' skills of seeking information and making use of searched information to solve the problem (Schacter, Herl, Chung, Dennis, & O'Neil, 1999). Two assessment systems employed microworlds that allowed students to work on scientific investigations of the depicted science phenomena with running trials in experiments (Bennet, Persky, Weiss, & Jenkins, 2010; Quellmalz, Timms, Silberglitt, & Buckley, 2012). While working on these microworlds, students were exposed to multiple representations of abstract, complex and dynamic phenomena, and created models at different temporal and spatial scales from what were experienced in everyday lives.

Together, the reviewed studies suggest that simulations not only presented science problems in variety of situations from realistic problems to abstract, dynamic and unobservable phenomena but also allowed students to demonstrate a wide range of problem solving skills in these problems. Simulations thus hold promise for measuring significant problem solving skills required in both academic and clinical settings.

2.2 Scoring procedures

Scoring of students' performances in simulations focused on the presence and the quality of response units (each unit comprising a series of actions in the simulation and centering on the same theme), and concerned with the combination of multiple units (response patterns). The identified simulations more commonly focused scoring on the presence of response units (6 of 11) and response patterns (4 of 11). Fewer (3 of 11) evaluated the quality of response units and only one assessment made judgments on the quality of the overall performance. Below, the two approaches concerning the presence of response units and patterns were discussed because automated scoring procedures were developed and implemented only with these approaches.

In the first approach, students' performances were most frequently evaluated by matching single units of responses to predetermined scoring keys representing experts' performance in the areas. The response units were quantifiable by simply matching transaction data with a checklist and thus feasible for automated scoring. Four of these 6 assessments adopted automated scoring for evaluating the response actions (Bersky, 1994; Clauser, Margolis, Clyman & Ross, 1997; Quellmalz, Timms, Silberglitt, & Buckley, 2012, Schacter, Herl, Chung, Dennis, & O'Neil, 1999). Variations of this approach gave different credits to different response units. For example, Clauser, Margolis, Clyman and Ross (1997) asked experts to determine necessary and appropriate treatments in the simulated practitioner-patient interaction and assigned higher credits to those treatments of importance for optimal care.

Four assessments evaluated performances by the second approach and through matching with specific combinations of response units either articulated by experts or derived from empirical data (Clauser, Margolis, Clyman & Ross, 1997; Stevens, Ikeda, Casillas, Palacio-Cayetano, & Clyman, 1999; Vendlinski, & Stevens, 2002). For example, Steven et al. (1999) reported applications of neuron networking modeling that extracted exemplars of the response patterns from cases at different levels of proficient competence so as to evaluate new performance data. Because complex path maps were used to entail information about the sequence of access to information resources during students' scientific reasoning, the result showed discernible strategy changes in the extracted exemplars representing proficient competence at different levels. The four assessments all adopted automated scoring, possibly suggesting only technology-enabled procedures can process numerous or even infinite possible combinations of response units taken by examinees in simulations.

2.3 Curriculum-embedded Assessments

Among 11 assessments, three assessments were designated as curriculum-embedded assessments since these assessments offered learning activities while providing information about learning progress through unobtrusive assessments. The first two assessments were developed by a situated assessment tool but focused on different principles of clinical conditions (Shaw, Effken, Fajen, Garrett, & Morris, 1997). The tool offered the potentials of immediate feedbacks from instructors through monitoring graphic paths representing dynamics of organ systems over time. The system also served as an online teaching assistant to give instructors advice signals based on ongoing evaluation of the changes of dynamics. The third assessment was a multi-level, balanced state science assessment system offering opportunities for students to engage scientific investigations in a microworld (Quellmalz, Timms, Silberglitt, & Buckley, 2012). In its formative assessment, the system offered adaptive scaffolding to students. Namely, the system included informative feedbacks and graduated levels of coaching that varied from identifying the error responses to pointing out simulated results of students' experiments. These three assessments thus suggested that simulations could be part of an integrated assessment closely tied to learning activities. Particularly, the assessment system could offer contingent feedbacks and coaching by making use of information that was generated by simulations in response to student involvement. Such contingencies only followed automated evaluation of interaction data that were instant and ongoing.

3. Conclusion

The findings of this study revealed that less than one sixth of the reviewed assessments adopted science simulations although simulations showed promise for taping a wide range of critical competences that were authentic and meaningful. The findings also suggested that scoring of the large amount of interaction data from simulations focused more on identifying meaningful performance features from both response units and response patterns. Automatizing such evaluation undoubtedly may facilitate applications of simulations into assessment practices. By including automated evaluation, assessments can make use of the interaction data to offer immediate and contingent scaffolding for students while they are working on simulations.

Acknowledgements

This research was supported by the National Science Council of Taiwan under NSC 100-2811-S-003-009 and NSC 100-2511-S-003-041-MY3.

References

[1] Bennett, R. E., Persky, H., Weiss, A. & Jenkins, F. (2010). Measuring problem solving with technology: A demonstration study for NAEP. *Journal of technology, Learning and Assessment, 8(8).* Retrieved 19, August, 2012 from http://www.jtla.org.

[2] Bersky, A.K. (1994). Effect of audiovisual enhancement on problem solving and decision making activities during a computerized clinical simulation test of nursing competence. *Evaluation & the Health Professions,17* (4), 446-464.

[3] Chung, G. K. W. K., Vries, L. F., Cheak, A.M., Stevens, R. H., Bewley, W. L. (2002). Cognitive process validation of an online problem solving assessment. Computers in Human Behavior, 18, 669-684.

[4] Fizergerald, J. T., Wolf, F. M., Davis, W., Barclay, M. L., Bozynski, M. E., Chamberlain, K. R., Clyman, S. G., Shope, T. C., Woolliscroft, J. O & Zelenock, D. B. (1994). *Evaluation & the Health Professions, 17* (3), 307-321.

[5] Means, B. & Haertel, G. (2001). Technology supports for assessing science inquiry. *Technology and Assessment: Thinking ahead*. (pp. 12-25). Washington, D.C.: National Academy Press. Retrieved 19, August, 2012 from http://www.nap.edu/catalog/10297.html

[6] National Rearch Council. (2011). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.

[7] Quellmalz, E. S., Timms, M. J., Silberglitt, M. D. & Buckley, B. C. (2012). Science assessments for all: Intergrating science simulations into balanced state science assessment systems. Journal of Research in Science Teaching, 48 (3), 363-393.

[8] Scalise, K., Timms, M., Moorjani, A., Clark, L., Holtermann, K. & Irvin, P. S. (2011). Student Learning in Science Simulations: Design Features that Promote Learning Gains. *Journal of Research in Science Teaching, 48*(9), 1050-1078.

[9] Schacter, J., Herl, H. E., Chung, G. K., Dennis, R. A., O'Neil, H. F., & Jr. (1999). Computer-based performance assessments: A solution to the narrow measurement and reporting of problem-solving. *Computers in Human Behavior , 15*, 403-418.

[10] Shaw, R.E., Effken, J, Fajen, B.R., Garrett, S.R. & Morris, A. (1997). An ecological approach to the on-line assessment of problem-solving paths: Principles and applications. *Instructional Science, 25* (2), 151–166.

[11] Stevens, R., Ikeda, J., Casillas, A., Palacio-Cayetano, J. & Clyman, S. (1999). Artificial neural network-based performance assessments. *Computers in Human Behavior, 15*, 295–313.

[12] Vendlinski, T., & Stevens, R. (2002). Assessing student problem-solving skills with complex computer-based tasks. *Journal of Technology, Learning, and Assessment, 1* (3). Retrieved 19, August, 2012 from http://www.jtla.org.