Performance Assessment Models and Tools for Complex Tasks

CSE Technical Report 682

Gregory K. W. K. Chung, Girlie C. Delacruz,
and William L. Bewley
CRESST/University of California, Los Angeles

June 2006

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
GSE&IS Bldg., Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported under the National Research and Development Centers, PR/Award Number R305B960002, as administered by the Institute of Education Sciences (IES), U.S. Department of Education, and under Office of Naval Research Award Number #N00014-02-1-0179, as administered by the Office of Naval Research.

The findings and opinions expressed in this report are those of the author(s) and do not necessarily reflect the positions or policies of the National Research and Development Centers, the Institute of Education Sciences (IES), the U.S. Department of Education, or the Office of Naval Research.

Acknowledgements

This report is re-published as a CRESST report with permission from the International Test and Evaluation Association (ITEA) Journal. Our very deep thanks to Beverly Schaeffer, ITEA Editor in Chief for granting permission and for her assistance. The original article was published in the ITEA Journal, March/April 2004, 25(1), 47-52. Additional information about ITEA may be found at http://www.itea.org/.

PERFORMANCE ASSESSMENT MODELS AND TOOLS FOR COMPLEX TASKS

Gregory K. W. K. Chung, Girlie C. Delacruz, and William L. Bewley, CRESST/University of California, Los Angeles

Abstract

Assessment models support the design of quality performance assessments. Assessment tools are being developed to enable easy and effective application of the models. Based on representations of assessment design knowledge and domain knowledge in ontologies, the tools provide guidance to assessment designers, and through constraint processing check the completeness and accuracy of designs. With the addition of Bayesian networks, the tools can also enable individualized instruction by identifying knowledge gaps and prescribing instruction to fill the gaps. This paper describes the technical approach to developing the tools and discusses applications of ontologies and Bayesian networks for assessment authoring and individualized instruction.

A key concern for any assessment of learner knowledge, skills, and abilities is the validity of the inferences drawn from the results with respect to the intended purposes (American Educational Research Association et al., 1999). Work at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) has focused on ensuring that measures used in assessment systems meet the technical requirements appropriate to their various purposes, and has led to the development of models for the design of assessments and tools supporting use of the models.

CRESST's assessment models have three major components: (1) identification of cognitive demands implied by performance objectives, including both cognitive strategies specific to the subject matter domain and those that are relatively domain independent; (2) analysis of the subject matter and content to be addressed, including identification of essential or desirable elements of competence, as well as elements that are common to multiple tasks; and (3) specification of behavioral demonstrations—the range of symbolic stimuli and response modes for demonstrations of levels of mastery (Baker, 1997). Models have been successfully applied to the development of assessments in domains ranging from middle school history, to reading comprehension in students Grades 2 through 9, to knowledge of fundamentals of rifle marksmanship in Marines (Delacruz, Chung & Bewley, 2003).

Studies have continued to elaborate the models, identifying attributes, representation strategies, and designs associated with different kinds of complex tasks (Baker, Sawaki, & Stoker, 2002) and the relationship of different cognitive demands to different kinds of assessment tasks (Niemi, Chung, & Bewley, 2003). These analyses outline the scope of permissible inferences one can draw from student performance on different kinds of tasks. For example, if a trainer is interested in evaluating how well a trainee understands the interrelationships among different topics in a domain, then a candidate task would be a knowledge map. If a trainer wants to evaluate whether trainees can carry out a particular procedure using a system, then a candidate task would require trainees to perform that procedure in the context of a simulation. Several different simulation contexts could be used to assess the depth and complexity of the trainees' procedural knowledge systems.

To support easy and effective application of the models, tools are being developed to represent assessment design knowledge and domain knowledge in ontologies and apply the knowledge to create and critique assessment designs. The use of ontologies to express domain knowledge for the purposes of augmenting users' knowledge has been implemented in many areas, including helping physicians diagnose patients' medical problems (e.g., Bernstam et al., 2000; de Clercq, Hasmon, Blom, & Korsten, 2001). CRESST is using ontologies in an analogous way by developing computational supports to help users implement a principled approach to assessment design and to link instruction more closely to assessment. The approach combines ontologies, constraint processing, and Bayesian networks. An assessment ontology representing the domain-independent dependencies that reflect ideal assessment properties is connected to a domain-dependent ontology representing content. Assessment designers are guided by the assessment ontology, and the content ontology supplies specific assessment content. The approach provides a constrained assessment design space enabling constraint processing to evaluate the fit between the user's design and the assessment ontology in the context of the domain ontology, alert the user to constraint violations, and show various options in the design that would satisfy all constraints. The content ontology can also be used to support individualized instruction with the addition of Bayesian networks that use information from assessments to infer gaps in the individual's knowledge as represented in the ontology, and then prescribe and deliver instructional content targeted to address specific knowledge gaps.

This paper describes the technical approach to developing the assessment tools, beginning with a definition of ontologies and an illustration of how an assessment model can be represented in an ontological framework. This is followed by an example of how an ontology could be instantiated using a problem-solving task. The paper ends with a discussion of potential applications of ontologies for assessment and instructional purposes.

Ontologies

An ontology provides an explicit representation of the knowledge in a domain as defined by an expert, usually represented as a network of concept nodes and links relating concepts. Different experts can have different views of domain knowledge, of course, and there can be alternative representations, so a particular ontology is a presumably accurate but not necessarily consensus representation, a commitment to a point of view on how a domain is structured (Chandrasekaran, Josephson, & Benjamins, 1999; McGuinness, 2002). It provides a common, explicit framework for sharing and using knowledge (Gruber, 1995), standardizing the terms and structure of the domain. This standardization makes it possible to share ontologies—and thus the knowledge they contain—for use across multiple computer platforms for different applications. They can be communicated to people and computational systems (Fensel, Hendler, Lieberman, & Wahlster, 2002), and tools such as Protégé (Gennari et al., 2002) make it possible to easily create and maintain them for use in assessment and instruction.

For the purpose of assessment design, an ontology is an explicit representation of the assessment model, providing a description of assessment parameters, the constraints governing relationships among the parameters, and computational access to the parameters and constraints. This representation can be used, for example, to provide guidance to assessment authors as they design assessments for particular purposes under particular constraints. If the ontology describes the conditions under which a simulation possessing particular characteristics is appropriate, then an assessment authoring system can, at minimum, check that the assessment designer uses the simulation task to measure the appropriate type of knowledge.

Representing Assessment Models with Ontologies

The specification of the different components of the assessment is the critical first step in developing an assessment ontology. Table 1 shows a partial listing of the assessment components required of an assessment of problem solving. Note that it represents one point of view of a problem-solving assessment model. Other perspectives will have different representations and constraints. The information in Table 1 can be mapped directly into an ontology, each major assessment component becoming a concept in the ontology, and each sub-component a property of the concept. Instances of the concept are created and assume specific values. For example, "Assessment purpose(s)" can be a concept of its own, and a specific instance of that concept is "Diagnostic." While this idea is not new (e.g., see Baker, 1997), what has changed is the availability of computational tools that make it feasible to computerize the approach (e.g., Noy et al., 2001). Further, an ontological representation in the form of a network provides the advantage of capturing the relationships among the assessment concepts.

Figure 1 shows a simplified representation of the problem-solving assessment model. Nodes represent concepts, and links represent the relationships among the different assessment concepts. Capturing the system of links is critical because the relation indicates some sort of dependency. For example, in Figure 1, two components of the problem-solving assessment ontology appear to have influential roles: assessment purpose and solution strategy. This set of relations highlights two ideas: first, the general idea that the purpose of the assessment should be explicitly linked to all aspects of the assessment, and second, in the case of the problem-solving assessment in particular, how solution strategy is constrained by other aspects of the assessment while other components are independent of each other. As will be discussed, a representation that explicitly captures these constraints can be leveraged for numerous purposes.

Table 1

Problem-solving assessment specifications

Problem-solving assessment concepts and properties	Possible values
Assessment purpose(s)	Diagnostic, readiness monitoring, certification, …
Problem scenario	
Context	Application specific
Constraints	Application specific
Situation	Application specific
Problem characteristics	Fixed, change usual sequence, improvise step(s)
Problem identification	
Explicitness	Stated, embedded, multiply masked, partial identification, barriers
Time constraints	Bounded, not bounded
Quality of information sources	Inconsistent data from multiple sources
Prior knowledge requirements	Application specific
Solution strategy	
Steps	Explicit courses of action, non-specified course of action
Grain-size	Problem subdivision
Contingency planning	Backup strategies
Help seeking	Required, not required
Cognitive strategies	Domain-independent cognitive strategies, domain-dependent cognitive strategies
Solution(s) characteristics	
Solution space	Convergent (single right answer), divergent (open-ended, with scoring criteria)
Solution correctness	Multiple acceptable solutions, partially acceptable solution
Sub-solution contingencies	Sequential, non-sequential

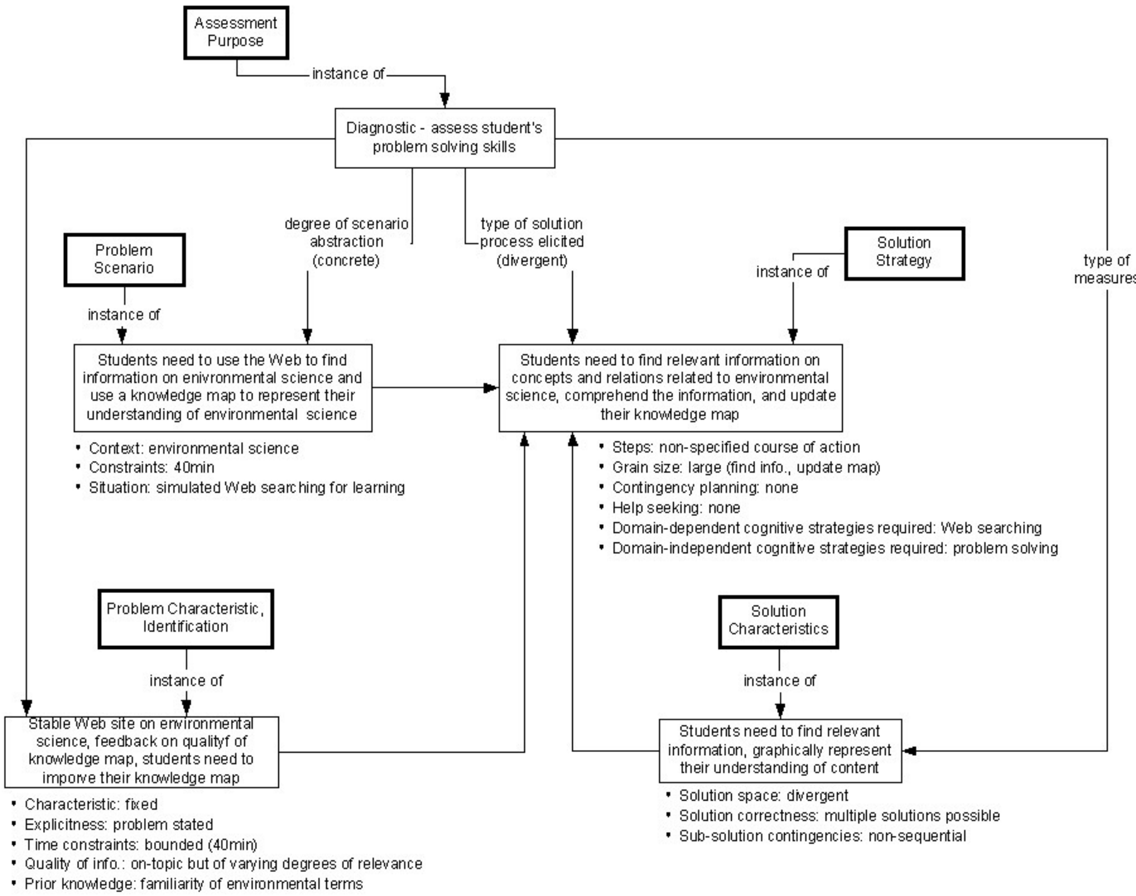

Figure 1. Example of upper level ontology for a problem-solving assessment. Links denote the presence of a relationship.

Figure 2 shows an example of how the problem-solving assessment ontology would be instantiated. The example is based on a task developed to measure problem solving (Schacter et al., 1999). Briefly, the purpose of the assessment was to measure students' information-seeking problem-solving skills. The assessment was implemented with a combination of knowledge mapping and Web searching. Students first created a knowledge map on environmental science. They were not given any supplementary information and thus their map was based on their existing prior knowledge of the subject. After completing their initial knowledge map, the maps were scored in real-time and general feedback was returned to the students about which concepts "needed work." At that point, students were given access to the information space—Web pages on environmental science. Students could search for information, modify their knowledge maps, and request feedback.

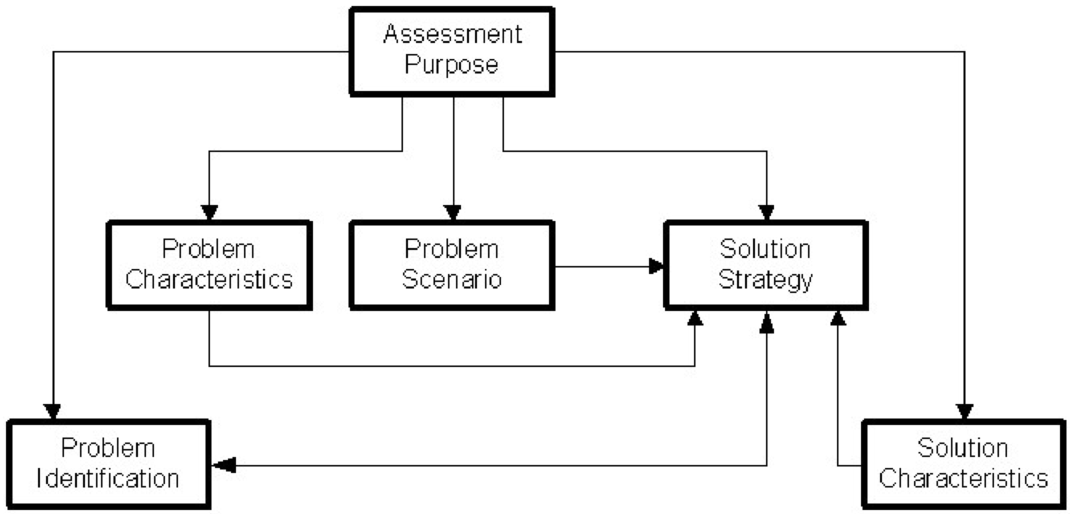

Figure 2. Example of a specific installation of the problem-solving assessment ontology. The bulleted items are the specific values of the properties bound to each concept. Unlabeled links represent constraints between concepts.

There are two important ideas to note in Figure 2. First is the idea that the representation is a particular instantiation of the assessment model ontology. Values are specified for each concept and link, and these values are task dependent. For example, the relationship between the instances for assessment purpose and solution strategy specifies a "divergent" strategy. This relationship is a particular value for the given assessment whereas the general relationship is the type of solution strategy elicited by the assessment. The second idea is the notion of constraint checking, indicated by the unlabeled links (unlabeled so that Figure 2 remains legible). The unlabeled links connect concepts that constrain each other—concepts that can have instances whose values are mutually acceptable or incompatible. As the complexity of the assessment model increases, the range of possible choices across all assessment concepts explodes. With constraint checking, the system can evaluate the state of the network on an ongoing basis and alert the user of incompatibilities as they arise.

Potential Applications

The preceding section described the use of ontologies to operationalize assessment models for use in computational environments. This section describes two applications for such online ontologies: assessment authoring support and individualized instruction.

Assessment Authoring Support

An ontology that explicitly represents an assessment model implies a given structure and constraints (via the relations among concepts). For assessment authoring purposes, structure is of very high utility because it allows the enforcement of a common and consistent framework. This structure can be leveraged to assist assessment authors (particularly non-experts) in designing assessments. Assessment authoring support could be in the form of (a) aiding assessment authors to populate the assessment ontology with values specific to the users' purposes, and (b) system constraint checking that would ensure that assessment authors are alerted to incompatible values.

To aid assessment authors, an authoring system can traverse the assessment ontology and gather information from users for only the relevant parameters. That is, because the ontology reflects all the relevant aspects of the assessment model, the user is queried for values for only those parameters that matter. The structure enforces the specification of important information and ignores variables outside the ontology. This scenario assumes the ontology is reasonably complete and accurate.

Constraint checking is carried out as the assessment author iteratively refines the values of the assessment parameters. An ontology network that converges to a steady-state condition implies that all constraints have been satisfied and all values for all assessment parameters are (simultaneously) acceptable. As an example, drawing from Table 1 and Figure 2, the problem-solving task was individual-based, short, academic (problem scenario = individual test of Web search skills), and explicitly stated (problem identification = stated), and remained fixed throughout the task (problem characterization = fixed). For this kind of task, it is unlikely that the task would require the use of contingency planning or help seeking. An assessment author who specifies such values would be alerted that the values are incompatible.

Additional authoring support could be provided by folding in assessment guideline information. Because of the flexibility of the ontology structure, slots could be added to bind guideline information to specific assessment concepts. The purpose of providing assessment guideline information would be to fill non-experts' (presumed) gaps in knowledge. Non-expert users are likely to have spotty knowledge of assessment in general—perhaps specific knowledge of only a few concepts and relations. Guidelines tied to concepts and relations should bolster what the assessment author knows about the domain, and just as important, what the person does not know. Similar applications of ontologies have been used to support physicians via clinical practice guidelines (Bernstam et al., 2000; de Clercq et al., 2001).

The use of ontologies to support assessment authoring seems particularly promising because of the nature of the anticipated users: non-experts who lack breadth and depth of knowledge of assessment. An ontology-based authoring system can impose structure on the authoring task as well as check that user-specified values simultaneously satisfy all constraints among the assessment components.

Individualized instruction

A second application of ontologies is to link student assessment information to relevant content for the purpose of individualized online instruction. One approach to individualizing instruction is to connect a particular set of observations within the assessment task to particular concepts and relations in a content ontology. During the assessment task, observations are taken and synthesized in real-time. Feedback and inferences about student knowledge can be made on the fly to tailor content delivery to an individual.

CRESST is currently exploring this idea by prototyping an ontology on rifle marksmanship for use by Marines (Chung, Delacruz, Dionne, & Bewley, 2003). Protégé 2000 is used to author the ontology (Noy et al., 2001). Currently, the ontology contains over 50 concepts that span seven core topics of rifle marksmanship and over 100 relationships (across 14 types of relations) that connect the concepts. These relationships capture in detail the causal and other relations that constitute deep knowledge in rifle marksmanship. Further, content—text, videos, and pictures—derived from experts, field manuals, and other information sources has been linked to particular concepts and relations in the ontology by type of knowledge (e.g., declarative, procedural). The information is chunked (e.g., definition, explanation, elaboration, procedure, picture of shot group, video of shooting position) and can be delivered in different packages of different grain-sizes. Thus, relevant content has been directly tied to the abstract representation of the ontology.

Before content can be recommended to the student, the assessment information needs to be synthesized. One way of synthesizing assessment observations to closely complement the content ontology is to use Bayesian inference networks. A Bayesian inference network, also known as an influence or probabilistic causal network, depicts the causal structure of a phenomenon in terms of nodes and relations (Jensen 2001). Nodes represent states, and links represent the influence relations among the nodes. Node states can be observable or unobservable.

The utility of a Bayesian inference network is that it yields the probability that an unobservable variable is in a particular state (e.g., understands trigger control) given observable evidence (student performance on different measures). Coupling content to performance can be achieved by associating concepts and relations in the ontology to unobservable variables in the Bayesian network. The probability of the unobservable variable being in a particular state is the system's inference about student performance, and the association and delivery of content (based on the variable state) and how individualized content delivery can be achieved.

Student performance on assessments of rifle marksmanship provides a concrete example of how such an approach would work. The example is based on the concept of trigger control (the skillful manipulation of the trigger that causes the rifle to fire without disturbing sight alignment). A particular student scored poorly on items that asked for (a) a simple definition of trigger control, (b) how trigger control relates to sight alignment, and (c) the pattern of shots for a shooter with poor trigger control. From this set of observations, one inference that could be drawn is that this student has little or no knowledge of trigger control.

The instructional remediation for this student could be to provide information on (a) trigger control—definition, explanation, and elaboration; (b) how trigger control is related to sight alignment (e.g., "A firm grip helps maintain good sight alignment because the grip helps ensure that the trigger is pulled straight toward the rear of the rifle."); and (c) the shot-dispersion pattern for poor trigger control with a picture and explanatory information.

Preliminary analyses of the pilot test results suggest that this approach is tractable and promising. For 10 high, medium, and low performers on the rifle marksmanship assessments, three independent raters of differing content knowledge were able to successfully recommend instructional content (i.e., what to deliver, how much to deliver, and what media format to use) based on the shooter's performance on a set of items (i.e., a short answer prior knowledge question on the topic, a shot-pattern depiction task on the same topic, and a short answer explanation task on the relation of the topic to sight alignment).

This finding is promising because it suggests that the approach to recommending instructional content can be tailored to an individual using assessment information. What makes the problem tractable is a content ontology with sufficient structure and detail that is consistent with the content and cognitive demands of the assessment and is associated with inferences about student performance.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Baker, E. L. (1997). Model-based performance assessment. *Theory Into Practice*, *36*, 247-254.

Baker, E. L., Sawaki, Y., & Stoker, G. (2002). The LEARNOME: A descriptive model to map and predict functional relationships among tests and educational system components." Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Bernstam, E., Ash, N., Peleg, M., Tu, S., Boxwala, A. A., Mork, M., Shortliffe, E. H., & Greenes, R. A. (2000). *Guideline classification to assist modeling, authoring, implementation and retrieval.* (Stanford Medical Institute Technical Report No. 2000-0848). Stanford University: Palo Alto, CA.

Chandrasekaran, R., Josephson, J. R., & Benjamins, V. R. (1999). What are ontologies, and why do we need them? *IEEE Intelligent Systems, 14*, 20–26.

Chung, G. K. W. K., Delacruz, G. C., Dionne, G. B., & Bewley, W. L. (2003). *Linking assessment and instruction using ontologies*. Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC): Orlando, FL.

de Clercq, P. A., Hasmon, A., Blom, J. A., & Korsten, H. H. M. (2001). The application of ontologies and problem-solving methods for the development of shareable guidelines. *Artificial Intelligence in Medicine*, *22,* 1–22.

Delacruz, G. C., Chung, G. K. W. K., & Bewley, W. L. (2003). Identifying learning phases using the human performance knowledge mapping tool (HPKMT) and microgenetic analysis. Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC): Orlando, FL.

Fensel, D., Hendler, J. Lieberman, H., & Wahlster, W. (Eds.). (2002). *Spinning the semantic web: Bringing the world wide web to its full potential*. Cambridge, MA: MIT Press.

Gennari, J., Musen, M. A., Fergerson, R. W., Grosso, W. E., Crubézy, M., Eriksson, H., Noy, N. F., & Tu., S. W. (2002). *The evolution of Protégé: An environment for knowledge-based systems development.* (Stanford Medical Institute Technical Report No. 2002-0943.) Stanford University: Palo Alto, CA.

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies, 43,* 907–928.

Jensen, F. V. (2001). *Bayesian networks and decision graphs*. New York: Springer-Verlag.

McGuinness, D. L. (2002). Ontologies come of age. In D. Fensel, J. Hendler, H. Lieberman, & W. Wahlster (Eds.), *Spinning the semantic web: Bringing the world wide web to its full potential*. Cambridge, MA: MIT Press.

Niemi, D., Chung, G. K. W. K., & Bewley, W. L. (2003). *Assessment design using ontologies: Linking assessment, content, and cognitive demands.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Noy, N. E., Sintek, M., Decker, S., Crubézy, M., Fergerson, R. W., & Musen, M. A. (2001). Creating semantic web contents with Protégé-2000. *IEEE Intelligent Systems*, *17*, 60–71.

Schacter, J., Herl, H. E., Chung, G. K. W. 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.