Issues in the Computer-Based Assessment of Collaborative Problem Solving

CSE Report 620

Harold F. O'Neil
University of Southern California / National Center for
Research on Evaluation, Standards, and Student Testing (CRESST)

San-hui Chuang
University of Southern California

Gregory K. W. K. Chung
University of California, Los Angeles / National Center for
Research on Evaluation, Standards, and Student Testing (CRESST)

February 2004

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

Project 2.1: Cognitively Based Models for Assessment and Instructional Design—Strand 2: Measures of Collaborative Problem Solving
Project Director: Harold F. O’Neil, University of Southern California / CRESST

The work reported herein was supported in part under the Educational Research and Development Centers Program, PR/Award Number R305B60002, as administered by the Institute of Education Sciences, U.S. Department of Education. The findings and opinions expressed in this report do not reflect the positions or policies of the National Center for Education Research, the Institute of Education Sciences, or the U.S. Department of Education.

The work reported herein was also supported 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 do not reflect the positions or policies of the Office of Naval Research.

ISSUES IN THE COMPUTER-BASED ASSESSMENT OF COLLABORATIVE PROBLEM SOLVING[1]

Harold F. O'Neil
University of Southern California/ National Center for
Research on Evaluation, Standards, and Student Testing (CRESST)

San-hui Chuang
University of Southern California

Gregory K. W. K. Chung
University of California, Los Angeles/National Center for
Research on Evaluation, Standards, and Student Testing (CRESST)

Abstract

Collaborative problem-solving skills are considered necessary skills for success in today's world of work and school. Cooperative learning refers to learning environments in which small groups of people work together to achieve a common goal, and problem solving is defined as "cognitive processing directed at achieving a common goal when no solution method is obvious to the problem solver" (R. E. Mayer & M. C. Wittrock, 1996, p. 47). Thus, collaborative problem solving is defined as problem-solving activities that involve interactions among a group of individuals. This paper will address several key issues (e.g., theory and measurement of collaborative problem solving and issues in measuring problem-solving processes). We rely on computerization of the administration, scoring, and reporting of collaborative problem-solving skills, thus potentially increasing reliability and validity.

In view of the changing needs in the workforce, the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) established a research program to identify and assess workforce skills. They examined the existing literature by reviewing five seminal studies on workforce readiness. All five studies identified higher order thinking, interpersonal and teamwork skills, and problem solving as the necessary generic skills needed for success in today's world (O'Neil, Allred, & Baker, 1997). Of these generic skills, university graduates and

---

[1] A revised version of this report is published as: O'Neil, H. F., Chuang, S., & Chung, G. K. W. K. (2003). Issues in the computer-based assessment of collaborative problem solving. *Assessment in Education*, *10,* 361-373.

employees rated thinking, decision making, communications skills, teamwork, and collaborative skills as the most important (Sinclair, 1997).

In addition, teachers and professors at all educational levels have recognized the changes in the requisite workforce skills, and some have modified their curriculum and instruction to prepare their students by incorporating problem solving and cooperative learning in daily instruction. Empirically, collaborative problem solving has been shown in educational research to enhance students' cognitive development (Webb, Nemer, & Chizhik, 1998; Zhang, 1998). Recently, many educational assessment programs have also used collaborative small-group tasks, in which students work together to solve problems or to accomplish projects, to evaluate learning results (Webb et al., 1998). Moreover, when students enter the workforce, they inevitably have to work in groups.

In this report, we will discuss the theory and measurement of problem solving, the theory and measurement of collaborative problem solving, current studies using the theoretic frameworks, and issues in measuring problem-solving processes, and will present a summary and a new conceptualization of our current research.

Theory and Measurement of Problem Solving

O'Neil (1999) defined problem-solving as consisting of three facets: content understanding, problem-solving strategies, and self-regulation. A good problem solver (a) understands the content well (content knowledge), (b) possesses specific intellectual skills (problem-solving strategies), and (c) is able to plan the use of her resources and skills and, during the process, monitors her own progress toward the end goal of solving the problem (self-regulation) (p. 256).

Most traditional assessment of problem solving relies on questionnaires, self-reporting, interviews, or naturalistic observations. However, those methods are not able to capture the essence of the problem-solving process (Chung, O'Neil, & Herl, 1999). Recently, with the emergence of computer technology, researchers have used computer simulation as an assessment tool for problem-solving skills (e.g., Herl, O'Neil, Chung, & Schacter, 1999; Hsieh & O'Neil, 2002; Schacter, Herl, Chung, Dennis, & O'Neil, 1999). Computer technology, with a powerful relational database, makes capturing problem-solving processes easier and less costly, especially in a large-scale test setting.

Content understanding can be measured using knowledge maps, which are graphical representations consisting of terms (concepts) and links (interrelationships between concepts). Knowledge maps have been extensively used in K-12 classrooms, especially in the teaching of science (Schau, Mattern, Zeilik, Teague, & Weber, 2001). Various research studies on knowledge maps also show them to be effective for teaching, learning, and assessment purposes (Herl et al., 1999; Hurwitz & Abegg, 1999; Ruiz-Primo, Schultz, & Shavelson, 1997; Schau et al., 2001).

Self-regulation can be defined as consisting of metacognition and motivation (O’Neil, 1999). O’Neil and Herl (1998) proposed examining metacognition in two aspects—planning and self-checking—and examining motivation as self-efficacy and effort. These four components make up our measure of self-regulation in problem solving. Planning is the first step, because one must have a goal and a plan to achieve the goal; self-monitoring or self-checking is assumed to be an essential mechanism to monitor the processes required for goal achievement. Self-efficacy is defined as one’s confidence in being capable of accomplishing a particular task (Bandura, 1997), and effort is the extent to which one works hard on a task. Self-regulation, both trait and state aspects, can be measured by questionnaires (O’Neil & Abedi, 1996; O’Neil & Herl, 1998).

Assessment of problem-solving strategies is more problematic than measuring content understanding or self-regulation. According to O’Neil (1999), problem-solving strategies are either domain independent or domain dependent. Domain-independent strategies may be applied over several subject areas, whereas domain-dependent strategies are more specific to an individual subject area. An example of a domain-independent strategy would be computer-based search strategies.

Computer-based information searching using the World Wide Web has attracted considerable attention in education and library science. Cyveillance (“Internet Exceeds 2 Billion Pages,” 2000) estimated in 2000 that the World Wide Web contained more than 2 billion publicly accessible information pages, and that it will continue to grow. According to the National Center for Education Statistics (2000), over 95% of public schools in the United States have computers. In addition, about 90% of classrooms have Internet access (Smith & Broom, 2003). More and more schools demand that students do research using search techniques on the World Wide Web so that they can satisfactorily complete their projects (Smith & Broom). In fact, research is the most common classroom Internet use (Becker, 1999).

However, merely getting online to the World Wide Web does not automatically result in getting the information needed. According to Smith and Broom (2003), students and teachers alike still lack basic information technology knowledge and skills. In addition, the current curriculum, instruction, and assessments do not adequately make use of the capabilities of today's networked information systems (Smith & Broom). Electronic information seeking is not just a single step, but a process of steps together. It involves determining the information needed, choosing topics to pursue, locating sites and locating information to increase overall domain understanding, analysis and evaluation of the information found, and finally, ending the search and returning to solving the problem (Lazonder, 2000).

In order to assess information-seeking strategies (or domain-independent problem-solving strategies) for a knowledge mapping task, CRESST created a simulated Internet Web space to search (Schacter et al., 1999). In one study, Schacter et al. used this simulated Internet Web space to measure individual problem-solving strategies and found that content understanding scores for participants increased significantly with access to the simulated Web space. Information-seeking processes such as browsing, searching, and accessing feedback improved students' mapping performance significantly on a knowledge map posttest.

Theory and Measurement of Collaborative Problem Solving

Currently, the terms *cooperative learning*, *collaborative learning,* and *group work* are used interchangeably in the education research literature to mean similar things. Cooperative learning refers to learning environments in which small groups of students work together to achieve a common goal (Underwood & Underwood, 1999). A *problem* is an unknown resulting from any situation where a person seeks to fulfill a need to accomplish a goal (Jonassen, 1997), and *problem solving* is "cognitive processing directed at achieving a goal when no solution method is obvious to the problem solver" (Mayer & Wittrock, 1996, p. 47). Collaborative problem solving thus refers to problem-solving activities that involve interactions among a group of individuals (Zhang, 1998).

In our research, we used the teamwork process model developed by CRESST researchers to measure collaborative learning processes. The CRESST teamwork process model was based on a teamwork model reported by Salas and his colleagues (e.g., Morgan, Salas, & Glickman, 1993; Salas, Dickinson, Converse, & Tannenbaum,

1992) and consists of six skills: (a) adaptability, (b) coordination, (c) decision making, (d) interpersonal, and (f) communication (Chung et al., 1999).

Adaptability

Adaptability refers to the group's ability to "monitor the source and nature of problems through an awareness of team activities and factors bearing on the task" (O'Neil, Chung, & Brown, 1997). That is, adaptability is used for the detection and correction of problems. In a knowledge mapping task, an adaptable team should detect problems with their knowledge map at a deep (semantic) level and at a surface level, by identifying inaccuracies as well as the strength and significance of relationships, and by recognizing that the given set of concepts and links should be included in their map, respectively. In our current studies, students are low in prior knowledge about environmental sciences; thus, we do not usually expect any effect of adaptability on performance.

Coordination

Coordination is defined as a group's "process by which group resources, activities, and responses are organized to ensure that tasks are integrated, synchronized, and completed with established temporal constraints" (O'Neil, Chung, & Brown, 1997, p. 413). Therefore, in a knowledge mapping task, coordinating strategies will include the domain expertise that group members use to determine the relationships between concepts, members' consciousness of the time constraints, and their ability to respond appropriately.

Decision Making

Decision making is defined as a group's "ability to integrate information, use logical and sound judgment, identify possible alternatives, select the best solution, and evaluate the consequences" (O'Neil, Chung, & Brown, 1997, p. 415). Effective teams or groups make decisions that take into consideration all available information, and thus decision making is regarded as playing a significant role in performance (Chung et al., 1999). In addition, Chung et al. (1999) indicated that, compared with group members who lacked of prior knowledge, group members with relevant prior knowledge might be more likely to engage in substantive discussions concerning the relationships between concepts.

In our current studies, participants are expected to have little prior knowledge in the domain of environmental science. However, we believe that, through seeking information from the simulated Web space, the participants will have the opportunity to engage in substantive discussion about the relationships. Thus, decision making is expected to have a positive effect on group performance.

Interpersonal Skill

Interpersonal skill is defined as “the ability to improve the quality of team member interactions through the resolution of team members’ dissent, or the use of cooperative behavior” (O’Neil, Chung, & Brown, 1997, p. 416). Interpersonal processes are important because they minimize intergroup conflict, as well as foster team interdependence (Weng, 1999). In our current studies, the group consists of only two or three members, and intergroup conflict should be highly unlikely. Therefore, no relationship between interpersonal skills and group performance is expected.

Leadership

Leadership is defined as “the ability to direct and coordinate the activities of other team members, assess group performance, assign tasks, plan and organize, and establish a positive atmosphere” (O’Neil, Chung, & Brown, 1997, p. 417). In our current studies, leadership is expected to have a positive effect on group outcome performance.

Communication

Communication is defined as “the process by which information is clearly and accurately exchanged between two or more team members in the prescribed manner and by using proper terminology, and the ability to clarify or acknowledge the receipt of information” (O’Neil, Chung, & Brown, 1997, p. 417). In our current studies, it is expected that communication should have a positive effect on group performance because the group cannot complete the task successfully without communicating with each other.

Current Studies Using the Theoretic Frameworks

Three CRESST studies used both the theory and measures that are outlined above. In all studies, the knowledge map construction task contained 18 environmental science concepts (i.e., atmosphere, bacteria, carbon dioxide, climate, consumer, decomposition, evaporation, food chain, greenhouse gases, nutrients, oceans, oxygen, photosynthesis, producer, respiration, sunlight, waste, water cycle), and seven relationships or links (i.e., causes, influence, part of, produces, requires, used for, uses). Students were asked to use these terms and links to construct knowledge maps on the computer.

In the first study, Schacter et al. (1999) used a simulated Internet Web space to measure individual problem-solving strategies on networked computers and found that content understanding scores for participants increased significantly with access to a simulated Web space. Information-seeking processes such as browsing, searching, and accessing feedback improved students' performance significantly on the knowledge map posttest.

In the second study, Chung et al. (1999) used a computer-based collaborative knowledge mapping task to measure team processes and team outcomes. Team processes were measured through predefined messages. Each message belonged to one of the six teamwork skills (e.g., adaptability or coordination) in the CRESST taxonomy of teamwork (O'Neil, Chung, & Brown, 1997). Unfortunately, in Chung et al.'s study, no significant positive relationships were found between most team processes and outcome measures; surprisingly, decision making and communication were found to have negative effects on outcome performance.

Hsieh and O'Neil (2002) hypothesized that the lack of useful feedback and the type of task involved (not a real group task) in Chung et al.'s (1999) study may have influenced the results. In a third study, based on these two hypotheses, Hsieh and O'Neil attempted to improve Chung et al.'s (1999) results by changing the nature of the task to a real "group task" and providing more extensive feedback. A group task is a task in which no single individual possesses all the resources and no single individual is likely to solve the problem or accomplish the task objectives without at least some input from the others in the group (Cohen & Arechevala-Vargas, 1987). Therefore, Hsieh and O'Neil modified the original task to be a real group task by assigning a specific role (either the leader, who constructs the knowledge map, or the searcher, who conducts the simulated Web space search and supplies information to the leader) to each group member in order to meet the group task requirements (neither the leader nor the searcher possesses all the knowledge, heuristic problem-solving strategies, materials, and skills, and neither individual is likely to solve the problem or accomplish the task objectives without at least some input from the other). In addition, in order to improve group performance, Hsieh and O'Neil compared two levels of feedback (knowledge of response feedback and adapted knowledge of response feedback) on group map construction processes.

Feedback accessing was demonstrated to be significantly related to students' performance in Schacter et al.'s (1999) studies.

By combining the methodologies of Schacter et al.'s (1999) and Chung et al.'s (1999) studies, Hsieh and O'Neil (2002) successfully demonstrated the use of adapted knowledge of response feedback to be better in group knowledge map performance than knowledge of response feedback. While doing so, they also evaluated student collaborative problem-solving skills and team processes on a computer-based knowledge mapping group task with a simulated Web space as an information source. They demonstrated that decision making and leadership were positively related to group knowledge map performance. However, searching performance in their study was unexpectedly negatively related to group knowledge map performance. We hypothesized that, even though feedback provided participants a direction as to "what" area to improve for search and task performance, the feedback did not give practical tips on "how" to improve the performance by providing more task-specific feedback on search strategies. Furthermore, participants did not know how to search efficiently or effectively in this task. Ongoing research in our lab is attempting to evaluate these hypotheses by providing more task-specific feedback to groups that have been trained in searching.

Issues in Measuring Collaborative Problem-Solving Processes

Online collaborative problem-solving tasks offer new measurement opportunities when information on what individuals and teams are doing is synthesized along the cognitive dimension (Baker, 1997; Baker & Mayer, 1999; Baker & O'Neil, 2002). In the following section, we describe the interface design features that can support online measurement, and then describe three approaches suitable for quantifying online collaborative problem-solving processes in an online context. For each type of analysis approach, we briefly describe the method with examples drawn from our prior work.

User Interface

The interaction between an individual and the computer interface and the interaction among group members can be rich sources of information when fused with task-related context information. Measures of what students are doing and when they are doing it, synchronized with measures of task-related variables, may provide enough information to let us draw inferences about learning processes. A key component in achieving this knowledge is the user interface. An interface that captures intentional acts that reflect good or poor judgment greatly strengthens inferences that can be made about cognition based on online behavior (Chung & Baker, 2002; Chung, de Vries, Cheak, Stevens, & Bewley, 2002). However, a considerable challenge remains: Given a continuous stream of student online behavior and task-related data, how should the data be synthesized to yield useful information about the individual and team? Our prior work in measuring teamwork skills and other work in developing online assessments offer some guidance.

Predefined Messages

Existing ideal approaches to measuring group processes rely almost exclusively on observational methods. Behavior that is videotaped or audiotaped and analysis of think-aloud protocols are the most common techniques used to measure group processes. However, these observations must be transcribed, coded, and analyzed post hoc, and such methods are neither practical nor cost-effective in online settings.

One interesting technique we have tested to measure group processes is to provide group members with predefined messages, which they use to communicate with each other (Chung et al., 1999; Hsieh & O’Neil, 2002; O’Neil, Chung, & Brown, 1997). Participants worked in online teams using a custom-developed computer conferencing system. The teams were required to jointly complete a task (a simulated negotiation or a knowledge map). Team members used the predefined messages to communicate with each other, and measures of teamwork processes were computed based on the quantity and type of messages used (i.e., each message was coded a priori as representing adaptability, coordination, decision making, interpersonal skill, or leadership). In general, participants were able to communicate using the predefined messages and to successfully complete the tasks, and the team processes and outcomes were measurable. The use of messages provides a tractable way of measuring communication and team skills. It allows real-time scoring and reporting of collaborative problem-solving skills.

Lag Sequential Analyses

If how a group interacts over time is of interest, then methods from observational research may be appropriate. One technique is lag sequential analyses (Bakeman, Adamson, & Strisik, 1995; Bakeman & Gottman, 1997; Gottman & Roy, 1990). Lag sequential analysis can be used to describe the sequential dependencies within a communication stream. For example, using the predefined message technique described earlier, lag sequential analyses can be used to contrast the communication patterns between more and less effective groups (e.g., Bowers & Jentsch, 1998; Hirokawa, 1980). In addition, if sequential dependencies exist, then such information could be used to simulate missing team members.

In an exploratory study, we applied this technique to quantify the communication patterns between effective and less effective teams in O'Neil, Chung, and Brown's (1997) study. Results showed a significant effect of message sending order, suggesting a dependence between successive messages sent among team members. Further, this relationship differed by groups that did and did not reach an agreement in the negotiation task. In general, the most likely message to be sent was a message in the same category as the preceding message, especially for coordination and interpersonal messages. Teams that did not reach an agreement were less likely to send coordination or interpersonal messages after an adaptability message, or an adaptability message followed by a coordination message. In contrast, teams that did reach an agreement were less likely to send a leadership or decision-making message after an adaptability message.

We found statistical support for the intuitive notion that sequential dependencies exist among team members' communications. The order of message usage and thus the use of team processes differed between effective and less effective groups. Interaction sequences within group communications are largely unexplored, and lag sequential analyses provide a way of analyzing such interactions.

Influence (Bayesian) Networks

Because the amount and type of data in online contexts can be overwhelming, a data fusion strategy is needed to cogently synthesize the data. One promising approach is the use of influence networks (also known as Bayesian networks, probabilistic networks, or causal networks). A Bayesian network graphically depicts the causal structure of a phenomenon as a network (Jensen, 2001) and expresses the phenomenon in terms of probabilities. Causal influences among the variables are expressed directly, which allows the testing of hypotheses about dependencies among variables. Further, probabilities for the existence of inferred variables in the model can be computed in real-time given observable evidence (e.g., the probability that collaboration is occurring given the amount and type of communication).

Chung, Delacruz, et al. (2002) developed a Bayesian network model to represent small-group collaborative processes given very simple physical data types for each group member (i.e., location, head orientation, and whether speech). The study is embedded in a larger study of wireless networked sensors in a classroom setting for instructional and assessment purposes. (Chen et al., 2002; Srivastava, Muntz, & Potkonjak, 2001).

Simulated collaborative group scenarios were used to test the approach. The data streams consisted of only position, head orientation, and speech on / off for each individual in a group, at 1-second resolution. Probabilities from the model were compared to a priori expectations of what was occurring at each second in the group. Results showed the model performed very well with scenarios that displayed prototypical small-group behavior (that is, students situated around a table where members were facing each other, in proximity to each other, paying attention to the speaker, and exhibiting turn-taking and other speaking behaviors). The simulation study suggested that high-level, complex phenomena could be captured and modeled using large amounts of very simple data types.

In summary, we have described several methods that appear promising as ways to analyze collaborative problem-solving processes in online environments. These approaches are suitable because they can take advantage of information provided by process data that have been typically ignored or viewed as too cumbersome to analyze.

Summary

Collaborative problem solving is considered a necessary skill for success in today’s world. Cooperative learning refers to learning environments in which small groups of students work together to achieve a common goal, and problem solving is defined as “cognitive processing directed at achieving a goal when no solution method is obvious to the problem solver” (Mayer & Wittrock, 1996, p. 47). Thus, collaborative problem solving is defined as problem solving activities that involve interactions among a group of individuals. Figure 1 shows the components and their relationship to each other in collaborative problem solving.

As seen in Figure 1, collaborative problem solving is first divided into two components: collaboration and problem solving. According to O’Neil, Chung, and Brown (1997), collaboration on team skills can be assessed by six skills: adaptability, coordination, decision making, interpersonal, leadership, and communication. According to O’Neil (1999), problem solving has three factors: content understanding, problem-solving strategies, and self-regulation. Problem-solving

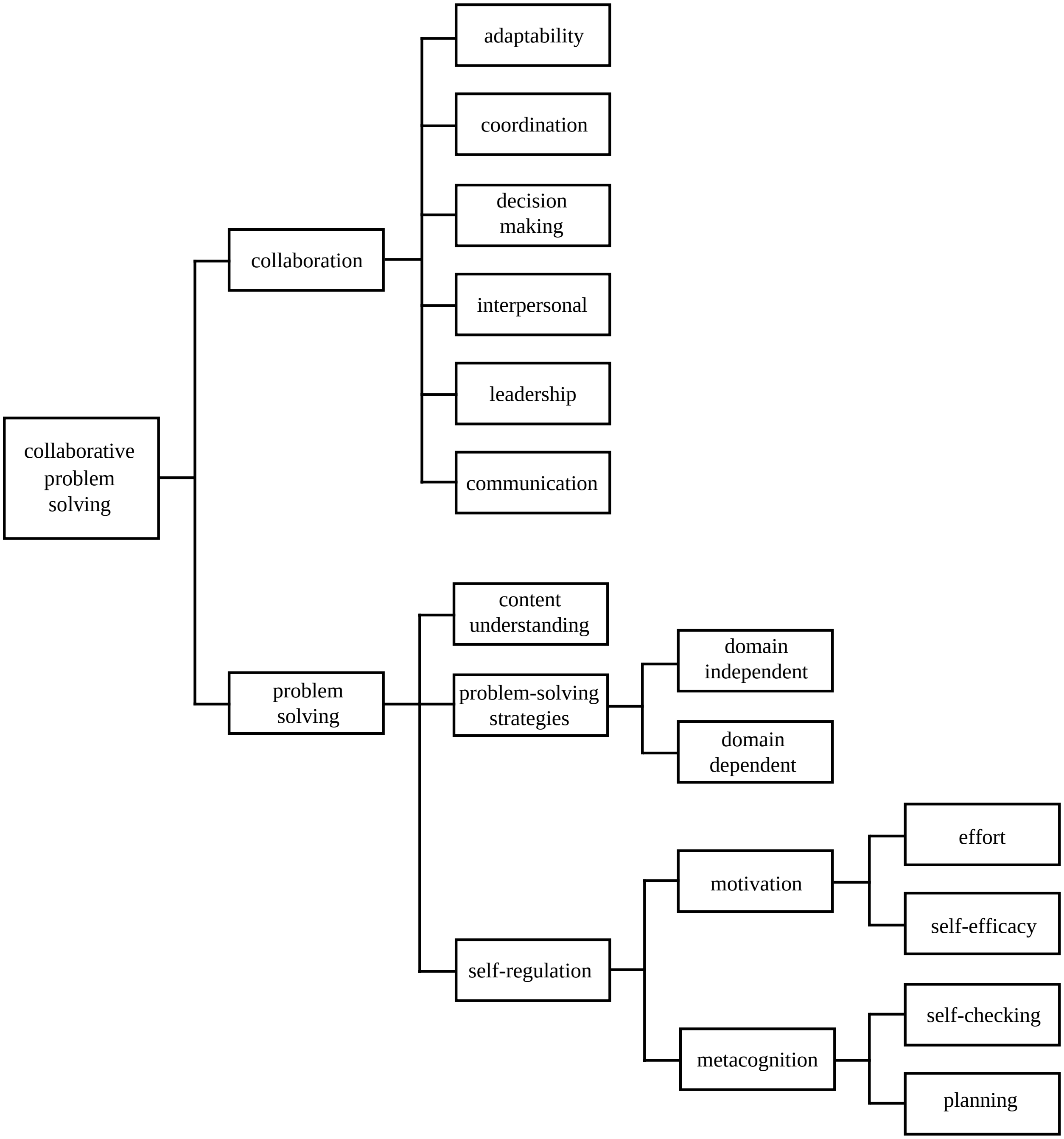

Figure 1. Collaborative problem-solving components.

strategies can be domain dependent or domain independent. Self-regulation has two main components—motivation and metacognition—and each of them has two components respectively. Motivation consists of effort and self-efficacy, and metacognition consists of self-checking and planning.

Another view that we are exploring in our lab is to conceptualize our research in terms of feedback in a dynamic testing environment. According to Grigorenko and Sternberg (1998), dynamic testing is “a collection of testing procedures designed to quantify not only the products or even the processes of learning but also the potential to learn” (p. 75). In order to fulfill the claims made for it, dynamic testing involves testing not only end products but also learning processes. This type of testing is quite different from traditional, static testing, which only assesses the learned end products. Another difference between dynamic testing and static testing is the role of feedback (Grigorenko & Sternberg). According to Grigorenko and Sternberg, in traditional static testing, feedback about performance is usually not given during a test. In dynamic testing, feedback is given during the test to help assess learning. In dynamic testing, an examiner presents a sequence of gradually more difficult tasks. After each student performance, the examiner gives the student feedback and continues until the student either solves the problem or chooses to give up. In our current studies, the “examiner” is the computer software. The basic goal of dynamic testing is to see, when feedback is given, whether test takers change and how they change. This is done through provision of feedback during the test; however, there are no agreed-upon ideas about how much information should be included in the feedback, nor about how quantitative measures could be derived. Currently, different approaches in dynamic testing vary in the amount of information contained in the feedback (Grigorenko & Sternberg). We believe such a framework will provide a theoretical rationale for our continuing research on the assessment of collaborative problem solving.

References

Bakeman, R., Adamson, L. B., & Strisik, P. (1995). Lags and logs: Statistical approaches to interaction (SPSS version). In J. M. Gottman (Ed.), *The analysis of change* (pp. 279-308). Mahwah, NJ: Erlbaum.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). New York: Cambridge University Press.

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

Baker, E. L., & Mayer, R. E. (1999). Computer-based assessment of problem solving. *Computers in Human Behavior, 15,* 269-282.

Baker, E. L., & O'Neil, H. F., Jr. (2002). Measuring problem solving in computer environments: Current and future states. *Computers in Human Behavior, 18,* 609-622.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: Freeman.

Becker, H. J. (1999). *Internet use by teachers: Conditions of professional use and teacher-directed student use. Teaching, learning, and computing: 1998 National survey: Report #1* [On-line]. Irvine: University of California, Irvine, Center for Research on Information Technology and Organizations, and University of Minnesota. Retrieved January 22, 2003, from the Center for Research on Information Technology and Organizations, University of California, Irvine, Web site, http://www.crito.uci.edu/TLC/FINDINGS/internet-use/startpage.htm

Bowers, C. A., & Jentsch, F. (1998). Analyzing communication sequences for team training needs assessment. *Human Factors, 40,* 672-679.

Chen, A., Muntz, R. R., Yuen, S., Locher, I., Park, S. I., & Srivastava, M. B. (2002). A support infrastructure for the Smart Kindergarten. *Pervasive Computing, 1*(2), 49-57.

Chung, G. K. W. K., & Baker, E. L. (2002). *Examining the feasibility of measuring problem-solving processes using a click-through interface* (Draft deliverable to the Center for Digital Innovation, UCLA). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Chung, G. K. W. K., de 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.

Chung, G. K. W. K., Delacruz, G. C., de Vries, L. F., Phan, C. H., Srivastava, M. B., & Alarcon, R. (2002, September). *Fusing wireless sensor data to measure small-group collaborative processes in real-time.* Presentation at the annual conference of the

Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles.

Chung, G. K. W. K., O’Neil, H. F., Jr., & Herl, H. E. (1999). The use of computer-based collaborative knowledge mapping to measure team processes and team outcomes. *Computers in Human Behavior, 15,* 463-494.

Cohen, B. P., & Arechevala-Vargas (1987). *Interdependence, interaction, and productivity* (Working Paper No. 87-3). Stanford, CA: Stanford University, Center for Sociology Research.

Gottman, J. M., & Roy, A. K. (1990). *Sequential analyses: A guide for behavioral researchers.* New York: Cambridge University Press.

Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin, 124,* 75-111.

Herl, H. E., O’Neil, H. F., Jr., Chung, G. K. W. K., & Schacter, J. (1999). Reliability and validity of a computer-based knowledge mapping system to measure content understanding. *Computer in Human Behavior, 15,* 315-333.

Hirokawa, R. Y. (1980). A comparative analysis of communication patterns within effective and ineffective decision-making groups. *Communication Monographs, 47,* 312-321.

Hsieh, I.-L., & O’Neil, H. F., Jr. (2002). Types of feedback in a computer-based collaborative problem solving group task. *Computers in Human Behavior, 18,* 699-715.

Hurwitz, C. L., & Abegg, G. (1999). A teacher’s perspective on technology in the classroom: computer visualization, concept maps and learning logs. *Journal of Education, 181,* 123-143.

Internet exceeds 2 billion pages. (2000). Retrieved January 23, 2003, from http://www.cyveillance.com/us/newsroom/pressr/000710.asp

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

Jonassen, D. H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning. *Educational Technology Research and Development, 45,* 65-94.

Lazonder, A. W. (2000) Exploring novice users’ training needs in searching information on the WWW. *Journal of Computer Assisted Learning, 16,* 326-335.

Mayer, R. E., & Wittrock, M. C. (1996). Problem-solving transfer. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 47-62). New York: Macmillan Library Reference USA, Simon & Schuster Macmillan.

Morgan, B. B., Salas, E., & Glickman, A. S. (1993). An analysis of team evolution and maturation. *Journal of General Psychology, 120,* 277-291.

National Center for Education Statistics. (2000). *Internet access in U.S. public schools and classrooms: 1994-1999* (NCES # 2000086). Washington, DC: U.S. Department of Education.

O'Neil, H. F., Jr. (1999). Perspectives on computer-based performance assessment of problem solving. *Computers in Human Behavior, 15,* 225-268.

O'Neil, H. F., Jr., & Abedi, J. (1996). Reliability and validity of a state metacognitive inventory: Potential for alternative assessment. *Journal of Educational Research, 89,* 234-245.

O'Neil, H. F., Jr., Allred, K., & Baker, E. L. (1997). Review of workforce readiness theoretical frameworks. In H. F. O'Neil, Jr., (Ed.), *Workforce readiness: Competencies and assessment* (pp. 3-25). Mahwah, NJ: Erlbaum.

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: Erlbaum.

O'Neil, H. F., Jr., & Herl, H. E. (1998, April). *Reliability and validity of a trait measure of self-regulation*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Ruiz-Primo, M. A., Schultz, S. E., & Shavelson, R. J. (1997). *Concept map-based assessment in science: Two exploratory studies* (CSE Tech. Rep. No. 436). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3-30). Norwood, NJ: Ablex.

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.

Schau, C., Mattern, N., Zeilik, M., Teague, K., & Weber, R. (2001). Select-and-fill-in concept map scores as a measure of students' connected understanding of science. *Educational and Psychological Measurement, 61,* 136-158.

Sinclair, K. E. (1997). Workforce competencies of college graduates. In H. F. O'Neil, Jr., (Ed.), *Workforce readiness: Competencies and assessment* (pp. 103-120). Mahwah, NJ: Erlbaum.

Smith, M. S., & Broom, M. (2003). The landscape and future of the use of technology in K-12 education. In H. F. O'Neil, Jr., & Perez, R. S. (Eds.), *Technology applications in education: A learning view* (pp. 3-30). Mahwah, NJ: Erlbaum.

Srivastava, M., Muntz, R., & Potkonjak, M. (2001). Smart kindergarten: Sensor-based wireless networks for smart developmental problem-solving environments. In *Proceedings of the ACM SIGMOBILE 7th Annual International Conference on Mobile Computing and Networking* (pp. 132-138). New York: AMC Press.

Underwood, J., & Underwood, G. (1999). Task effects on co-operative and collaborative learning with computers. In K. Littleton & P. Light (Eds.), *Learning with computers* (pp. 10-23). New York: Routledge.

Webb, N. M., Nemer, K. M., & Chizhik, A. W. (1998). Equity issues in collaborative group assessment: Group composition and performance. *American Educational Research Journal, 35,* 607-651.

Weng, A. L.-B. (1999). *A reliability and validity study of Chinese version of a teamwork skills questionnaire.* Unpublished doctoral dissertation, University of Southern California, Los Angeles.

Zhang, J. (1998). A distributed representation approach to group problem solving. *Journal of the American Society for Information Science, 49,* 801-809.