DOCUMENT RESUME

ED 361 207 SE 053 623

AUTHOR	Helgeson, Stanley L.; Kumar, David D.
TITLE	Technological Applications in Science Assessment.
INSTITUTION	National Center for Science Teaching and Learning, Columbus, OH.
SPONS AGENCY	Department of Education, Washington, DC.
PUB DATE	Apr 93
CONTRACT	RI17Q00062
NOTE	11p.; Paper presented at the Annual Meeting of the National Science Teachers Association (Kansas City, KS, April 1-4, 1993).
PUB TYPE	Information Analyses (070) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Educational Technology; Elementary Secondary Education; Evaluation Methods; Formative Evaluation; *Hypermedia; Microcomputers; *Science Education; Testing
IDENTIFIERS	*Science Achievement; Science Education Research

ABSTRACT

Educational technology has been a focus of development and research in science teaching and learning. This document reviews research dealing with computer and hypermedia applications to assessment in science education. The paper reports the findings first for computer applications for assessment and then for hypermedia applications in assessment. Computer applications include access to large test banks, evaluation of student data collected in laboratory exercises, microcomputer-administered diagnostic testing, effects of computer use on student motivation during computer-administered tests, and computerized adaptive testing. Hypermedia applications include interactive videodisc technology for teaching biology, assessing teachers' knowledge for certification purposes, problem-based learning in chemistry found in the "Hyperequation" project, and administering figural response test items to cell and molecular biology students. The summary describes three advantages for incorporating computer-assisted assessment: immediate student feedback, ease of test taking, and availability of another form of formative assessment. Users of computer-assisted assessment are cautioned that computer-generated tests can lull the teacher into simply testing low level recall knowledge, and that the linear nature of computer testing does not allow the student to go back and reflect upon particular items. (MDH)

TECHNOLOGICAL APPLICATIONS IN SCIENCE ASSESSMENT

Stanley L. Helgeson and David D. Kumar

The National Center for Science Teaching and Learning
The Ohio State University
104 Research Center
1314 Kinnear Road
Columbus, OH 43212.

Introduction

This paper will review the role of educational technology in assessment in science education. Documents were initially identified for this review by conducting a search of the ERIC data base. Next, documents were identified from known sources. Each document thus identified was then subjected to a systematic review. Those articles dealing with computer and hypermedia applications to assessment in science education were selected for inclusion.

Educational technology has been a focus of development and research in science teaching and learning (Grandgenett et al., 1992; Kumar, 1991a). Similarly, identifying or devising alternative strategies for assessment has recently been an emphasis in educational testing and development (Stiggins and Bridgeford, 1985; Shavelson, Carey and Webb, 1990; Swain, 1991; King and Brathwaite, 1991). It has been suggested that perhaps the best way to test students' understanding is to put them in a laboratory, pose a problem, and let them solve the problem using the resources of the laboratory. However, large-scale, hands-on testing in laboratories is too expensive in time, human resources, and equipment. For this reason researchers need to build a new knowledge base and a new technology base for achievement testing in science (Shavelson, Carey and Webb, 1990).

Educational technologies such as computers and hypermedia are in the forefront of this endeavor, offering "the closest approximation to hands-on performance evaluation that can be group administered" (Shavelson, Baxter, Pine, Yure, Goldman and Smith, 1990, p. 5). For example, computers and hypermedia applications can provide multi-dimensional environments to study the process of learning and problem solving, and to represent knowledge structures (Jonassen, 1988; Champagne and Klopfer, 1984; Bower and Hilgard, 1981). Thus, computers and hypermedia not only find applications in the development of alternative

Paper presented at the annual meeting of the National Science Teachers Association, Kansas City, KS, April 1-4, 1993

assessment technologies but also provide environments for understanding the processes involved in assessment in science education.

Computer Applications in Assessment

In summarizing research findings on computer-based education, Waugh and Currier (1986) found that: (1) groups experiencing some kind of computer-based education attained test scores which were on average between .25 and .44 standard deviations higher than their comparison groups; (2) there was evidence favoring the use of computer-based education with academically disadvantaged students; (3) long term retention was no better for computer-based education than for other modes of instruction; (4) secondary students who experienced computer-based education had more positive attitudes toward computers than did their peers who did not experience computer-based education; and, (5) there was significantly less time required for computer-based education compared to conventional instruction. It should be noted that many of the studies summarized relied heavily on drill and practice modes of instruction. Such programs depend upon immediate feedback as a major function. While this may not fit the common perception of assessment, it is clear that it does in fact function in such a manner and that the immediate feedback may well have a positive impact on learning.

A common use of computers in assessment is to provide teachers with access to large banks of test items. These may range from specific topics such as medical biochemistry (Aesche and Parslow, 1988) for instructors of a given course, to a test bank designed for state assessment (Willis, 1988), to a broad range of juried test items which teachers anywhere in the country may access and download onto their own computers (Dawson, 1987). Once the item banks are in place, the computer may then be used to devise unique combinations of test items for each student and to use the results of those tests to develop remedial learning activities for each student. In each case, the computer can administer the quizzes, grade and record the results, and provide the student with immediate feedback (Dunkleberger, 1980). Use of the computer to file test questions, assemble examinations, handle all records, produce and grade tests, and guide students to what should be done next enables testing to be done with an efficiency not possible from any teacher (Summers, 1984; Vogel, 1985; Heikkinen and Dunkleberger, 1985).

One type of formative assessment might be to use the computer to evaluate student data collected in laboratory exercises. Such checking of data and calculations is repetitive, prone to error, and not cost effective when done by humans. Computers, on the other hand, excel at this type of task (Harrison and Pitre, 1983, 1988). Programs used in this way are designed to check for realistic values, a range of data, and values clearly outside acceptable limits. When incorrect answers are given, students may be asked to redo their calculations and submit revised figures (May, Murray and Williams, 1985). The programs also may be designed to tentatively accept answers within a certain range, but to suggest that students return to places of potential error and check their work (Harrison and Pitre, 1988).

As part of a project to integrate computer-generated homework into physical science college courses, Milkent and Roth (1989) used computer-generated problems as homework assignments and monitored student progress with computer-generated multiple choice quizzes. They found that the use of the computer-generated homework significantly reduced the effectiveness of ACT scores as predictors of course achievement. Put in other words, as a result of the homework approach, students had greater opportunities for achieving mastery and for minimizing the potential influence of entry level aptitude and prior academic preparation.

Incorporation of computers into science instruction often takes the form of microcomputer-based laboratories (MBL) with assessment as a part of the system. However, in some cases this means simply presenting multiple choice questions by means of the computer screen (Bross, 1986). If immediate feedback is not available, no learning gains may accrue to such computer use. Easier data collection and processing may still make this approach to testing of value to the instructor. A more useful approach might be that described by Browning and Lehman (1988) for identifying student misconceptions in genetics problem solving. Four computer programs were presented and the students' responses were recorded and analyzed for evidence of misconceptions and difficulties in the problem solving process. Three main problem areas were identified: difficulties with computational skills, difficulties in the determination of gametes, and inappropriate application of previous learning to new problems. Evaluation of this type would seem to show considerable promise for remedial instruction and improved student learning.

Collins (1984) conducted a study to determine whether learning would be improved with computerized tests. The students (n=210) were enrolled in a one-semester introductory biology course. Students in the computer section took computer generated tests in addition to the tests taken by students in the other sections. Students taking the computer tests were given immediate feedback on their scores, then told which responses were correct and which were incorrect. In addition, the computer recorded student data on disk, allowing for later analysis by the instructor. Collins concluded that computer testing led to enhanced learning as indicated by higher scores on weekly in-class written tests, the midterm examination, the final examination, and final class marks.

Collins and Earle (1989-90) examined the effects of computer-based learning and computer-administered testing in an introductory biology class. They found that the greatest benefit was attained by those using the computer units in addition to attending regular lectures. Taking weekly computer-administered multiple choice tests also appeared to benefit students of middle and upper ability but not students of lower ability levels. Although students benefitted from using either the computer learning units or the computer tests, the use of the two together did not result in even greater gain, as might have been expected. Frequency of use of the units appeared to be a factor in that the "frequent" user group achieved a much higher mean score and higher pass rate than did the "infrequent" user group.

The effects of microcomputer-administered diagnostic testing on both student achievement and attitudes were of concern to Waugh (1985). Students in one group were given the unit objectives and responded to a computer-administered diagnostic test consisting of one item per objective. The other group received the objectives and were assigned an out-of-class task of completing an objective specific mini-project. The results showed that microcomputer-administered diagnostic testing could positively influence the immediate achievement of students in science. Evidence did not, however, support the hypothesis that an exposure to diagnostic testing might influence continuing achievement. The findings indicated that the use of microcomputer-administered diagnostic testing was successful in increasing student achievement in science by an average of six percent with no loss of positive attitude toward school, learning, or science. The evidence further indicated that diagnostic testing might have played a role in arousing student interest in microcomputers.

The possibility that students were being disadvantaged by taking computer tests instead of written paper forms of the same tests was studied by Fletcher and Collins (1986-87). They found that students' mean scores on the computer-administered test and the written forms of the same test were roughly equivalent, and concluded that the students were not disadvantaged by taking the computer tests. The students indicated that most of them favored the computer-administered tests and cited several major advantages: (1) immediacy of scoring; (2) immediate feedback on incorrect answers; (3) more convenient, straight forward and easy-to-use; and (4) faster than written tests. Two major disadvantages were noted by the students: (1) not being able to review all their responses at the end of the test and make changes; and, (2) not being able to skip questions and come back to answer them later (p. 42).

The opposite case was studied by Jackson (1988) who attempted to discover whether a computer could give any significant educational advantage to the pupil. That is, could the computer improve pupil motivation during the test, by giving instant feedback and marking, thus improving understanding and hence produce an enhanced score in a future test? The middle school science students who were tested by computer and given immediate feedback scored significantly higher on a later test using the same material than did those students who were tested using the traditional paper and pencil method.

Computerized adaptive testing is emerging as a more efficient way to assess student knowledge. A unique characteristic of computer adaptive testing is that each examinee is given an individualized test comprised of questions from a content valid item bank. The computer program is designed to select questions that provide the most information about the examinee based on his/her current estimated ability measure. After answering each question the examinee's ability is re-estimated. If the question is answered correctly, the measure increases and the next question administered is more difficult. If the question is answered incorrectly, the examinee's ability measure decreases and the next question presented is easier. Thus, a tailored test is designed for each individual. A pilot study of the effectiveness of computerized testing for certification in five medical technology fields revealed that 50 to 100 questions served to provide the necessary pass/fail information compared to 109 written questions. The computerized test took two to two and a half hours to complete compared to four hours required for the written test. Other advantages of computerized adaptive testing included shorter turn-around time of test results, improved security and data collection, and less chance of cheating due to individualized tests (Herb, 1992).

Hypermedia in Assessment

The impact of emerging interactive videodisc technology was studied by Huang and Aloi (1991) in a first year biology course. The interactive video involved 17 menu driven chapters integrating computer text with laser disc images and computer graphics. The researchers compared the proportion of students getting A, B, C, D, F, and W (withdraw) for 11 semesters prior to using interactive video with the proportions during the 5 semesters following its use. They found that the proportion receiving A's increased significantly ($p<.005$) following use of the interactive video. The percentage changes were: A's, 6% before and 18% after; B's, 21% before and 32% after; C's, 20% before and 36% after; D's 10% before and 4% after; F's did not change. Retention of students was also increased. The proportion of W's was 33% before interactive video use and 24% after. Thus, the use of interactive videodisc resulted in increased proportions of success at nearly all levels of achievement.

Interactive videodisc was also used as a tool in assessing science teachers' knowledge of safety regulations in school laboratories for purposes of teacher certification by the Connecticut State Department of Education (Lomask, Jacobson and Hafner, 1992). The program simulates a typical lab activity in a secondary school general science course and shows four student performing a simple lab experiment to identify unknown materials. The IVD assessment includes two stages: stage one deals with safety equipment and storage of chemicals and stage two deals with students' laboratory practices. The examinees are asked to assume the role of the lab teacher by viewing an interactive videodisc simulated classroom. The teachers are then asked to identify safety violations and to suggest preventive or corrective measures. Subjects' responses are recorded for later analysis ana scoring (p. 1).

An emerging application of the hypermedia in assessment involving problem-based learning in chemistry is found in the "Hyperequation" (Kumar, 1991b) project at the National

Center for Science Teaching and Learning at The Ohio State University (Cognosos, 1991). Hyperequation is an assessment software developed in the HyperCard (TM) on a Macintosh platform to study student performance in balancing stoichiometric chemical equations. Studies using the Hyperequation are underway in a high school in Columbus, Ohio.

Hyperequation (in its pilot stage) has the following features. It is easy to operate through the computer-mouse interface. It has been programmed to provide immediate feedback and motivation, and to register some pertinent information involved in the process of balancing stoichiometric equations. One of the purposes of this software is to simulate similar tasks involving traditional paper-pencil methods of assessment, in addition to providing a non-linear visual environment for problem solving. For example, Hyperequation records the number of attempts and the order in which responses are made by each student, including the total time on-task. Also the Hyperequation can display on the screen as well as provide a print-out of an overall and item by item record of each student's performance on the problem task. Due to confidentiality of student performance records, only the classroom teacher, through a password, has access to this information in the Hyperequation.

Prima facie evidence from pilot data reveals that Hyperequation provided all of the information mentioned above pertaining to the problem space in solving stoichiometric chemical equations. Both expert and novice high school chemistry students who participated in the study made significantly different scores using the Hyperequation than at the same task using traditional paper-pencil methods. In addition, the Hyperequation was found to differentiate between hard and easy problems previously established by paper-pencil methods.

Martinez (1991) reported a similar hypermedia environment using an "IBM-compatible computer interface delivery" platform for administering "figural response" test items to cell and molecular biology students. With a computer-mouse interface, a set of computer screen tools are activated by buttons (e.g., "move object," "rotate," "draw line"). For example, chromosomes and molecular groups are moved on the screen by students to respond to various questions such as "Given the D-glucose below, construct its L-glucose stereoisomer using the template shown" (p. 387).

A similar work in physics was reported by Shavelson, Baxter, Pine, Yure, Goldman and Smith (1990). For example, using a simulation "Electric Mysteries" on a Macintosh platform, a hands-on environment for assessment in electric circuits was replicated. Students have to find out the circuitry among five possible circuit designs from five "mystery boxes" by manipulating icons on the Macintosh computer, instead of physically manipulating bulbs, batteries and wires. Every move made by the student is recorded by the computer which is later used for assessment. The findings indicate that expert students performed significantly better on the electric mysteries problem than novices.

Summary

There appear to be several advantages to incorporating some form of computer assistance in assessment. Immediate feedback to the students seems to be a consistent factor in increased achievement. Ease of test taking, together with improved record keeping, suggest improved efficiency for both students and teachers. The availability of large test item banks makes possible several intermediate quizzes with achievement gains appearing to result from this practice. Another type of formative assessment is made possible through the use of computers to monitor homework and laboratory activities. Such formative evaluation serves both as a diagnostic tool and as a remediation device, indicating where corrections are needed. The data collection capability of computer testing also permits more extensive data analysis, especially in the area of test item analysis, which in turn should yield more reliable and, presumably, more valid assessment. Two cautions must be noted, however. First, the simplicity of devising multiple choice, true/false, matching, and other objective tests can lull the teacher into simply doing a better job of assessing low level recall knowledge. Second, the linear nature of most computer testing does not allow the student to go back and reflect upon a particular item, nor to view the completed test as a whole to check for consistency of responses. The increased improvement and implementation of such emerging technologies as interactive video and hypermedia (Kumar, 1991b) show high promise for overcoming both difficulties by providing opportunities for both improved levels of questions and increased flexibility in the testing process.

While the research evidence is limited, it appears that some tentative conclusions may be drawn. The first, and possibly most important, finding is the positive effect on achievement of immediate feedback and its attendant reinforcement. A second outcome is the increased ease and simplicity of test-taking and data collection and analysis. Next, there is an increased facility to do formative or intermediate assessment with accompanying remediation. Finally, with the emergence of hypermedia, there is increased flexibility of assessment allowing for a potentially better match between the way in which humans construct knowledge and methods for assessing such learning.

Acknowledgement

This research was supported in part by the National Center for Science Teaching and Learning under OERI Grant No. R117Q00062, U.S. Department of Education. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the sponsoring agency.

References

Aesche, D. W. and Parslow, G. R. (1988). Use of questions from the medical biochemistry question bank with the 'Q' instruction package. Biochemical Education, 16(1): 24-28.

Bross, T. R. (1986). The microcomputer-based science laboratory. Journal of Computers in Mathematics and Science Teaching, 5(3): 16-18.

Browning, E. and Lehman, J. D. (1988). Identification of student misconceptions in genetics problem solving via computer program. Journal of Research in Science Teaching, 25(9): 747-761..

Bower, G. H. and Hilgard, E. R. (1981). Theories of learning. Englewwod Cliffs, NJ: Prentice-Hall.

Champagne, A. B. and Klopfer, L. E. (1984). Research in science education: Cognitive psychology perspective. In Holdzkom, D. and Lutz, P. B. (eds.), Research within reach: Science Education. Washington, DC: National Science Teachers Association.

Cognosos. (1991), 1(1): 5.

Collins, M. A. J. (1984). Improved learning with computerized tests. The American Biology Teacher, 46(3): 188-190.

Collins, M. A. J. and Earle, P. (1989-90). A comparison of the effects of computer-administered testing, computer-based learning and lecture approaches on learning in an introductory biology course. Journal of Computers in Mathematics and Science Teaching, 9(2): 77-84.

Dawson, G. (1987). 10,000 science questions by modem. The Science Teacher, 54(3): 41-44.

Dunkleberger, G. E. (1980). Applying 20/20 hindsight to self-pacing. The Science Teacher. 47(8): 32-34.

Fletcher, P. and Collins, M. A. J. (1986-87). "Computer-Administered Versus Written Tests--Advantages and Disadvantages." Journal of Computers in Mathematics and Science Teaching, 6(2): 38-43.

Grandgenettn N., Ziebarth, R., Koneck, J., Farnham, M. L., McQuillan, J. and Larson, B. (1992). An investigation of the anticipated use of multimedia by pre-sevice teachers. Journal of Educational Multimedia and Hypermedia, 1: 91-101.

Harrison, D. and Pitre, J. M. (1983). Computerized lab test on error analysis. The Physics Teacher, 21(9): 588-589, December.

Harrison, D. and Pitre, J. M. (1988). Computerized pre-checking of laboratory data. The Physics Teacher, 26(3): 156.

Heikkinen, H. and Dunkleberger, G. E. (1985). On disk with mastery learning. The Science Teacher, 52(7): 26-28.

Herb, A. (1992). Computer adaptive testing-. ADVANCE, (Newsletter of the Medical Laboratory Professionals). February, 17, 1992.

Huang, S. D. and Aloi, J. (1991). The impact of using interactive video in teaching general biology. American Biology Teacher, 53(5): 281-284.

Jackson, B. (1988). A comparison between computer-based and traditional assessment tests, and their effects on pupil learning and scoring. School Science Review, 69(249): 809-815.

Jonassen, D. H. (1988). Designing structured hypertext and structuring access to hypertext. Educational Technology, 28(11): 13-16.

King, W. and Bathwaite, W. (1991). School-based assessment in science - a Caribbean perspective. School Science Review, 72(261): 127-131.

Knight, C. W. and Dunkleberger, G. E. (1977). The influence of computer-managed self-paced instruction on science attitudes of students. Journal of Research in Science Teaching, 14(6): 551-555.

Kumar, D. D. (1991a). Hypermedia: A tool for STS education? Bulletin of Science, Technology & Society, 11(6): 331-332.

Kumar, D. D. (1991b). A Note on "Hyperequation." An occasional paper. National Center for Science Teaching and Learning, The Ohio State University, Columbus, September.

Lomask, M., Jacobson, L. S. and Hafner, L. P. (1992). Interactive Videodisc as a Tool for Assessing Science Teachers' Knowledge of Safety Regulations in School Laboratories. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Boston, MA, March 21-25.

Martinez, M. E. (1991). Figural response in science and technology testing. In Kulm, G. and Malcom, S. M. (eds.), Science assessment in the service of reform. Washington, DC: AAAS.

May, P. M., Murray, K. and Williams, D. R. (1985). CAPE: A computer program to assist with practical assessment. Journal of Chemical Education, 62(4): 310-312.

Mihkelson, A. et al. (1984). "The Application of computer technology in the assessment of large undergraduate classes. Journal of Chemical Education, 61(7): 631.

Milkent, M. M. and Roth, W. M. (1989). Enhancing student achievement through computer-generated homework. Journal of Research in Science Teaching, 26(7): 567-573.

Shavelson, R. J., Carey, N. B. and Webb, N. M. (1990). Indicators of science achievement: Options for a powerful policy instrument. Phi Delta Kappan, pp. 692-697, May.

Shavelson, R. J., Baxter, G. P., Pine, J., Yure, J., Goldman, S. R. and Smith, R. (1990). Performance indicators for large-sclae science assessment. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA, April.

Stiggins, R. J. and Bridgeford, N. J. (1985). The ecology of classroom assessment. Journal of Educational Measurement, 22(4): 271-286.

Summers, G. (1984). Testor. Journal of College Science Teaching, 13(5): 356-358.

Swain, J. R. L. (1991). The nature and assessment of scientific explorations in the classroom. School Science Review, 72(260): 65-77.

Vogel, G. (1985). A microcomputer-based system for filing test questions and assembling examinations. Journal of Chemical Education, 62(11): 1024-1027.

Waugh, M. L. (1985). Effects of microcomputer-administered diagnostic testing on immediate and continuing science achievement and attitudes. Journal of Research in Science Teaching, 22(9): 793-805.

Waugh, M. L. and Currier, D. (1986). Computer-based education: What we need to know. Journal of Computers in Mathematics and Science Teaching, 5(3): 13-15.

Willis, J. A. (1988). Learning outcome testing: A statewide approach to accountability. Technological Horizons in Education, 16(1): 67-93.