Feasibility of Machine Scoring of Concept Maps

CSE Technical Report 460

Harold F. O'Neil, Jr.
University of Southern California/CRESST

Davina C. D. Klein
CRESST/University of California, Los Angeles

December 1997

Center for the Study of Evaluation
National Center for Research on Evaluation,
Standards, and Student Testing
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported under the Educational Research and Development Center Program, PR/Award Number R305B60002, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

FEASIBILITY OF MACHINE SCORING OF CONCEPT MAPS

Harold F. O'Neil, Jr.
University of Southern California/CRESST

Davina C. D. Klein
CRESST/University of California, Los Angeles

The purpose of this letter report is to document our progress on the feasibility of scoring concept maps using technology. Technology can clearly play a role in the assessment process, including administration of the assessment, scoring, and reporting of results, thus providing consistently high quality assessments, possibly at a reduced cost. Our approach to using technology for assessment is called "An Integrated Simulation Approach to Assessment." CRESST has assembled a suite of performance assessment tasks (our integrated simulation) onto which have we mapped the types of learning expected of students.

The design of this integrated simulation performance assessment has the following characteristics: (a) relevant, project-based scenarios that include meaningful, real-world tasks; (b) individual and team processes and products; and (c) a technology base using Web-based networked systems. The integrated simulation we have developed includes both collaborative and individual concept mapping tasks, a problem-solving search task, and a questionnaire to measure metacognition and motivation (self-efficacy and effort).

Additional documentation on an integrative approach to assessment is provided by a series of reports on the Department of Defense's Computer Aided Education & Training Initiative (CAETI) (see Baker & O'Neil, 1996; Herl, Baker, & Niemi, 1996; Herl et al., 1996; Herl & O'Neil, 1996; Klein, O'Neil, & Baker, 1996) and papers presented at the 1997 annual meeting of the American Educational Research Association (Chung, O'Neil, Herl, & Dennis, 1997; Herl, O'Neil, Chung, & Dennis, 1997; Klein, O'Neil, Dennis, & Baker, 1997; Schacter et al., 1997).

One element of our integrated simulation is an online concept mapping construction and scoring system. This letter report will focus on the use of technology to score concept maps. A concept map is a graphical representation of information consisting of nodes and labeled lines; nodes correspond to concepts within a particular subject area or domain, lines indicate a relationships between pairs of concepts (or nodes), and labels on each line explain how two concepts are related (refer to Jonassen, 1996, and Jonassen, Beissner, and Yacci, 1993, for more in-depth coverage of concept mapping). We use concept maps to measure content understanding. The potential advantages and disadvantages of using concept maps versus multiple-choice tests are specified in Table 1. Our assumptions in Table 1 are (a) availability of task analysis, experts, students for formative evaluation; and (b) if existing Intranet or Internet capability is in place, cost is very low; if networks not in place, cost is high. In general, concept maps may provide most of the advantages of both multiple-choice and performance testing with few of the disadvantages. Because the online concept mapping construction and scoring system is very new, little empirical research has been conducted on the issues in Table 1. Thus, the comparisons reflect our judgment and not the results of empirical studies.

Table 1

Type of Testing Comparisons (Ideal)

	Multiple choice	Performance assessment	Concept maps
Time to develop "items"	High	Medium	Low
Time to develop "rubrics"	Low	High	Low
Time to score	Low	High	Low
Logistics to administer	Low	High	Medium
Reliability	High	Medium	High
Validity	High	High	High
Credibility (parents)	High	Low	Unknown
Fairness	Medium	Medium	Unknown
Deep understanding	Low	High	High
Problem solving	Low	High	High
Work in teams	Low	Medium	High
Knowledge representation	Low	Low	High
Cost	Low	High	Varies[b]
Language dependent	Medium	High	Low

We have designed and developed concept mapping software (a) to permit students to both individually and collaboratively construct concept maps on the computer, and (b) to provide real-time scoring and feedback to students based upon an expert's map.

An expert criterion concept map is used to score students' concept maps in real time. Preliminary results underscore the ease with which students were able to learn the computerized concept mapping tool and the simplification of concept map scoring. Further research on scoring, reliability, and validity issues regarding the concept mapping assessment are ongoing at CRESST. Frequently-asked questions and answers regarding concept maps are presented in Table 2.

In our research, we also examined the viability of using collaborative concept mapping in a networked computer environment as an assessment tool. A particularly novel feature of our work is that we are refining an approach that employs networked computers to capture, measure, and report—in real time—team processes for individual students. The team processes are (a) adaptability—

Table 2

Frequently-Asked Questions

Why replace our multiple-choice tests with concept maps?

To allow better measurement of what is important—content understanding and problem solving.

Is it hard to teach people to make computerized concept maps?

Our experience with high school students is that it takes about 10 minutes.

How long does it take the students?

About 30 minutes.

Are concept maps related to other types of assessments?

On same topic, concept maps and essays correlate about .7, which indicates reasonable evidence that they are measuring similar things.

What is the process for getting the expert map?

Identify the expert(s); interview (1 hour) to generate key concepts based on existing job/task analysis; we provide links; teach expert(s) how to make concept map (10 minutes); have expert(s) construct the concept map(s) (30-40 minutes); we digitize the map(s).

recognizing problems and responding appropriately; (b) communication—the exchange of clear and accurate information; (c) coordination—organizing team activities to complete a task on time; (d) decision making—using available information to make decisions; (e) interpersonal—interacting cooperatively with other team members; and (f) leadership—providing structure and direction for the team.

We have conducted one study with the collaborative concept mapping tool and have found it to be feasible. The current work, coupled with our past efforts, suggests that our computer-based assessment approach is feasible, and will be reliable and valid (e.g., O'Neil, Chung, & Brown, 1997; Chung et al., 1997).

A Feasible Solution for Schools: Paper-and-Pencil Administration With Machine Scoring

Clearly, computer administration of the kind described above is not feasible in schools without extensive computer resources. However, computerized scoring of paper-and-pencil administered concept maps should be a reasonable solution. By using some form of paper-and-pencil approach in which students construct their concept maps, and by then using a computer to score the maps against an expert's map, we can capitalize on both the innovative approach of using concept mapping as an assessment tool and the cost-effectiveness associated with computer scoring of multiple-choice tests.

We have explored several technologies for this purpose involving either scanning technology or voice recognition. With respect to a scanning technology, we have explored a relationship with National Computer System (NCS), a leading test form designer and the largest scorer of multiple-choice forms (e.g., the Iowa Tests of Basic Skills [ITBS] multiple-choice tests) in the industry. One approach is for us to co-design the preprinted concept map form, with CRESST providing the scoring software. The student would "draw" his or her concept map on the preprinted concept map form. The form would be mailed to NCS, scanned, and a digital file would be created and scored using CRESST software. Results would be then distributed to the school system.

With NCS we have done a preliminary analysis to estimate the costs of a computer-scored, paper-and-pencil-administered concept map (Figure 1). These figures represent direct costs, without university overhead. The student estimates are CRESST planning figures for a potential implementation in the Los

NCS Scannable Concept Maps

Estimate based on assumption of *fully operational,* districtwide implementation.

Students per grade level (estimates)

Elementary	50,000
Middle school	45,000
High school	45,000

Expected number of versions and forms

4	Four content areas possible: history, language arts, math, science
3	Each content area will test at three grade levels
2	In addition, elementary school versions will be in both English and Spanish
1	Middle and high school versions will be English-only
1	At operation, expecting only one concept map per student (per content area)
1	At operation, expecting all students within grade/content to take same task (plus year-round version for separate administration?)

Total number of distinct versions/grade level = #content areas x #languages x #distinct maps/student x #different administrations

Elementary school versions	8	(however, each student takes only 4)
Middle school versions	4	
High school versions	4	
Total number of versions	16	
Total number of forms	560,000	(on average; 40,000-50,000 forms per version needed)

NCS costs

Printing		
Design/typeset	$1,200	for first 6-bubble layout
	$45	for each new version (text change only)
	$1,200	for first 10-bubble layout
	$45	for each new version (text change only)
Printing	$4,000	per 50,000 of same form
Total printing	$67,030	
Software development		
First form	$25,000	
Each additional form	$900	
Total software	$38,500	
Processing and delivery of scan file to CRESST		
	$0.25	per document
Total processing	$140,000	

Figure 1. Costs estimates of a computer-scored, paper-and-pencil-administered concept map using preprinted forms.

Project administration	
For one year	$15,000
Total administration	$15,000
Shipping	
To CRESST	$250
Total shipping	$250
CRESST costs after NCS processing	
Printing	$100
Scoring software	$8,000
Project administration	$10,000
Total CRESST	$17,900
Total costs for NCS venture	
Total	$278,680
Cost per form	$0.50

Figure 1. (continued)

Angeles Unified School District (LAUSD) system. The purpose of such detail is to provide a baseline or notional context to allow estimates of costs and rough comparisons of other technologies to achieve the same goal. As may be seen from the bottom line (the last line in the figure), the cost estimate using preprinted forms to be scored by NCS is $0.50 per concept map or assessment form.

An alternative to optical scanning of student maps that we have explored is machine entry of concept maps via voice input. A digital file is then created, which can be scored by CRESST software. In this approach, students would create their concept maps using paper and pencil (but not using preprinted scanning forms); then the maps would be "read" into a computer file by data entry personnel using voice commands and off-the-shelf discrete speech understanding software. Figure 2 shows an analysis of costs using voice input. The cost categories do not include costs of computer data entry stations or licensing of the voice entry software. Further, university overhead is not included. The cost estimate indicates $0.25 per form.

Voice Scannable Concept Maps

Estimate based on assumption of *fully operational*, districtwide implementation.

Students per grade level (estimates)

Elementary	50,000
Middle school	45,000
High school	45,000

Expected number of versions and forms

4	Four content areas possible: history, language arts, math, science
3	Each content area will test at three grade levels
2	In addition, elementary school versions will be in both English and Spanish
1	Middle and high school versions will be English-only
1	At operation, expecting only one concept map per student (per content area)
1	At operation, expecting all students within grade/content to take same task (plus year-round version for separate administration?)

Total number of distinct versions/grade level = #content areas x #languages x #distinct maps/student x #different administrations

Elementary school versions	8	(however, each student takes only 4)
Middle school versions	4	
High school versions	4	
Total number of versions	16	

Total number of forms	560,000	(on average; 40,000-50,000 forms per version needed)

In-house costs

Printing

Copying/form	$0.03
Total printing	$16,800

Software development

Scoring software	$8,000
Voice additions	$10,000
Total software	$18,000

Processing and delivery of scan file

Voice entry	60	seconds/form		
Time needed	9333	hours	# weeks	233
			with staffers	30
			total weeks	8
Staff pay	$10.00	per hour		
Work station costs	TBD			
Total processing	$93,999			

Figure 2. Costs of a computer-scored, paper-and-pencil-administered concept map using voice input.

Project administration	
Administration	$10,000
Report printing	$100
Total administration	$10,1000
Total costs for CRESST-only venture	
Total	$138,233
Cost per form	$0.25

Figure 2. (continued)

Figure 3 shows a cost comparison of these two technologies. As previously mentioned, the student data are based on a scenario regarding a possible implementation in the Los Angeles Unified School Districts (LAUSD). In the analysis in Figure 3, voice recognition appears to be more cost effective than the CRESST/NCS forms approach. However, these figures are a first cut of direct costs (e.g., no overhead has been added), and more extensive analysis is needed. We need to validate the assumptions underlying these analyses and also conduct a sensitivity study of the cost drivers.

Although these estimates are still quite rough, it is clear that use of this technology makes paper-and-pencil concept mapping administration with machine scoring a feasible assessment solution, especially when compared with more expensive forms of performance assessment (e.g., scoring a written essay costs about $5.00 per student using the commercial services of the Iowa Tests of Basic Skills, and scoring a hands-on performance measure in science can cost $90.00/student/test).

Final Issues

Student testing is necessary to ensure that such concept mapping approaches are reliable and valid. Further, more work needs to be done to enhance our reporting techniques once scoring is complete. Nonetheless, the approaches outlined in this report capture the positive features of performance assessment, while incorporating the cost-effective scoring approaches of multiple-choice testing. They should thus be seriously considered as viable and effective strategies in the assessment of students' knowledge.

Variables	NCS form approach	Voice recognition approach
Cost	Current estimate: $0.50 per concept map	Current estimate: $0.25 per concept map
Turn-around time	Dependent upon NCS (and in-house CRESST scoring and reporting)	Dependent upon number of data entry people (and in-house CRESST scoring and reporting)
Piloting concerns	Forms have not yet been piloted with students; special forms are required for any piloting	Addition of letters/numbers has not yet been piloted with students but can be accomplished with relative ease
Fidelity of data	Dependent upon optical character recognition and correct "bubbling"; greater fidelity means more human intervention (= higher cost)	Dependent upon data entry personnel who can be well trained to check for voice recognition errors; software "trainable" to each individual data enterer's voice
Critical cost variables	Processing and delivery of scan file *and* printing of forms—half of total estimate goes to NCS's processing and delivery of scan file (price is on a per-concept-map basis); in addition, 25% of total estimate goes to printing of forms (price is on a per-form-type basis)	Processing and delivery of scan file—60% of total estimate goes to data entry personnel (price is based on time needed to enter each concept map)
Flexibility of maps	Low flexibility: standard form necessary, creation of map difficult, letters and numbers *only*, bidirectionality not possible	High flexibility: student can dictate form and content of map, letters and numbers required in addition to terms/link labels, bidirectionality supported

Figure 3. Comparison of preprinted form and voice-scannable approaches to concept map scoring.

References

Baker, E. L., & O'Neil, H. F., Jr. (1996). *Design report for CAETI assessment plans (Year 1)* (Report to ISX/DODEA). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Chung, G. K. W. K., O'Neil, H. F., Jr., Herl, H. E., & Dennis, R. A. (1997, March). Use of networked collaborative concept mapping to measure team processes and team outcomes. In H. F. O'Neil, Jr. (Chair), *An integrated simulation approach to assessment*. Symposium presented at the annual meeting of the American Educational Research Association, Chicago.

Herl, H. E., Baker, E. L., & Niemi, D. (1996). Construct validation of an approach to modeling cognitive structure of U.S. history knowledge. *Journal of Educational Research, 89*, 206-218.

Herl, H. E., & O'Neil, H. F., Jr. (1996). *Final report. CAETI STS1 technologies baseline data collection (Year 1)* (Report to ISX/DODEA). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Herl, H. E., O'Neil, H. F., Jr., Chung, G. K. W. K., & Dennis, R. A. (1997, March). Feasibility of an on-line concept mapping construction and scoring system. In H. F. O'Neil, Jr. (Chair), *An integrated simulation approach to assessment*. Symposium presented at the annual meeting of the American Educational Research Association, Chicago.

Herl, H. E., O'Neil, H. F., Jr., Chung, G. K. W. K., Dennis, R. A., Klein, D. C. D., Schacter, J., & Baker, E. L. (1996). *Measurement of learning across five areas of cognitive competency: Design of an integrated simulation approach to measurement. Year 1 report* (Report to ISX/DODEA). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Jonassen, D. H. (1996). *Computers in the classroom*. Englewood Cliffs, NJ: Prentice Hall.

Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Klein, D. C. D., O'Neil, H. F., Jr., & Baker, E. L. (1996). A *cognitive demands analysis of innovative technologies* (Report to ISX/DODEA). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Klein, D. C. D., O'Neil, H. F., Jr., Dennis, R. A., & Baker, E. L. (1997, March). The five families of cognitive learning: A context in which to conduct cognitive demands analyses of innovative technologies. In H. F. O'Neil, Jr. (Chair), *An integrated simulation approach to assessment*. Symposium presented at the annual meeting of the American Educational Research Association, Chicago.

O'Neil, H. F., Jr., Chung, G. K. W. K., & Brown, R. (1997). Use of networked simulations as a context to measure team competencies. In H. F. O'Neil, Jr. (Ed.), *Workforce readiness: Competencies and assessment* (pp. 411-452). Mahwah, NJ: Lawrence Erlbaum Associates.

Schacter, J., Herl, H. E., Chung, G. K. W. K., O'Neil, H. F., Jr., Dennis, R. A., & Lee, J. L. (1997, March). Feasibility of a Web-based assessment of problem solving. In H. F. O'Neil, Jr. (Chair), *An integrated simulation approach to assessment*. Symposium presented at the annual meeting of the American Educational Research Association, Chicago.