Final Report for Validation of Problem-Solving Measures

CSE Technical Report 501

Howard E. Herl,[*] Harold F. O'Neil, Jr.,[†] Gregory K. W. K. Chung,[*]
Cecilia Bianchi,[*] Shu-ling Wang,[*] Richard Mayer,[‡]
Charlotte Yeh Lee,[*] Angela Choi,[*] Thomas Suen,[*] and Aileen Tu[*]

[*]University of California, Los Angeles / CRESST
[†]University of Southern California / CRESST
[‡]University of California, Santa Barbara

March 1999

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 by contract #85300-97-0057 from Statistics Canada, Government of Canada, to the National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles.

FINAL REPORT FOR
VALIDATION OF PROBLEM-SOLVING MEASURES

Howard E. Herl,* Harold F. O'Neil, Jr.,† Gregory K. W. K. Chung,*
Cecilia Bianchi,* Shu-ling Wang,* Richard Mayer, ‡
Charlotte Yeh Lee,* Angela Choi,* Thomas Suen,* and Aileen Tu*

Executive Summary

The purpose of this final report is to summarize several studies regarding a new measure of problem solving. It is assumed that the reader is familiar with the International Life Skills Survey project.

Baker (1997), citing Mayer and Wittrock (1996), described CRESST's approach to measuring problem solving as "cognitive processing directed at achieving a goal when no solution method is obvious to the problem solver" (p. 47). This definition can be further analyzed into components suggested by the expertise literature, that is, content understanding or domain knowledge (measured by a knowledge map), domain-specific problem-solving strategies (measured by response to a troubleshooting prompt), and self-regulation (measured by a trait self-report survey). To be a successful problem solver, one must know something (content knowledge), possess intellectual tricks (problem-solving strategies), be able to plan and monitor one's progress towards solving the problem (metacognition), and be motivated to perform (effort and self-efficacy).

In three separate studies, CRESST has administered knowledge maps to measure content understanding, short-answer questions to measure problem-solving strategies, and self-report questionnaires to measure self-regulation traits. This set of measures constitute our new measure of problem solving.

This report focuses on CRESST data collection efforts described in *Proposal for Concept Map Variables: Problem-Solving Study* (Baker, 1997). The work presented here summarizes several issues in our research, including measurement of problem solving, and investigation of the impact of providing content information (Yes / No) and scoring criteria (Yes / No) for knowledge mapping tasks. Reliability and validity issues related to the use of knowledge maps for measuring content understanding, domain-specific problem-solving prompts for measuring participants' problem-solving strategies, and self-regulation questionnaires for measuring levels of

---

*University of California, Los Angeles / CRESST, †University of Southern California / CRESST,
‡University of California, Santa Barbara

participants' metacognition and motivation are presented within the larger context of measuring problem solving. There were two pilot studies and a main study.

Two pilot studies (one in a Taiwan high school and one in a southern California university) were conducted to assess the feasibility of our approach to measuring problem solving across different contexts, and to try out and refine measures and procedures for the main study. The main study was conducted in southern California. For the main study, 129 adults were recruited from two different temporary employment agencies in Los Angeles, CA, and paid to participate in CRESST and Statistics Canada's problem-solving study.

Data analyses performed on main study data showed that providing scoring instructions to participants did not significantly affect their performance on knowledge mapping tasks, including measures of content understanding, number of terms of used, and number of links constructed. There were no gender effects found for knowledge mapping measures for the bicycle tire pump task. However, there were significant effects found for two knowledge mapping task measures for the respiratory system task, where males constructed significantly greater numbers of links and achieved significantly higher content understanding scores. Finally, there was an impact of providing content information. Its provision resulted in higher performance scores. In general there was sufficient reliability and validity to use these new measures of problem solving for the International Life Skills Survey.

FINAL REPORT FOR
VALIDATION OF PROBLEM-SOLVING MEASURES

Howard E. Herl,* Harold F. O'Neil, Jr.,† Gregory K. W. K. Chung,*
Cecilia Bianchi,* Shu-ling Wang,* Richard Mayer, ‡
Charlotte Yeh Lee,* Angela Choi,* Thomas Suen,* and Aileen Tu*

This final report for problem solving summarizes several issues: our theoretical framework of problem solving, a way of measuring the construct, and a focused discussion of a series of research studies on the reliability / validity of our new problem-solving measures.

Our definition of problem solving is based on Mayer (Mayer & Wittrock, 1996; see also Mayer, 1992): "Problem-solving is cognitive processing directed at achieving a goal when no solution method is obvious to the problem solver" (p. 47) As may be seen in Figure 1, this definition can be further analyzed into components suggested by the expertise literature, that is, content understanding or domain knowledge (which can be measured by a knowledge map), domain-specific problem-solving

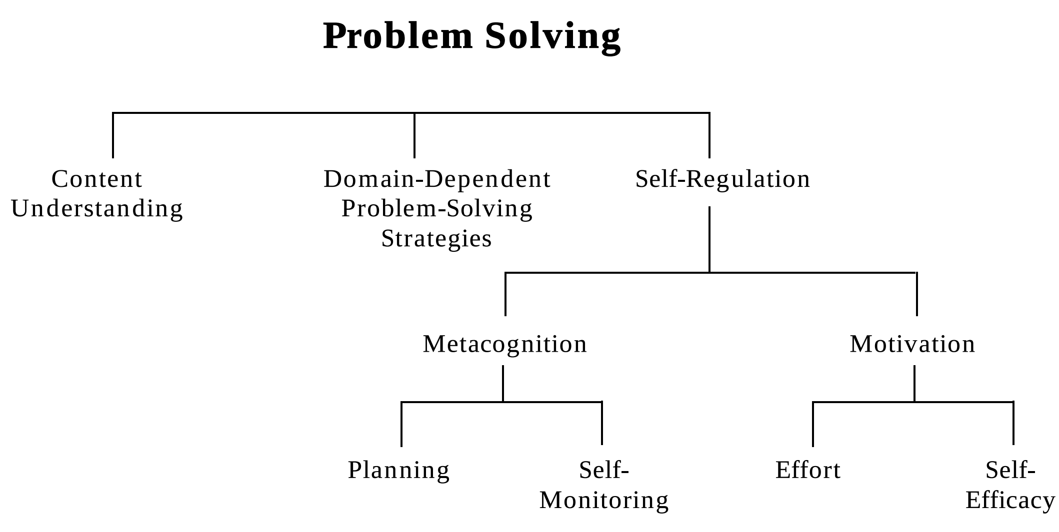


Figure 1. CRESST model of problem solving.

*University of California, Los Angeles / CRESST, †University of Southern California / CRESST,
‡University of California, Santa Barbara

strategies (measured by response to a troubleshooting prompt), and self-regulation (measured by a trait self-report survey). In summary, to be a successful problem solver, one must know something (content knowledge), possess intellectual tricks (problem-solving strategies), be able to plan and monitor one's progress towards solving the problem (metacognition), and be motivated to perform (effort and self-efficacy).

Our approach to measuring problem solving is to use paper-and-pencil knowledge maps to assess content understanding (Herl, Baker, & Niemi, 1996), explanation lists to assess domain-dependent problem-solving strategies, and finally a questionnaire to measure self-regulation (see Table 1). All of these facets of problem solving would eventually be measured in 30 minutes.

Prior CRESST Research in Problem Solving

Currently, the ideal assessment of problem solving is based on think-aloud protocols (Ericsson & Simon, 1993; Voss & Post, 1988; Voss, Tyler, & Yengo, 1983) or performance assessments that require extensive human rater scoring (Baxter, Elder, & Glaser, 1996). However, such assessments are expensive and time consuming and

Table 1
Measurement of Problem Solving

Issues	Approach
Definition	"Problem solving is cognitive processing directed at achieving a goal when no solution method is obvious to the problem solver." (Mayer, 1992; Mayer & Wittrock, 1996)
Task area	Troubleshooting (e.g., bicycle tire pump, respiratory system)
Facets of definition	
Content understanding	Measure with knowledge map (concept map)
Problem-solving strategies	Measure with prompt to elicit troubleshooting responses
Self-regulation	Measure with trait metacognition scale (planning, self-monitoring) Measure with trait motivation scale (effort, self-efficacy)
Rubrics	
Content understanding	Score knowledge (concept) map with existing CRESST software —e.g., content score, number of concepts, links
Problem-solving strategies	Rubric for an explanation list based on Mayer's (1992) framework—i.e., compare agreement with expert list of problems / solutions
Self-regulation	Rubric exists (see O'Neil & Herl, 1998)

result in delayed (up to months / years) reporting to parents, students, and teachers. In our work, we have taken a different approach to measuring exploratory problem-solving processes and outcomes. Instead of conducting interviews, naturalistic observation or think-aloud protocols while students problem solve, our prior R&D in problem solving focused on the feasibility of using computer-based technology for assessment of student problem solving. For example, for problem-solving strategies and measurement we rely on server access logs that automatically record all information-seeking and retrieval behaviors the student engages in. Expert problem-solving performance was collected and analyzed to infer the criteria used for scoring students' knowledge maps. Our approach is to computerize the administration, scoring, and reporting of problem solving, thus facilitating timely reporting and potentially increasing reliability and validity (Herl, O'Neil, Chung, & Dennis, 1997; Martinez, 1993; O'Neil, Chung, & Brown, 1997; Schacter et al., 1997).

In our prior computer-based research, content knowledge was assessed by knowledge maps. Domain-specific problem-solving strategies were measured on a search task by looking both at search behavior and at how newly found information was utilized. Self-regulation (metacognition and motivation) was assessed by a questionnaire. The basic design involved having the student (a) create a knowledge map, (b) receive feedback on it, and then (c) using a simulated Web site, search for information to improve it, and (d) construct a final knowledge map. The final knowledge map served as the outcome content understanding measure. Finally, self-regulation (metacognition and motivation) was assessed by a paper-and-pencil instrument.

We believe that content understanding and problem-solving strategies are best assessed domain-specifically whereas metacognition and motivation are best assessed as domain-independent constructs. We have created measures of trait metacognition and motivation that can be administered in 10 minutes. Using constructs from state-trait anxiety theory (Spielberger, 1975) as an analogy, we have formulated a set of self-report, domain-independent trait and state measures of metacognition and motivation. We find the state versus trait distinction useful for both cognitive and affective measurement. Thus, we have generalized the key constructs from an affective domain (e.g., state and trait anxiety) to a cognitive domain (e.g., state and trait metacognition).

However, our current computer-based approach to measurement of problem solving is not feasible for the International Life Skills Survey. Thus, a paper-and-pencil version was recommended.

Paper-and-Pencil Measurement of Problem Solving

Our approach to the measurement of problem solving for the International Life Skills Survey was the following: For content understanding, we asked participants to represent a mechanical system (e.g., a bicycle tire pump) by creating a paper-and-pencil knowledge map that was scored with software. This item would constitute the content understanding measure (see Mayer, 1997, for detail on the tire pump task). For problem-solving strategies, we asked participants to respond with a list to a prompt regarding domain-specific problem-solving process (e.g., troubleshooting: Imagine that the tire pump does not pump air to the hose. What could be wrong?). For self-regulation, we administered a trait self-regulation scale (see O'Neil & Herl, 1998). These items would be summed to create multiple scores, for metacognition (planning, self-monitoring) and motivation (effort, self-efficacy). Problem solving would be a profile of three scores (content understanding, problem-solving strategies, and self-regulation).

Construct Validity Studies for Problem Solving

Critical Issues

Our problem-solving conceptual framework represents a synthesis of research-based ideas, but the measurement of problem solving in this context is breaking new ground. Thus, there is limited reliability / validity information. Data collection was designed in a generalizability theory format (e.g., Shavelson & Webb, 1991).

A major conceptual issue was how critical should prior knowledge be in the assessment of problem solving. Our definition explicitly makes the measurement of such knowledge via a knowledge map an essential component of problem-solving assessment. Thus, if one does not know about the content (e.g., tire pumps), then one will not be able to exhibit problem-solving strategies in response to a prompt (e.g., troubleshoot the tire pump). Most definitions of problem solving in knowledge-rich tasks (e.g., not games) assume such knowledge. However, an alternative technical approach is to provide some level of information regarding the content (e.g., a schematic of a tire pump; see Figure 2) and then assess the responses

Bicycle Tire Pump

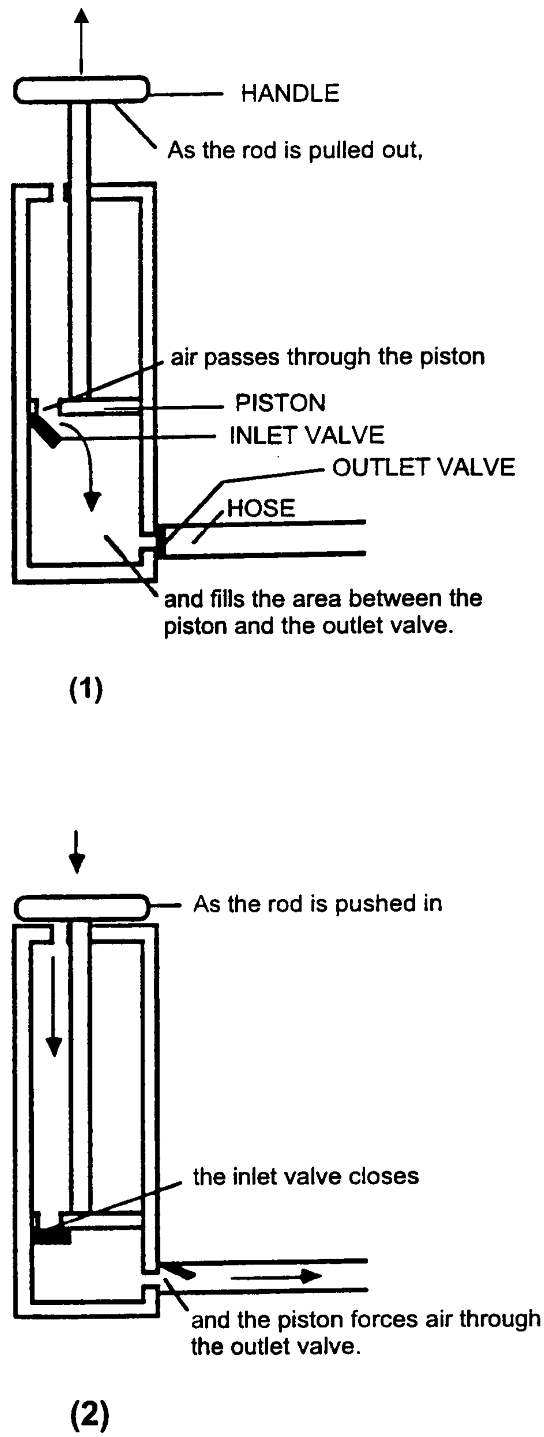


Figure 2. Schematic of a tire pump.

to the troubleshooting questions. One would lose information on the representation of the problem and the degree of content understanding possessed by the participants. However, one would gain more knowledge from troubleshooting strategies as some level of content would allow most students to respond to the prompt.

An additional issue is whether to provide participants with knowledge of how knowledge maps (our measure of content understanding) are scored—that is, should we provide test-taking strategies? For example, since knowledge maps will be new to most participants, should we tell them how such maps are scored? We were also interested in whether there would be any gender effects.

In summary, as may be seen in Table 2, the major questions of this research involve reliability issues as well as the impact of gender and whether providing scoring information influences performance. A critical question is how many topics and questions to administer to be reliable. Finally, there are two control issues that should be addressed: (a) Is there any effect of order? and (b) How shall procedural content knowledge be represented, that is, as one state or two states? We have collected data for both English-speaking and Chinese-speaking participants.

Table 2

Problem Solving: Assessment Questions

Questions
How many topics do we provide to be reliable?
How many problem-solving strategy questions do we provide to be reliable?
What is the impact of providing scoring information?
What is the role of gender?
How difficult are the topics?
Control issues
What is the effect of counterbalancing?
Should we use a one-state or two-state model?

Method

Design

Two pilot studies were conducted from December 1997 to February 1998, with different samples, in order to try out and refine the measures and procedures for the main study and to examine whether there were order effects that needed to be accommodated in the main study. Data for the first pilot study were collected at a high school in Taiwan in December 1997; data for the second pilot study were collected in January 1998 at a university in southern California. The main study was conducted at CRESST during March, April, and May of 1998.

For the pilot studies, a brief discussion is presented in the Methods, Administration Procedures, Scoring Procedures, and Analyses sections of this report; a more thorough presentation is provided for main study results and discussion.

Pilot Studies

Designs for the pilot studies targeted analyses of content order (counterbalancing order of tire pump and respiratory system tasks), scoring instructions (presence and absence of information pertaining to how knowledge maps would be scored), and the single vs. binary state model (states represented either as single entities or as one part of a dual state component). We were interested in content order (i.e., tire pump task given first, then the respiratory task versus respiratory task first, followed by the tire pump task); because knowledge mapping would be a relatively novel task for most participants, one might expect better performance on the second task due to practice in knowledge mapping independent of content. Such order effects are not desirable in testing. We manipulated scoring instructions for two reasons: It is common in some countries (e.g., in European countries) to be explicit about how tests are scored. It is seen as an ethical issue. Further, given the novelty of knowledge maps, few students would know how they are scored as compared to multiple-choice tests. Thus, we were interested in whether the provision of scoring instructions would improve performance.

The final design issue—single versus binary process model—refers to a knowledge mapping representation of procedural knowledge issue. When one represents the functioning of a device (e.g., a switch), the switch can be conceptualized as having two states: It is either *on* or *off*. Further, any analogous information can also be represented in this fashion (i.e., digitally). However, in our knowledge mapping technique, the switch can be represented in two different ways (see Figure 3). It can be represented as two distinct objects, that is, switch-on and switch-off; or it can be represented as one object with two states, on and off. Our preference was for the two-state version, but we were interested in whether this variable (one state vs. two states) would affect learning.

Pilot Study Participants

High school in Taiwan. The first pilot study was an initial attempt at assessing the feasibility of large-scale administration of paper-and-pencil problem-solving tasks, including knowledge mapping and problem-solving strategy questions. In December 1997, 143 high school students (50.4% males and 49.6% females) in Taiwan were selected to participate in the first pilot study on problem solving.

University in southern California. In January 1998, 41 undergraduate students were selected from a local university in the southern California area to participate in the second pilot study. Approximately 61% of the participants were females, and 39% were males.

Main Study

In February 1998, prior to the main study, CRESST personnel contacted a local temporary agency, requested a group of 9 participants (4 male and 5 female) for a specific date and time and prepared facilities to serve as a "friendly" environment in which adults could perform our problem-solving tasks.

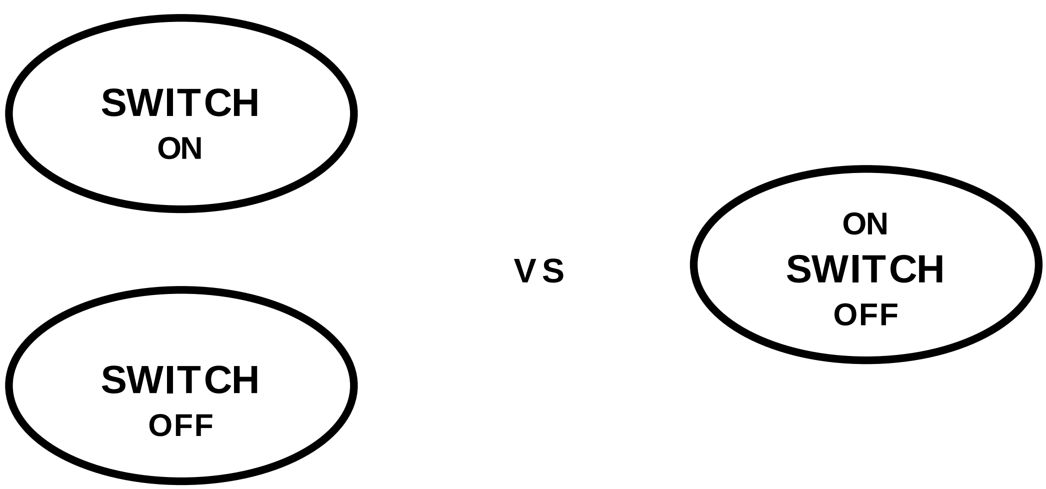


Figure 3. Two ways of representing the functioning of items.

Beginning in March 1998, adult participants were recruited from two temporary employment agencies in Los Angeles, CA. Most participants were obtained using the agencies' databases of temporary employees, while the rest were recruited from agencies' placement of ads in local newspapers.

Efforts were made to obtain as diverse as possible a sample of the adult population. Three important criteria for sampling were age, gender, and ethnicity. Tables 3, 4, and 5 present frequency data for age, education, and gender, respectively. Table 3 shows that the distribution of age is fairly uniform across experimental groupings, except for the 56-65 years category, where 9.9% of the sample was represented. Due to the random group assignment method employed, a smaller than expected proportion of participants ages 16-25 years were assigned to experimental group 1 (5.3%), whereas a larger than expected proportion of that age level were randomly assigned to experimental group 2 (35%). Chi-square analysis ($\chi^2 = 19.5$, $df = 20$) revealed no significant differences ($p > .05$) for age levels across the six experimental groupings.

There was some variability in the level of educational attainment by participants (Table 4), where over 40% had achieved either bachelor's or graduate degrees. Educational attainment levels were spread across the experimental groups fairly well, with the lone exception being experimental group 4, where only 20% possessed bachelor's or graduate degrees. Analyses revealed a nonsignificant difference across experimental groups based on the criterion of having at least a bachelor's degree.

Table 3

Percentages at Each Age Level by Experimental Grouping

	Experimental group						
Age	1	2	3	4	5	6	Total
16-25	5.3	35.0	9.5	25.0	26.3	18.8	19.8
26-35	21.1	30.0	33.3	37.5	26.3	25.0	28.8
36-45	42.1	15.0	14.3	18.8	26.3	43.8	26.1
46-55	15.8	10.0	28.6	12.5	10.5	12.5	15.3
56-65	15.8	10.0	14.3	6.3	10.5	0.0	9.9

Table 4

Percentages at Each Educational Level by Experimental Grouping

	Experimental group						
Education	1	2	3	4	5	6	Total
High school	33.3	52.6	27.8	60.0	35.3	33.3	40.4
AA	16.7	10.5	16.7	20.0	11.8	8.3	14.1
Bachelors	38.9	21.1	44.4	13.3	41.2	33.3	32.3
MS, MA, PhD	5.6	10.5	5.6	6.7	5.9	25.0	9.1
Other	5.6	5.3	5.6	0.0	5.9	0.0	4.0

Table 5 shows that more females participated in the study than males. Chi-square analysis (χ^2 = 7.4, df = 5) revealed no significant differences (p > .05) for gender across the six experimental groupings.

Table 6 presents the job titles that participants reported having held in the last five years of their work experiences. The majority of the participants worked in the area of administrative support services, where 58% reported having served in that capacity at some time in the last five years. Respondents also reported that they had almost 16 years of full-time work experience during their careers, and a little over 4 years of part-time experience. Although we did not collect this information as it pertains to the last five years, we would assume that the majority of the part-time work has come more recently, since all of our participants were recruited from temporary agencies.

Table 5

Percentages of Males and Females by Experimental Grouping

	Experimental group						
Gender	1	2	3	4	5	6	Total
Male	57.9	40.0	47.6	26.3	56.3	25.0	42.3
Female	42.1	60.0	52.4	73.7	43.8	75.0	57.7

Table 6

Percentages of Participants' Work Experiences in the Last 5 years

Job title	%
Administrator	23
Professional	35
Technician	10
Protective services	10
Administrative support	58
Skilled craft worker	12
Service maintenance	16

Note. Percentages do not add to 100 due to multiple category selection by participants.

The sample was ethnically diverse, although it was not meant to be representative of southern California or United States populations. Thirty-six percent of participants in the sample were Caucasian, 39% were African American, 14% were Latino, 4% were Asian, and 7% were of other ethnic origins.

Measures

There were several different types of tasks administered for this study, including background information and self-regulation questionnaires, paper folding tasks, content understanding (knowledge mapping) tasks, and problem-solving (troubleshooting and design prompts) tasks. The nature of participants' tasks varied either slightly or drastically depending on group assignment.

Background and preliminary questionnaires. CRESST researchers presented an overview of the study approximately two and a half minutes in duration to participants at each session. After this introduction, participants completed a background questionnaire (Appendix A), a self-regulation trait questionnaire (Appendix B), and a paper folding test (Appendix C).

Participants required a mean time of a little more than three minutes to complete the background questionnaire. This questionnaire contained questions regarding gender, age, ethnicity, education (including major and minor areas of study, certification, credentialing, and licensing), and numbers of years of part-time and full-time work experience (including various job titles they may have held).

Using constructs from state-trait anxiety theory (Spielberger, 1975) as an analogy, we have formulated a set of self-report, domain-independent trait and state measures of metacognition and motivation. The self-regulation trait questionnaire (O'Neil & Herl, 1998) used in this study contains 32 Likert-scale questions and provides trait measures of metacognition (planning, self-monitoring) and motivation (effort, self-efficacy). The metacognition (planning and self-monitoring) and effort items were written by our research group. The self-efficacy items were adapted from Pintrich and De Groot's (1990) intrinsic motivation scale (with Dr. Pintrich's permission). Students indicate how they generally think or feel by responding to a 4-point Likert scale, ranging from *almost never*, to *sometimes, often*, and *almost always*. We define the constructs as follow: *Planning*: One must have a goal (either assigned or self-directed) and a plan to achieve the goal. *Self-monitoring or self-checking*: One needs a self-checking mechanism to monitor goal achievement. *Self-efficacy* is one's confidence in being able to accomplish a particular task. *Effort* is the extent to which one works hard on a task. In multiple studies, the data indicate acceptable reliability and validity (O'Neil & Herl, 1998). Participants needed approximately six minutes on average to complete this measure.

The paper folding test (Ekstrom, French, Harman, & Derman 1976) administered in this study is a spatial ability test containing 10 problems, and was the first timed task that participants were asked to complete. Participants were given two minutes to read the instructions and three minutes to complete the paper folding test.

Content understanding. Knowledge mapping tasks were designed to measure one of five learning areas defined by CRESST's model of learning (Baker, 1995), namely content understanding. For the research presented in this report, conceptual and procedural content understanding are represented in the form of knowledge mapping (Baker et al., 1994; Baker & Niemi, 1991; Baker, Niemi, Gearhart, & Herman, 1990; Baker, Niemi, Novak, & Herl, 1992; Dansereau & Holley, 1982; Herl, Baker, & Niemi, 1996; Holley & Dansereau, 1984; Jonassen, 1996; Lambiotte, Dansereau, Cross, & Reynolds, 1989). Knowledge maps (or concept maps) are comprised of nodes and links, with nodes representing concepts or their attributes, and links expressing the semantic relations among those concepts. Our research indicates that knowledge maps correlate about .70 with essay tests on the same topic (Herl, 1995). A closed map construction system contains finite sets of nodes and links, from which constructors must choose in order to construct their maps, and it provides the model for the construction system used for these studies. The knowledge mapping tasks ask examinees to map how the parts of a system are related.

Content understanding was measured by comparing participants' knowledge maps to experts' maps (Herl, O'Neil, et al., 1996). Knowledge mapping tasks were constructed for two different kinds of systems: bicycle tire pumps and the human respiratory system, both of which are types of pressure systems. These knowledge mapping tasks were designed to measure prior content knowledge for each of these systems.

Each system's mapping task had six main components and three links for representing relationships among those components. The components of the tire pump were cylinder, handle, hose, inlet valve, outlet valve, and piston. The components of the human respiratory system were air sacs, bloodstream, chest cavity, diaphragm, lungs, and rib muscles. These components were characterized to have a binary state design to represent their functions. For example, for the tire pump knowledge map task, handles and pistons were represented as being either *up* or *down*; inlet and outlet valves were either *open* or *closed*; cylinders were either *high pressure* or *low pressure*; and hoses had *airflow* or *no airflow*. For the human respiratory task, air sacs had *oxygen in* or *carbon dioxide out*; the bloodstream had either *increased oxygen* or *decreased carbon dioxide*; chest cavities were either *expanded* or *reduced*; diaphragms were either *up* or *down*; lungs were either *expanded* or *contracted*; and rib muscles were either *relaxed* or *contracted*.

Initially, knowledge mapping tasks were designed to fit our existing models of knowledge mapping, where a "term," or concept, is represented by a single entity. Even though the systems were designed to contain binary state entities, a 1-state design was also included in the early pilot studies. The 2-state design definitely fit the particular nature of these systems, and results from pilot study data analyses supported its eventual inclusion in the main study design. Appendices D and E contain the mapping task label sheets for the 1-state and 2-state bicycle tire pump and human respiratory system tasks. Appendix F contains instruction sheets for the knowledge mapping tasks that were used in the main study. The first four pages of Appendix F comprise instructions for experimental groups mapping onto diagrams of the systems; the last four pages of this appendix were used for groups mapping onto blank sheets.

The inclusion of a 2-state mapping task allowed CRESST researchers to design a new measure for the main study. Reproductions of the diagrams used for the problem-solving questions (see Problem-Solving Tasks in this section) were employed to provide a measure of system component identification in the knowledge mapping task. The texts in the diagrams were removed and only the picture elements were overlaid onto the mapping sheet. Map constructors were asked first to look at the mapping page and correctly identify each part of the system, and then to place each component label on the mapping page in order to correctly identify each part of the system. There were no distractors captioned in the diagrams; therefore, each system's diagram had six possible areas where participants could place labels. Appendix G contains knowledge mapping sheets with tire pump and human respiratory system diagrams, respectively. Participants in experimental groups 1 and 2 were asked to perform the same knowledge mapping task as those in groups 3 and 4, except they mapped onto pictures of a tire pump or the human respiratory system.

Knowledge Map Construction Training Sessions

Knowledge map training sessions were modeled after those of previous studies (Baker et al., 1990; Baker & Niemi, 1991; Herl, 1995; Herl, O'Neil, et al., 1996) employing knowledge mapping tasks. The training sessions emphasized both the conceptual and procedural skills necessary for constructing knowledge maps. Each group of participants received a sheet of mapping labels containing both terms and links for the training and actual tasks. Training sessions were tailored to the different types of experimental groups. Participants were asked to read silently a sheet containing written instructions as the administrator read the instructions aloud to the group.

Instructions and a demonstration were given to participants concerning the method for constructing knowledge maps. After the instructions were read, a brief demonstration was conducted using a simple cause-and-effect model (Appendix H), so that the group could participate in developing a knowledge map. Participants were allowed to construct several links using the demonstration labels on their mapping task sheets.

After the 10-minute hands-on training session was completed, participants were allowed to begin constructing their knowledge maps. Ten minutes were given to participants in order to construct knowledge maps. Considering it was the first time that any of these individuals had constructed a knowledge map, up to three minutes of extra time was allotted. For the second knowledge mapping task, very little instructional time was necessary (less than a minute to answer one or two questions), and participants completed the second map in an average of 10 min 44 sec. Recording of actual start and stop times by CRESST researchers also enabled us to reevaluate estimated times and project administration times for future studies.

Problem-Solving Strategy Questions

The problem-solving strategy task consisted of a diagram of a bicycle tire pump (Appendix I) or the human respiratory system (Appendix J) annotated with text, accompanied by two short-answer troubleshooting questions, and one short-answer design question for a particular system. These questions were used, or were similar to those used, in previous research (Mayer & Gallini, 1990, Mayer & Sims, 1994). Appendix K contains the three troubleshooting and design questions for the bicycle tire pump task, and Appendix L contains the three troubleshooting and design questions for the human respiratory task.

Participants were given two and a half minutes to examine the diagram. For each question, participants were given two and a half minutes to read the question and write down as many answers as they could think of to answer that particular question. One of the bicycle tire pump troubleshooting questions was “Suppose you push down and pull up on the handle several times but no air comes out of the bicycle pump. What could be wrong?” This type of question is designed to elicit all possible responses as to why no air comes out of the hose. A second troubleshooting question was included in the main study, and was similar in nature to the first troubleshooting question. In essence, participants were asked to pinpoint potential causes of faults in the system.

Design questions, on the other hand, required participants to present alternative designs to create more efficient systems. One of the tire pump design questions was “How can you increase the efficiency of the bicycle pump? That is, how can you move more air through the pump?” The respiratory system design question was “Suppose you are a scientist trying to improve the human respiratory system. How could you get more oxygen into the bloodstream faster?” Participants’ suggestions for the respiratory system design question turned out to be much more difficult to score than the suggestions for the tire pump task design question.

Administration Procedures

For all studies, participants completed the background questionnaire, self-regulation trait questionnaire, and paper folding test (collected as information for another CRESST study and reported elsewhere). Between 15 and 20 minutes were required for participants to complete these tasks. Because each pilot study was designed to address specific problem-solving issues, different designs were employed for content understanding and problem-solving questions.

High school in Taiwan. After the materials were constructed in English, they were translated into Chinese, and used in Taiwan for the first pilot study. This pilot study comprised a 2 x 2 x 2 x 2 experimental mixed model design, where effects of scoring instructions (see Appendix F for examples from the main study), number of states (one vs. two), and task order (first vs. second) were tested as between-subjects variables. The within-subjects part of the design tested for the effects of type of task (bicycle tire pump task vs. respiratory system task). System diagrams were not used with the knowledge mapping tasks for this study. Only two problem-solving questions (one troubleshooting and one design) were given to students during this study. A complete description of the method, procedures, and results can be found in Lee (1998).

University in southern California. The design for the second pilot study replicated the first pilot study on a smaller scale. Because interactions were not important in this design, the sample size was adequate for testing simple main effects.

Main study. For the main study, participants were randomly assigned to one of six experimental groups. Six to 12 people were tested at a time depending on participant availability. The design for these groups included two research criteria: presence or absence of scoring instructions, and presence or absence of diagrams for knowledge mapping tasks. Participants in experimental groups 5 and 6 were included in the design to obtain generalizability coefficients for problem-solving questions with and without accompanying illustrations. Table 7 displays a representation of the design showing which measures members of each experimental group completed.

Table 7

Order of Tasks by Experimental Group

Group	1	2	3	4			5	6			7
1	BQ	SR	PF	Map	Picture given	Scoring instructions given	DQQQ	Map	Picture given	Scoring instructions given	DQQQ
2	BQ	SR	PF	Map	Picture given	No scoring instructions given	DQQQ	Map	Picture given	No scoring instructions given	DQQQ
3	BQ	SR	PF	Map	No picture given	Scoring instructions given	DQQQ	Map	No picture given	Scoring instructions given	DQQQ
4	BQ	SR	PF	Map	No picture given	No scoring instructions given	DQQQ	Map	No picture given	No scoring instructions given	DQQQ
5	BQ	SR	PF	DQQQ			DQQQ	Map	Picture given	Scoring instructions given	
6	BQ	SR	PF	QQQ			QQQ				

Note. BQ = background questionnaire. SR = self-regulation trait questionnaire. PF = paper folding task. D = diagram for problem solving tasks. Q = problem-solving question.

Table 8 displays the estimated and actual administration times for each of the tasks in the study. Despite unfamiliarity with the knowledge mapping task on the part of most participants, the average time needed for training sessions was only slightly more than the estimated time. For the second mapping task, only a brief overview was required to refresh participants' memories as to the important parts of the task.

Table 8

Time Requirement for Activities

Activity	Estimated time (min)	Actual time
Background questionnaire	3	3 min 14 sec
Self-regulation trait questionnaire	8	5 min 49 sec
Paper folding test	5	5 min 08 sec
Knowledge mapping task training session	10	12 min 30 sec
Knowledge mapping task	10	12 min 52 sec
Problem-solving questions	10	10 min 05 sec

Scoring Procedures

Self-regulation trait questionnaire. The 32-question trait self-regulation questionnaire consists of four subscales (Appendix B). The responses are summed to provide subscale measures for planning, self-checking, effort, and self-efficacy. The range of possible scores for each subscale is 8-32. Planning and self-checking subscale scores can be summed to form a trait metacognition scale, and effort and self-efficacy subscale scores can be summed to form a trait motivation scale. The metacognition and motivation scales have a range of possible scores from 16-64. The scoring key for the trait self-regulation questionnaire is found in Appendix M.

Paper folding test. Participants received one point for each correct answer they circled on the paper folding test. The range of scores for this measure is 0-10.

Knowledge mapping tasks. There are three descriptive statistics associated with the knowledge maps presented here: (a) semantic content score, (b) number of terms used, and (c) number of links. Semantic content score was computed using the experts' knowledge maps as sources of criteria; number of terms used and number of links were calculated using the participants' maps. A term was considered to be used in a knowledge map when there was at least one link connected to it. The number of links is also a countable statistic and is defined as the number of links constructed.

Composite knowledge maps were constructed a priori for both the tire pump and human respiratory system by CRESST personnel. For the tire pump (Appendix N), 12 relationships were theorized to represent the two main processes of the system (pulling the handle up to start the first process, and pushing the handle down to start the second process). For the respiratory system (Appendix O), 10 relationships were theorized to represent three processes of respiration, namely, breathing in (diaphragm down to start this process), breathing out (diaphragm up to start this process), and exchange (involving the exchange of oxygen and carbon dioxide).

These two a priori knowledge maps were used to decide how the scoring instructions would be constructed. For each of the two systems, participants received scoring instructions informing them of the correct number of relationships, and that they should construct no more than that number. Unfortunately, there was a clerical error and slightly different scoring instructions were given for each map. The respiratory map had, as intended, the following additional sentence, "There is no penalty for guessing." The tire pump map scoring instructions did not have this sentence. Thus, the two maps differed unintentionally in scoring instructions. Ten of 77 (13%) maps constructed by participants receiving scoring instructions contained more than the maximum number of links. For each of these maps, the total number of relationships was randomly reduced so that all participants receiving scoring instructions would have tire pump and respiratory maps with at most 12 and 10 links, respectively.

During the main study data collection period, 21 journeymen with expertise in air conditioning system repair were asked to construct knowledge maps about bicycle tire pumps in order to test the hypothesis that 12 links could reasonably represent a model to explain the operation of a tire pump. Table 9 displays descriptive statistics for the main study and expert participants on the bicycle tire pump knowledge mapping task. Six of these experts' tire pump knowledge maps were not used because other responsibilities prevented these experts from finishing their maps. The table shows that the average number of terms used and the number

Table 9

Means and Standard Deviations for Knowledge Mapping Task Measures for Tire Pump System for Adults and Experts

Measure	Adults[a]	Experts[b]	Model[c]
Number of terms used			
M	10.41	12.00	12.00
SD	2.03	0.00	
Number of links			
M	8.39	12.23	12.00
SD	4.09	3.78	
Number of CAUSES links			
M	4.43	4.31	10.00
SD	3.02	2.50	
Number of CONTRIBUTES links			
M	2.64	7.00	2.00
SD	2.24	3.44	
Number of PRIOR TO links			
M	1.32	0.92	
SD	1.64	1.44	

[a]n = 76. [b]n = 15. [c]n = 1.

of links constructed are close to those of the model scoring rubric created by project personnel. One major difference is the average frequency of CAUSES links used by experts compared with our model. The experts (*M* = 4.31) did not use the CAUSES link nearly as often as we did in the model (10). Instead, experts (*M* = 7.00) used CONTRIBUTES much more often than we had anticipated (2).

We had theorized that the operation of a bicycle tire pump system contained more direct causal events than the experts indicated. Experts' maps showed that they thought influences (by using CONTRIBUTES) on system processes described the operation of the pump more accurately than direct causal (CAUSES) relationships. On average, experts (*M* = 0.92) constructed few PRIOR TO links, and 10 of the 15 experts did not construct any PRIOR TO links in their maps.

The 15 journeymen's tire pump maps were used as expert criteria to compute tire pump map scores for main study participants. The correlation between those map scores and map scores calculated using our a priori model was .90 (p = <.01). The model did not contain any PRIOR TO links because, for the most part, experts did not construct important links using PRIOR TO. For example, two experts' maps contained the INLET VALVE-closed prior to OUTLET VALVE-open link, but these terms were not seen as having a causal relationship by other experts. Based on this kind of analysis, we decided to use the a priori pump model to score tire pump knowledge maps. However, we were not able to collect respiratory system maps from experts, so we could not perform similar analyses for that topic. Because we decided to use the a priori pump model to score participants' tire pump maps, the choice to use the a priori respiratory system model was also adopted to score the participants' respiratory system maps.

We modified our a priori models to categorize CAUSES and CONTRIBUTES as one entity. One point was awarded if a link in the participant's map matched the link in the model's map. For example, HANDLE-up causes PISTON-up and HANDLE-up contributes PISTON-up links were awarded 1 point, but HANDLE-up prior to PISTON-up was scored as incorrect. This scoring method provided map scores for both systems with a range of 0-12.

Participants constructing knowledge maps onto system diagrams were given scores for correct identification of system components. One point was awarded for every component correctly identified, resulting in a range of scores from 0-6.

Problem-solving strategy tasks. The rubrics used to score participants' responses to problem-solving questions had been used in previous research (Mayer & Gallini, 1990, Mayer & Sims, 1994). Rubrics were constructed for the new troubleshooting questions added to the main study. A small selection (n = 10) of participants' answers was randomly sampled for each of the six problem-solving questions. Pairs of raters scored them, disagreements in scoring were discussed and resolved, and minor adjustments were made to the rubrics when necessary. All problem-solving question responses were subsequently scored by at least two raters. One point was awarded for each answer that agreed with possible answers in the rubrics. Appendix P contains the scoring rubrics for the bicycle tire pump and human respiratory system questions.

For each system, scores obtained by participants on the first and third questions were added together to form a troubleshooting scale. The second question for each of the two systems was used to form a one-item design scale.

Analyses

This section presents descriptive statistics for content understanding and problem-solving strategy tasks, and results for reliability analyses of each measure.

Descriptive Statistics

Knowledge mapping task. There are three descriptive statistics associated with knowledge maps reported here: (a) content understanding, (b) number of terms used, and (c) number of links. The first statistic was computed using a priori model maps as the source of scoring criteria; the second and third statistics were generated solely from student performance on the mapping task. For the knowledge mapping tasks used in this study, a term is defined as a component and its associated binary state (e.g., *handle up* and *handle down* are considered to be different terms). A term was considered to be used in the knowledge map when there was at least one link connected to it. Therefore, for both the tire pump and respiratory system tasks, there are 12 terms that can be used. The number of links is a countable statistic and is defined as the number of valid links constructed.

Pilot studies. Table 10 displays the descriptive statistics computed for the pilot study knowledge mapping tasks. Inspection of the means shows that university students scored higher on the knowledge mapping tasks than high school students, and this finding is certainly not a surprise.

Table 10

Means and Standard Deviations for Knowledge Mapping Task Measures for Tire Pump and Respiratory System for High School and University Students, Pilot Studies

	Tire pump		Respiratory system	
Measure	High school[a]	University[b]	High School[c]	University[d]
Content understanding				
M	3.14	4.52	2.09	3.10
SD	2.30	3.42	1.89	2.13
Number of terms used				
M	10.58	11.52	9.52	11.71
SD	2.90	1.21	1.54	1.13
Number of links				
M	9.16	10.81	8.45	10.74
SD	3.45	3.13	2.78	2.83

[a]*n* = 107. [b]*n* = 41. [c]*n* = 105. [d]*n* = 41.

As stated previously, one purpose of the pilot studies was to test specific hypotheses concerning format (1-state vs. 2-state) of the knowledge mapping tasks. In the Taiwanese high school sample, students performed significantly better on the 1-state map. In the university sample, 16 students scored higher on semantic content for both the 2-state tire pump task (*M* = 4.93) and the 2-state respiratory system task (*M* = 3.80) than 15 other university students constructing 1-state representations on the tire pump task (*M* = 4.13) and the respiratory system task (*M* = 2.44). Although these differences were not significant, trends in the analyses led us to believe that the 2-state model would provide a more intuitive context for U.S. students in which to construct knowledge maps about these systems.

Results were inconclusive when examining differences in knowledge map task performance based on task order and provision of scoring instruction criteria. On the second task, the Taiwanese high school students who did not receive scoring instructions performed better that those who did receive scoring instructions, but there were no significant differences on the first task. Groups of university students scored slightly higher on the tire pump knowledge mapping task and slightly lower on the respiratory system knowledge mapping task when scoring instructions were provided than their respective counterparts not receiving the scoring instructions. High school and university males performed at a higher level than high school and university females on both knowledge mapping tasks; however, these differences proved to be nonsignificant.

Generalizability analyses were performed for both pilot studies. Results showed that the generalizability coefficient for the two knowledge mapping tasks was lower for Taiwanese high school students (g = .34) than for university students (g = .65).

Main study. Groups 1 through 4 (see Table 7) were designed to contain the experimental conditions for the knowledge mapping tasks and therefore were used to test for differences concerning provision of scoring instructions, mapping onto system diagrams, and gender effects. Table 11 presents overall descriptive statistics for knowledge mapping tasks for groups 1 to 4.

Table 12 displays the descriptive statistics computed for the knowledge mapping task for groups either receiving scoring instructions or not receiving scoring instructions. It appears that providing scoring instructions to participants did not affect their performance on the knowledge mapping task. It should be noted that the overall performance levels are very low. Thus, these participants may not have had sufficient knowledge to use our provision of the scoring instructions to improve their performance. Examination of the statistics in Tables 11 and 12 shows that participants were able to use more terms and construct more links when constructing tire pump knowledge maps than when constructing respiratory system knowledge maps, regardless of whether scoring instructions were given or whether participants mapped onto a diagram. However, these additional links did not

Table 11

Means and Standard Deviations for Knowledge Mapping Task Measures for Tire Pump and Respiratory System

Measure	Tire pump[a]	Respiratory system[b]
Content understanding		
M	2.53	2.15
SD	2.28	1.98
Number of terms used		
M	10.41	10.00
SD	2.03	2.61
Number of links		
M	8.39	6.79
SD	4.09	4.03

[a]n = 76. [b]n = 73.

Table 12

Means and Standard Deviations for Knowledge Mapping Task Measures for Tire Pump and Respiratory System by Scoring Instructions Provided

Measure	Tire pump: Scoring instructions[a]	Tire pump: No scoring instructions[b]	Respiratory system: Scoring instructions[c]	Respiratory system: No scoring instructions[d]
Content understanding				
M	2.53	2.53	2.43	1.86
SD	2.17	2.42	2.42	1.36
Number of terms used				
M	10.12	10.72	9.86	10.14
SD	2.03	2.01	2.20	3.00
Number of links				
M	8.50	8.28	6.65	6.94
SD	4.10	4.12	2.43	3.82

[a]n = 40. [b]n = 36. [c]n = 40. [d]n = 36.

produce higher content scores for either tire pump or respiratory system tasks when scoring instructions were given. Respiratory map constructors receiving scoring instructions achieved higher content scores ($M = 2.43$) than those not receiving scoring instructions ($M = 1.86$); however, this difference is not significant ($p > .05$).

Analyses of descriptive statistics found in Table 13 revealed that there are differences when considering whether or not map constructors constructed their relationships onto map sheets containing a diagram of the system. Participants mapping onto diagrams of tire pumps ($M = 3.00$) scored significantly higher ($p < .05$) than those mapping onto blank mapping sheets ($M = 2.03$), regardless of whether they received scoring instructions or not. It is very interesting that tire pump map constructors mapping onto a diagram actually constructed fewer links on average ($M = 7.90$) than those not receiving the diagram ($M = 8.92$), yet the diagram group achieved significantly higher content scores. Interestingly, participants mapping onto a picture of the human respiratory system ($M = 2.05$) scored lower than those mapping onto a blank sheet, although the difference is nonsignificant ($p = .68$).

Appendices Q and R contain examples of high-scoring knowledge maps constructed by main study participants for the bicycle tire pump and respiratory system, respectively.

Table 13

Means and Standard Deviations for Knowledge Mapping Task for Tire Pump and Respiratory System by Diagram or No Diagram

Measure	Tire pump: Diagram provided[a]	Tire pump: No diagram provided[b]	Respiratory system: Diagram provided[c]	Respiratory system: No diagram provided[d]
Content understanding				
M	3.00	2.03	2.05	2.27
SD	2.43	2.02	2.16	1.75
Number of terms used				
M	10.33	10.49	9.43	10.70
SD	1.98	2.10	2.88	2.07
Number of links				
M	7.90	8.92	6.45	7.21
SD	3.76	4.39	4.41	3.53

[a]n = 39. [b]n = 37. [c]n = 39. [d]n = 37.

Table 14 presents the descriptive statistics for knowledge mapping tasks for males and females. Surprisingly, there were no gender effects found at all for the bicycle tire pump content understanding measure. There were some significant

Table 14

Means and Standard Deviations for Knowledge Mapping Task for Tire Pump and Respiratory System by Gender

Measure	Tire pump: Males[a]	Tire pump: Females[b]	Respiratory system: Males[c]	Respiratory system: Females[d]
Content understanding				
M	2.97	2.19	2.84	1.63
SD	2.36	2.17	2.57	1.13
Number of terms used				
M	10.79	10.12	10.13	9.90
SD	1.71	2.22	2.78	2.53
Number of links				
M	9.33	7.67	8.00	5.85
SD	5.18	2.86	4.68	3.24

[a]n = 33. [b]n = 43. [c]n = 33. [d]n = 43.

effects for the respiratory system knowledge mapping tasks. Although males and females used approximately the same number of terms in their maps, females (*M* = 5.85) constructed significantly ($p < .05$) fewer links than males (*M* = 8.00).

Table 15 presents the mean number of components identified correctly on the knowledge mapping tasks for the tire pump and the respiratory system. Analyses showed that participants identified bicycle tire pump components better than they identified respiratory system components. Overall, participants correctly identified 4.70 tire pump components and 3.97 respiratory system components out of 6. The difference between males (*M* = 5.11) and females (*M* = 4.32) was nonsignificant for the bicycle tire pump task. The difference between males (*M* = 4.00) and females (*M* = 3.95) was negligible for the respiratory system task.

Problem-solving strategy questions. Due to time constraints, only two problem-solving questions were administered for the pilot studies—one troubleshooting question and one design question. For the main study, a second troubleshooting question was added; therefore, troubleshooting was measured by two questions instead of one.

Pilot studies. Table 16 displays descriptive statistics for troubleshooting and design problem-solving questions for both pilot studies. Once again, the trends for scores on problem-solving questions are the same as they are for knowledge mapping scores. Table 16 also shows that university students scored higher on both troubleshooting and design measures, which again is not surprising.

It was surprising to find that the generalizability coefficient for problem-solving questions was much lower for university students (g = .09) than for Taiwanese high school students (g = .57). There is no clear explanation as to why the

Table 15

Means and Standard Deviations for Number of Components Correctly Identified for Tire Pump and Respiratory System by Gender

	Tire pump		Respiratory system	
Measure	Males[a]	Females[b]	Males[c]	Females[d]
M	5.11	4.32	4.00	3.95
SD	1.45	1.77	1.91	1.90

[a]*n* = 18. [b]*n* = 19. [c]*n* = 19. [d]*n* = 19.

Table 16

Means and Standard Deviations for Knowledge Mapping Task Measures for Tire Pump and Respiratory System for High School and University Students

Measure	Tire pump		Respiratory system	
	High school[a]	University[b]	High school[c]	University[d]
Troubleshooting				
M	1.53	2.42	1.68	2.61
SD	1.12	1.24	1.03	1.26
Design				
M	1.10	2.02	1.43	2.22
SD	0.72	0.90	0.91	1.11

[a]$n = 107$. [b]$n = 41$. [c]$n = 105$. [d]$n = 41$.

generalizability coefficient was so low for that sample. Examination of the two topics separately revealed correlations of .37 and .71 for the two tire pump and two respiratory system questions, respectively. However, analyses showed that scores for tire pump questions and respiratory system questions did not correlate significantly ($p > .05$). It appears that for university students, the problem-solving questions for the two tasks measured completely different things.

Main study. Table 17 displays descriptive statistics for troubleshooting and design problem-solving questions for main study participants. Table 17 shows that averages for the second troubleshooting questions for both the bicycle tire pump task and respiratory system task were much lower than those for the first troubleshooting questions (which were the only troubleshooting questions used in pilot studies). Possible explanations for the lower performance on these questions are that (a) participants experienced more fatigue when responding to more questions, and (b) the second troubleshooting questions are much more difficult than the first troubleshooting questions.

It was hypothesized that the conditions under which participants were asked to construct knowledge maps would have little or no effect on participants' performances on the problem-solving questions. Analyses supported this hypothesis, revealing no significant differences between knowledge map testing conditions on results for problem-solving questions. This finding is important because analysis would be problematic if the knowledge mapping experimental groupings contaminated the reliability of problem-solving strategy questions.

Table 17

Means and Standard Deviations for Problem Solving Questions for Tire Pump and Respiratory System

Measure	Tire pump[a]	Respiratory system[b]
Troubleshooting Q1		
M	1.37	1.18
SD	1.13	0.99
Troubleshooting Q2		
M	0.64	0.56
SD	0.75	0.80
Design		
M	1.04	0.27
SD	0.93	0.40

[a]n = 76. [b]n = 76.

Participants in experimental groups 5 and 6 (see Table 7) completed their problem-solving strategy questions before constructing any knowledge maps. Therefore, map performance did not have any effect on problem-solving responses. Designs for groups 5 and 6 were employed to test for differences in problem-solving performance in the absence of diagrams. One group (group 5) was allowed to study the diagrams prior to answering problem-solving questions, whereas the other group (group 6) was not.

It was hypothesized that participants studying the diagram for two and a half minutes would perform better on the problem-solving questions than those who did not. Indeed, participants studying diagrams listed more correct answers to both the troubleshooting and design questions for both the tire pump and the respiratory system (see Table 18). Because participants sampled for this study would not be classified as experts, it is not surprising that participants receiving the diagram scored higher on the problem-solving questions than those not receiving the diagram. However, only some of differences were significant. Participants who studied the respiratory system diagram scored significantly higher ($p < .05$) on the respiratory system troubleshooting measure (M = 1.80) than those who did not (M = 0.97). The tire pump diagram group scored significantly higher ($p < .05$) on the design question (M = 1.08) than the nondiagram group (M = 0.50). Other differences were nonsignificant.

Table 18

Means and Standard Deviations for Problem Solving Questions for Tire Pump and Respiratory System by Diagram or No Diagram Provided

	Tire pump		Respiratory system	
Measure	Diagram provided[a]	No diagram provided[b]	Diagram provided[c]	No diagram provided[d]
Troubleshooting				
M	1.55	0.78	1.80	0.97
SD	1.55	0.81	1.45	0.76
Design				
M	1.08	0.50	0.78	0.69
SD	1.04	0.55	0.80	0.66

[a]*n* = 20. [b]*n* = 16. [c]*n* = 20. [d]*n* = 16.

Table 19 presents the descriptive statistics for problem-solving strategy questions for tire pump and respiratory system tasks for males and females completing knowledge mapping tasks prior to answering the questions. The only significant gender difference was found in the design question, where males (M = 1.29) scored significantly higher ($p < .05$) than females (M = 0.85).

Table 19

Means and Standard Deviations for Problem-Solving Questions for Tire Pump and Respiratory System by Gender

	Tire pump		Respiratory system	
Measure	Males[a]	Females[b]	Males[c]	Females[d]
Troubleshooting				
M	2.33	1.77	1.56	1.36
SD	1.57	1.55	1.10	1.15
Design				
M	1.29	0.85	0.61	0.52
SD	0.98	0.85	0.73	0.86

[a]*n* = 33. [b]*n* = 43. [c]*n* = 33. [d]*n* = 43.

Reliability and Generalizability

Results are reported for the knowledge mapping and problem-solving strategy tasks. Generalizability coefficients for the knowledge mapping and problem-solving strategy questions were in the .60-.70 range. Because the problem-solving strategy questions were scored by multiple raters, alpha reliabilities are presented for both the tire pump and respiratory system tasks.

Knowledge mapping tasks. Generalizability coefficients using the content understanding scores from the tire pump and human respiratory system tasks are consistent for the university pilot study (g = .65) and the main study (g = .65). The generalizability coefficient for the Taiwanese high school pilot study (g = .34) was low relative to the other two studies. The knowledge mapping tasks require the greatest amount of verbal communication between the administrator and the participants. Language differences concerning translation of our written instruction from English to Chinese may have contributed to the low coefficient obtained during the first pilot study. Instructions were oral for the second pilot study and the main study.

Problem-solving strategy tasks. The troubleshooting and design questions were scored by pairs of raters using modified rubrics. Results show that interrater reliabilities ranged from .62 (respiratory system troubleshooting question 2) to .86 (tire pump design question 1). A rule of thumb is that reliabilities of .80 or above are generally acceptable. However, it is a concern that the reliability for the second respiratory system troubleshooting question is so much lower than the others.

In addition to the mean differences found for the tire pump and respiratory system troubleshooting and design questions, use of diagrams had important effects on the reliabilities of those systems. Cronbach alpha reliabilities for the tire pump and respiratory system tasks were .73 and .58, respectively, when participants received diagrams prior to responding to the prompts, and .53 and .36, respectively, when they did not receive diagrams prior to responding to the prompts. One may infer that less knowledge leads to less reliability.

The generalizability coefficient for the problem-solving strategy questions for both the tire pump and respiratory system tasks was .66 in the main study.

Latent-variable modeling. It has been suggested that latent-variable modeling is an appropriate alternative for estimating generalizability (Abedi & Baker, 1995). Abedi and Baker (1995) reported that the application of latent-variable models resulted in significant improvement for topic generalizability. In this study, an estimate of generalizability was obtained through the use of confirmatory factor analysis/latent-variable modeling as implemented by EQS (Bentler, 1992). In this model, problem solving was defined as three factors representing content understanding, domain-dependent problem-solving strategies, and self-regulation. Two content understanding measures (tire pump and human respiratory system map scores), four problem-solving strategy questions (tire pump and human respiratory system troubleshooting and design questions), and four self-regulation subscale measures (effort, self-efficacy, planning, and self-monitoring) were used as observed variables.

A chi square of 73.10 with 51 degrees of freedom and a chi-square ratio of 1.43 indicated a good fit for this model. Rules of thumb for these measures would indicate a good fit if the χ^2 was nonsignificant and the chi-square ratio was less than 3. Similarly, the Bentler's Normed-Fit Index (NFI) (Bentler & Bonett, 1980) was .71, the Non-Normed Fit Index (NNFI) (Bentler & Bonett, 1980) was .85 and the Comparative Fit Index (CFI) (Bentler, 1992) was .88, which all were indications of a relatively poor fit to the data. Rules of thumb for these measures would indicate a good fit of .90 or better. Factor loadings for content understanding were both about 0.5, and ranged from 0.3 to 0.8 for problem-solving strategies, and from 0.5 to 0.9 for self-regulation. We hoped that the three latent variables would all be positively correlated. It can be seen from Figure 4 that the content understanding and problem-solving strategies latent variables are correlated 0.84, which is high but not unexpected. We expected that knowledge of content would facilitate problem-solving strategies. However, the correlations among those same latent variables and the self-regulation latent variable are negative or zero, which was unexpected. We had expected high positive correlations. These data, taken collectively, indicate that the problem-solving measures should be reported as a profile of three scores and not collapsed into a single score, as the latent-variable correlations do not justify a single score.[1]

---

[1]The authors wish to thank Dr. Jamal Abedi, CRESST/UCLA, for his assistance on the latent variable analyses.

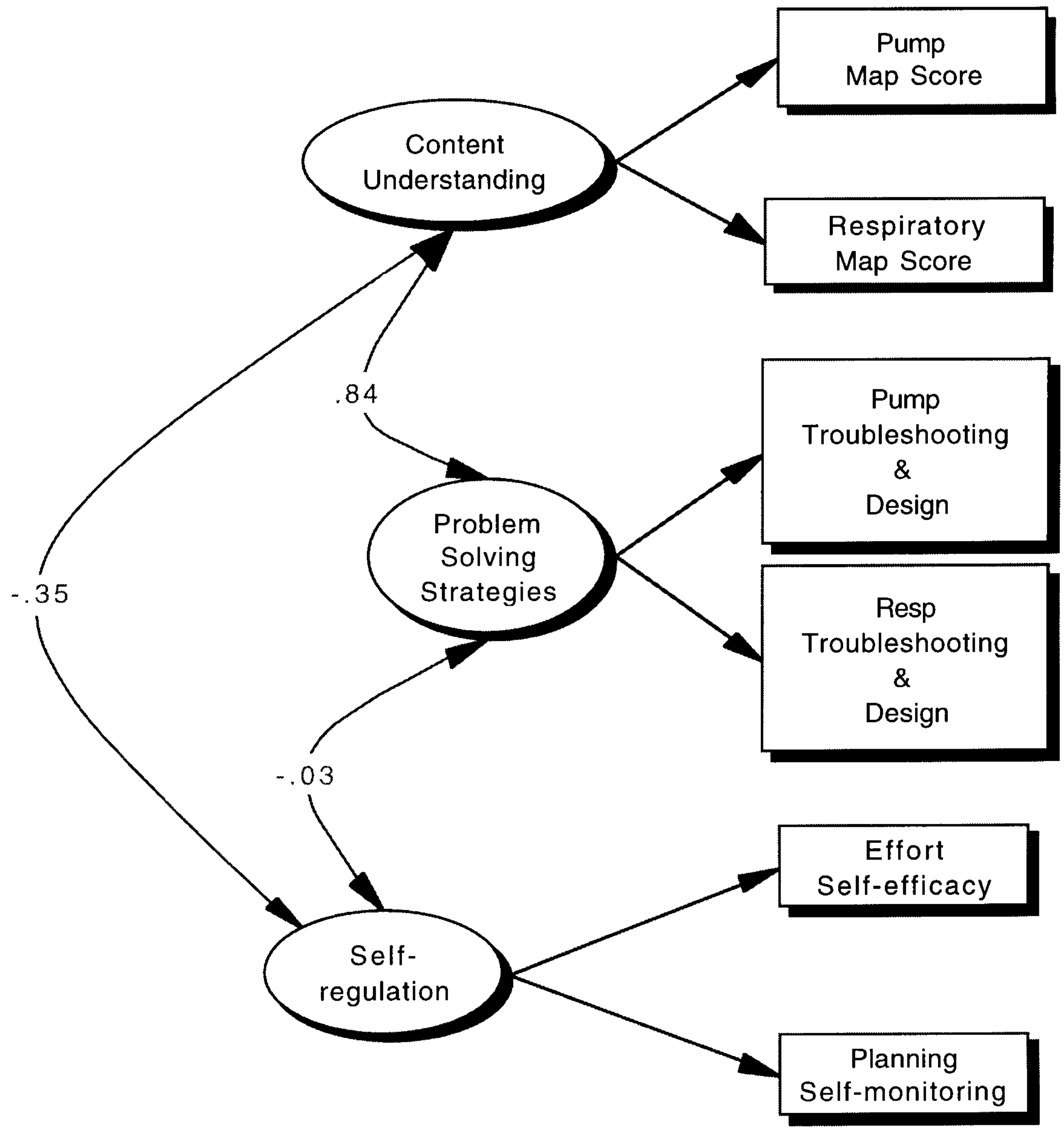


Figure 4. Latent variable model.

Discussion

A major conceptual issue is how critical should prior knowledge be in the assessment of problem solving. It was hypothesized that participants studying a supplied diagram for 2.5 minutes would perform better on the problem-solving questions than participants who did not. Indeed, participants studying diagrams listed significantly more correct answers to both the troubleshooting and design questions for both the tire pump and respiratory system tasks (see Table 18).

Because participants sampled for this study would not be classified as experts, it is not surprising that participants receiving the diagram scored higher on the problem-solving questions than those not receiving the diagram.

An additional issue is whether to provide participants with knowledge of how knowledge maps (our measure of content understanding) are scored—that is, should we provide test-taking strategies? For example, because knowledge maps will be new to most participants, should we tell them how these maps are scored? Our results show that providing scoring instructions informing adult participants about the number of correct relationships did not have any positive effect on their knowledge map performance. In fact, it was surprising that participants receiving scoring instructions stating the number of correct links constructed far fewer links than they were told were possible (M = 8.50 for the tire pump task, M = 6.65 for the human respiratory system task).

A third issue is the role of gender in assessing problem-solving skills using mechanical/physical systems. It might be expected that males would perform better than females. There were some surprising findings concerning males and females and their performance on the knowledge mapping tasks. It was hypothesized that males would score significantly higher on the bicycle tire pump map task than females; however, analyses revealed that although males did score higher, those differences were nonsignificant. The human respiratory system map proved more difficult for females (M = 1.63), and they scored much lower on that map than did males (M = 2.84). Males achieved knowledge map scores for the respiratory system task (M = 2.84) that were similar to their tire pump map scores (M = 2.97).

A final issue is whether our technical approach has sufficient reliability and validity for use in an international assessment context. In general, the results indicate the feasibility of paper-and-pencil administration and computer scoring of knowledge maps, and reasonable reliability and validity for cross-national use but not for individual use. If the use of the generalizability information is to compare various countries, one is generating a score for each country. Such a score has sufficient reliability for cross-national inferences for both the knowledge maps and the problem-solving strategies questions. However, if one is generating a score for an individual participant, such scores would not be reliable. The g coefficients are sufficiently robust to estimate a country-level measure but not sufficient to estimate an individual-level measure. These findings are analogous to those for the area of performance assessments in general. For example, in general, the reliability of performance assessments is sufficient for a school-level score but not an individual-level score. The latent variable analysis also provided a different model for interpreting generalizability coefficients for multiple topics. Such modeling presents a consistent and somewhat more optimistic view. Table 20 presents a summary table of the studies' findings.

What's Next?

The research supported by Statistics Canada has indicated that the CRESST problem-solving technology (e.g., knowledge maps) is feasible. Moreover, the samples upon which this conclusion is based were U.S. English speakers and Taiwanese Chinese (Mandarin) speakers. Thus, there is some cross-cultural generalizability. However, all the problem-solving assessments were developed at CRESST with CRESST personnel. An important question for an international study using multiple languages is whether such CRESST technology can be used or successfully modified by other countries using their own experts. A feasibility study should be conducted to test whether we can successfully collaborate with another country's experts to conduct a joint study.

Table 20

Conclusions

How many topics do we provide to be reliable?
2 topics; we can tighten up training and rubrics to improve reliability. However, G = .65, although low for individual student estimates, is acceptable for country estimates; also need to look at standard error (Bob Linn, personal communication, 4/25/98).
How many problem solving questions do we provide to be reliable?
2-3 questions (G = .66, G = .78).
What is impact of providing scoring information?
Minimal.
What is role of gender?
Continuing concern.
How difficult is the topic?
In our samples, the difficulty ranged from 24% to 38% for the knowledge maps and from 6% to 23% for the problem-solving strategy questions.
Revision of self-regulation questionnaire
Excellent shape; revised version has 20 items.

References

Abedi, J., & Baker, E. L. (1995, October). A latent-variable modeling approach to assessing interrater reliability, topic generalizability, and validity of a content assessment scoring rubric. *Educational and Psychological Measurement, 55*, 701-715.

Baker, E. L. (1995, September). *Finding our way*. Presentation at the annual conference of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles.

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

Baker, E. L., & Niemi, D. (1991, April). *Assessing deep understanding of history and science through hypertext*. Paper presented at annual conference of American Educational Research Association, Chicago.

Baker, E. L., Niemi, D., Gearhart, M., & Herman, J. L. (1990, April). *Validating a hypermedia measure of knowledge representation*. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Baker, E. L., Niemi, D., & Herl, J. (1994). Using HyperCard technology to measure understanding. In E. L. Baker & H. F. O'Neil, Jr. (Eds.), *Technology assessment in education and training* (pp. 133-152). Hillsdale, NJ: Lawrence Erlbaum Associates.

Baker, E. L., Niemi, D., Novak, J., & Herl, H. (1992). Hypertext as a strategy for teaching and assessing knowledge representation. In S. Dijkstra, H. P. M. Drammer, & J. J. G. van Merriënboer (Eds.), *Instructional models in computer-based learning environments* (pp. 365-384). Heidelberg: Springer-Verlag.

Baxter, G. P., Elder, A. D., & Glaser, R. (1996). *Assessment and instruction in the science classroom* (CSE Tech. Rep. No. 418). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Bentler, P. M. (1992). *EQS: Structural equations program manual*. Los Angeles: BMDP Statistical Software, Inc.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588-606.

Dansereau, D. F., & Holley, C. D. (1982). Development and evaluation of a text mapping strategy. In A. Flammer & W. Kintsch (Eds.), *Discourse processing*. Amsterdam: North Holland Publishing.

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). *Kit of factor-referenced cognitive factors*. Princeton, NJ: Educational Testing Service.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (rev. ed.). Cambridge, MA: MIT Press.

Herl, H. E. (1995). *Construct validation of an approach to modeling cognitive structure of experts' and novices' U.S. history knowledge.* Unpublished doctoral dissertation, University of California, Los Angeles.

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,* 206-219.

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.

Holley C. D., & Dansereau, D. F. (1984). The development of spatial learning strategies. In C. D. Holley & D. F. Dansereau (Eds.), *Spatial learning strategies. Techniques, applications, and related issues* (pp. 3-19). Orlando, FL: Academic Press.

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

Lambiotte, J. G., Dansereau, D. F., Cross, D. R., & Reynolds, S. B. (1989). Multirelational semantic maps. *Educational Psychology Review, 1,* 331-367.

Lee, C. Y. (1998). *Reliability and validity of problem solving assessments for Taiwanese high school students.* Unpublished doctoral dissertation, University of Southern California, Los Angeles.

Martinez, M. E. (1993). Problem-solving correlates of new assessment forms in architecture. *Applied Measurement in Education, 6,* 167-180.

Mayer, R. E. (1992). Teaching of thinking skills in the sciences and mathematics. In D. F. Halpern (Ed.), *Enhancing thinking skills in the sciences and mathematics* (pp. 95-116). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32,* 1-19.

Mayer, R. E., & Gallini, J. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 82*, 715-726.

Mayer, R. E., & Sims, V. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86,* 389-401.

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.

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.

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

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82,* 33-40.

Schacter, J., Herl, H. E., Chung, G. K. W. K., O'Neil, H. F., Jr., Dennis, R. A., & Lee, J. L. (1997). 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.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer.* Newbury Park, CA: Sage Publications.

Spielberger, C. D. (1975). Anxiety: State-trait process. In C. D. Spielberger & I. G. Sarason (Eds.), *Stress and anxiety* (Vol. 1, pp. 115-143). Washington, DC: Hemisphere.

Voss, J. F., & Post, T. A. (1988). On the solving of ill-structured problems. In M. T. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise* (pp. 261-287). Hillsdale, NJ: Lawrence Erlbaum Associates.

Voss, J. F., Tyler, S. W., & Yengo, L. A. (1983). Individual differences in the solving of social science problems. In R. F. Dillon & R. R. Schmeck (Eds.), *Individual differences in cognition* (pp. 204-232). New York: Academic Press.

This page intentionally blank.

Appendix A

Background Questionnaire

Questionnaire

Please answer all of the questions below. Your answers will only be used for the purpose of this study and will be kept confidential.

Gender M F (Please circle one)

Age ________________

Ethnicity ____________________________ Native Language __________________________

Education (Please circle one):

High School A.A. Bachelors M.S., M.A., PhD., Other _______________

Major (if applicable) ____

Minor (if applicable) ____

Major (if applicable) ____

Minor (if applicable) ____

Certification ____

Certification____

Credentials ____

Credentials ____

Licensing ____

Work Experience

Years of work experience: full time? ________ part time? _______

Check all the job titles that apply to your work experiences in the last 5 years.

❑ Administrator ❑ Professional ❑ Technician ❑ Protective Services

❑ Administrative Support ❑ Skilled Craft Worker ❑ Service-Maintenance

Others (please list) ____

Appendix B

Self-Regulation Trait Questionnaire

Directions: A number of statements which people have used to describe themselves are given below. Read each statement and indicate how you generally think or feel on learning tasks by marking this sheet. There are no right or wrong answers. Do not spend too much time on any one statement. Remember, give the best answer that seems to describe how you generally think or feel.

		Almost never	Sometimes	Often	Almost always
1.	I determine how to solve a task before I begin	1	2	3	4
2.	I check how well I am doing when I solve a task.	1	2	3	4
3.	I work hard to do well even if I don't like a task.	1	2	3	4
4.	I believe I will receive an excellent grade in this course.	1	2	3	4
5.	I carefully plan my course of action.	1	2	3	4
6.	I ask myself questions to stay on track as I do a task.	1	2	3	4
7.	I put forth my best effort on tasks.	1	2	3	4
8.	I'm certain I can understand the most difficult material presented in the reading of this course.	1	2	3	4
9.	I try to understand task before I attempt to solve them.	1	2	3	4
10.	I check my work while I am doing it.	1	2	3	4
11.	I work as hard as possible on tasks.	1	2	3	4
12.	I'm confident I can understand the basic concepts taught in this course.	1	2	3	4
13.	I try to understand the goal of a task before I attempt to answer.	1	2	3	4
14.	I almost always know how much of a task I have to complete.	1	2	3	4

		Almost never	Sometimes	Often	Almost always
15.	I am willing to do extra work on tasks to improve my knowledge.	1	2	3	4
16.	I'm confident I can understand the most complex material presented by the teacher in this course.	1	2	3	4
17.	I figure out my goals and what I need to do to accomplish them.	1	2	3	4
18.	I judge the correctness of my work.	1	2	3	4
19.	I concentrate as hard as I can when doing a task.	1	2	3	4
20.	I'm confident I can do an excellent job on the assignments and tests in this course.	1	2	3	4
21.	I imagine the parts of the task that I have to complete.	1	2	3	4
22.	I correct my errors.	1	2	3	4
23.	I work hard on a task even if it does not count.	1	2	3	4
24.	I expect to do well in this course.	1	2	3	4
25.	I make sure I understand just what has to be done and how to do it.	1	2	3	4
26.	I check my accuracy as I progress through a task.	1	2	3	4
27.	A task is useful to check my knowledge.	1	2	3	4
28.	I'm certain I can master the skills being taught in this course.	1	2	3	4
29.	I try to determine what the task requires.	1	2	3	4
30.	I ask myself, how well am I doing, as I proceed through tasks.	1	2	3	4
31.	Practice makes perfect.	1	2	3	4
32.	Considering the difficulty of this course, the teacher, and my skills, I think I will do well in this course.	1	2	3	4

Appendix C

Paper Folding Test

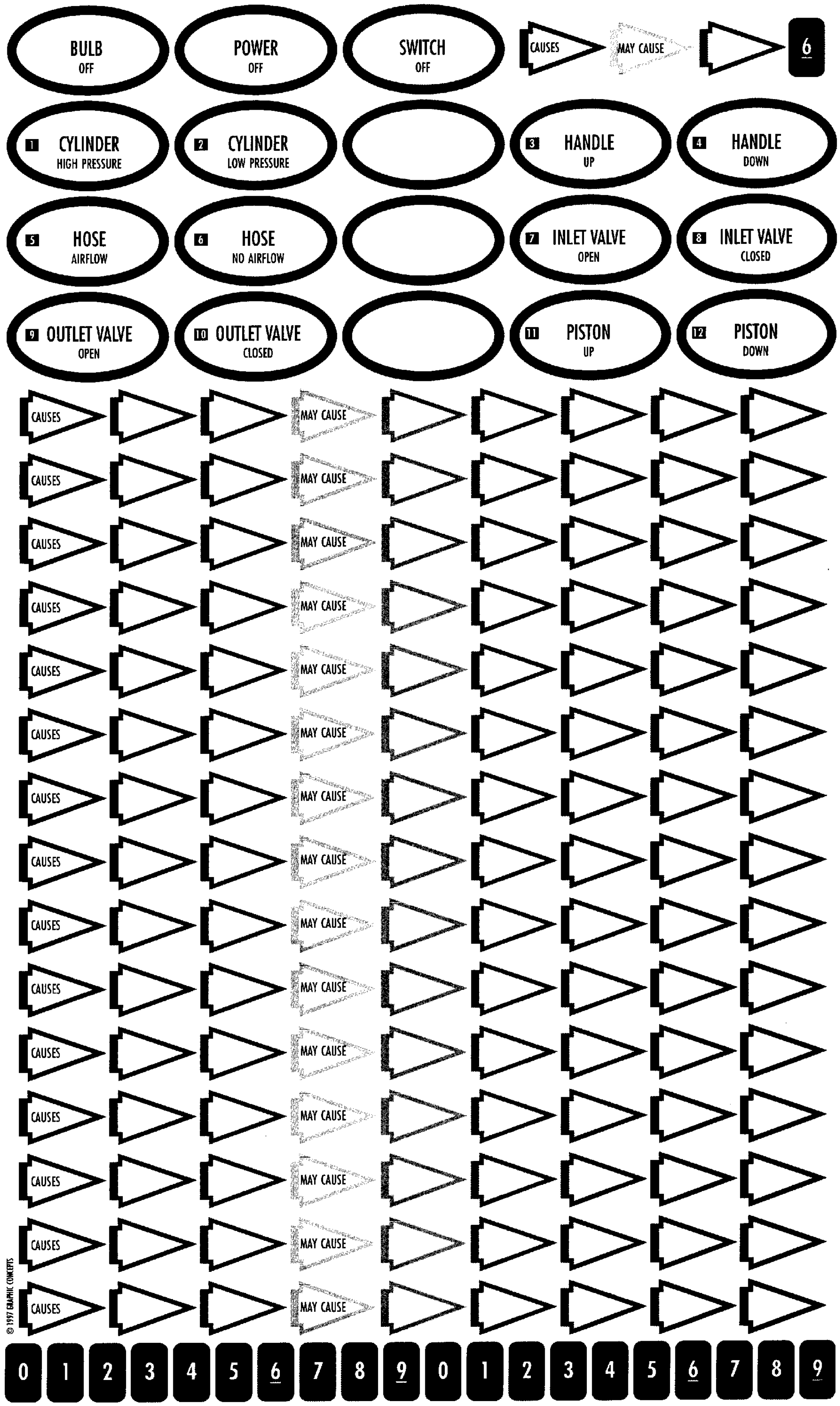
Refer to:

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Appendix D

Knowledge Mapping Task Label Sheets

1-State Bicycle Tire Pump Task
1-State Human Respiratory System Task


BULB OFF
POWER OFF
SWITCH OFF
CAUSES
MAY CAUSE
6
1 CYLINDER HIGH PRESSURE
2 CYLINDER LOW PRESSURE
3 HANDLE UP
4 HANDLE DOWN
5 HOSE AIRFLOW
6 HOSE NO AIRFLOW
7 INLET VALVE OPEN
8 INLET VALVE CLOSED
9 OUTLET VALVE OPEN
10 OUTLET VALVE CLOSED
11 PISTON UP
12 PISTON DOWN
CAUSES
MAY CAUSE
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9

BULB
OFF

POWER
OFF

SWITCH
OFF

CAUSES

MAY CAUSE

6

1 AIR SACS
OXYGEN IN

2 AIR SACS
CARBON DIOXIDE OUT

3 BLOODSTREAM
INCREASED OXYGEN

4 BLOODSTREAM
DECREASED CARBON DIOXIDE

5 CHEST CAVITY
EXPANDED

6 CHEST CAVITY
REDUCED

7 DIAPHRAGM
UP

8 DIAPHRAGM
DOWN

9 LUNGS
EXPANDED

10 LUNGS
CONTRACTED

11 RIB MUSCLES
RELAXED

12 RIB MUSCLES
CONTRACTED

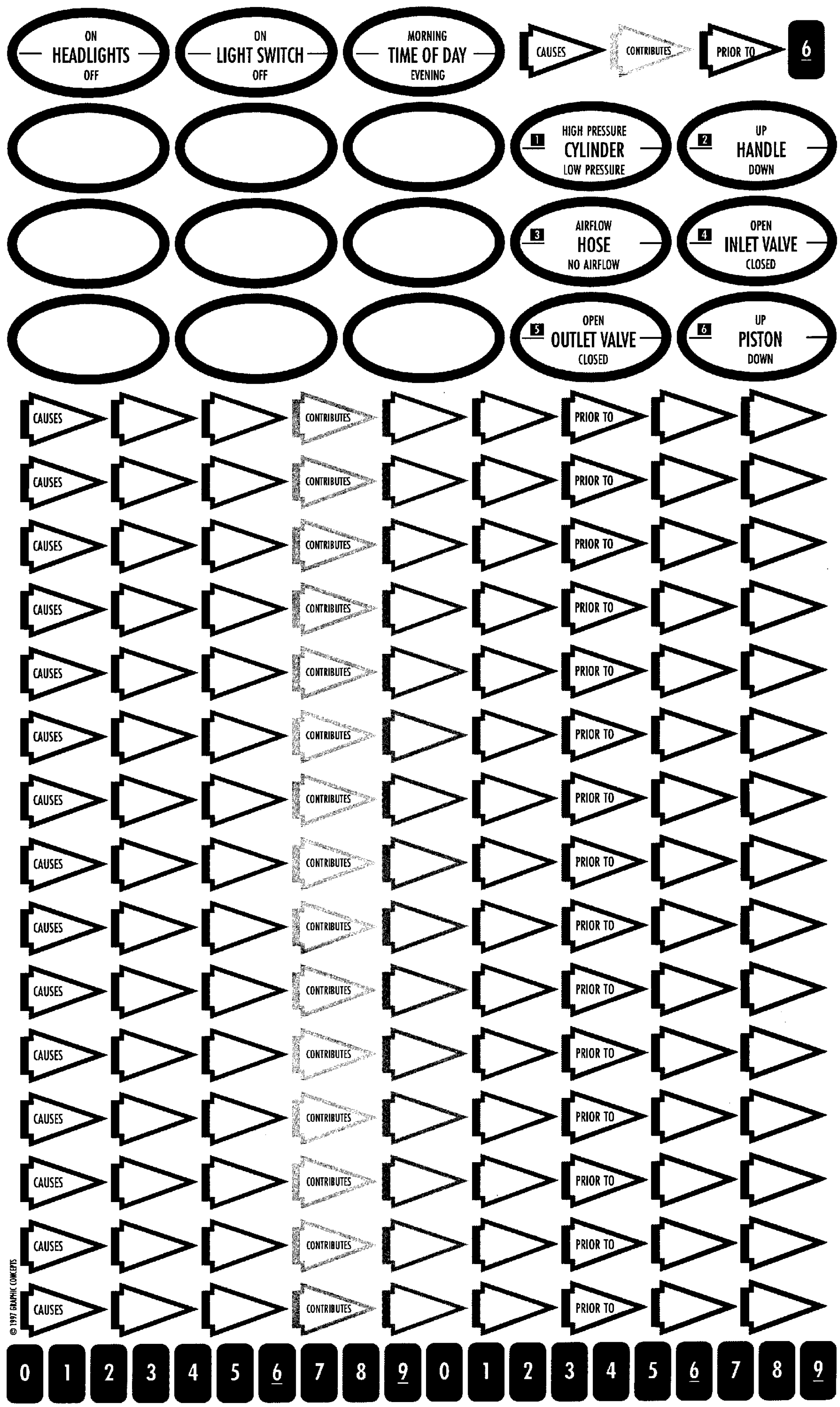
CAUSES MAY CAUSE
CAUSES MAY CAUSE
CAUSES MAY CAUSE
CAUSES MAY CAUSE
CAUSES MAY CAUSE
CAUSES MAY CAUSE
CAUSES MAY CAUSE
CAUSES MAY CAUSE
CAUSES MAY CAUSE
CAUSES MAY CAUSE
CAUSES MAY CAUSE
CAUSES MAY CAUSE
CAUSES MAY CAUSE
CAUSES MAY CAUSE
CAUSES MAY CAUSE

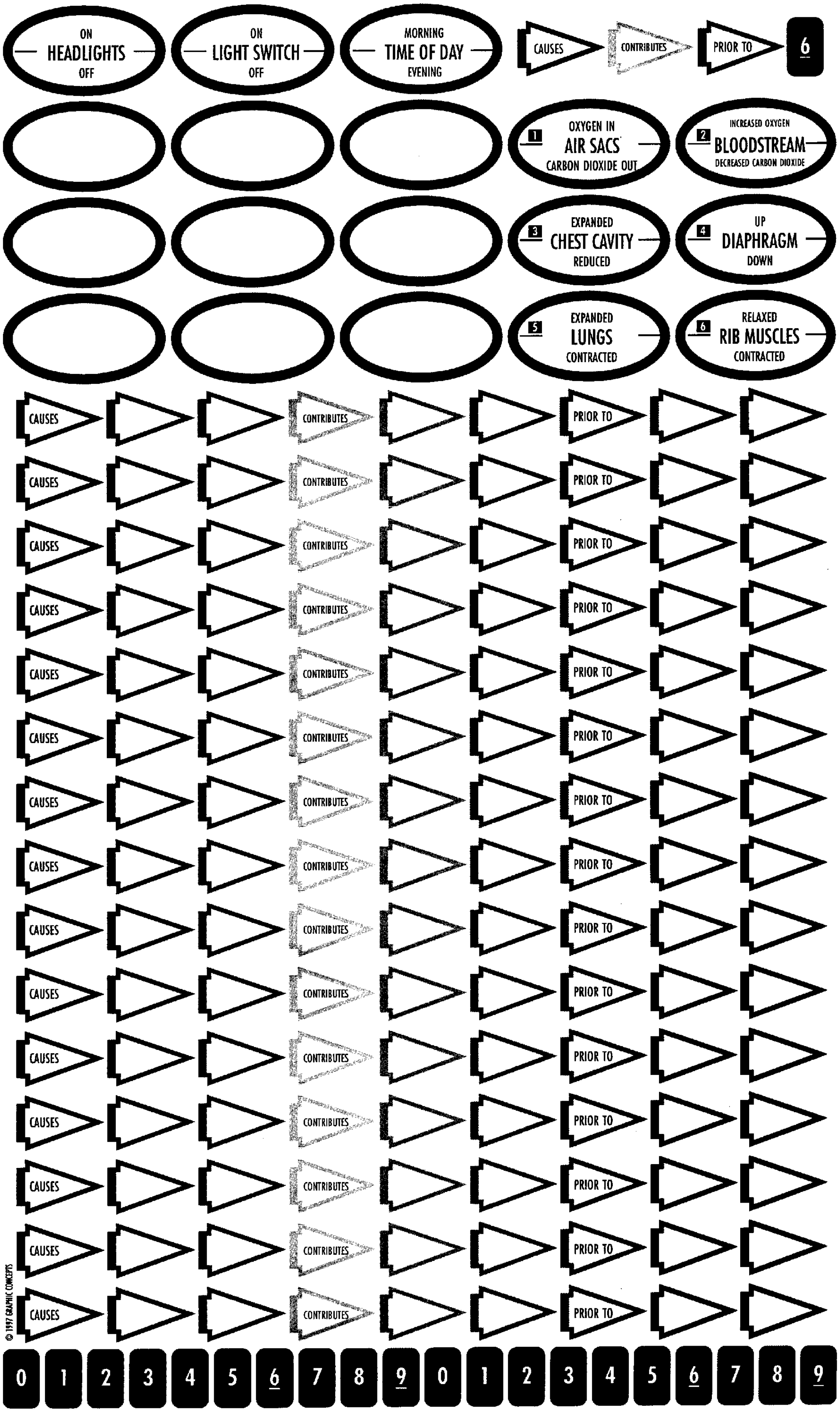
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9

Appendix E

Knowledge Mapping Task Label Sheets

2-State Bicycle Tire Pump Task
2-State Human Respiratory System Task


ON HEADLIGHTS OFF
ON LIGHT SWITCH OFF
MORNING TIME OF DAY EVENING
CAUSES
CONTRIBUTES
PRIOR TO
6
1 HIGH PRESSURE CYLINDER LOW PRESSURE
2 UP HANDLE DOWN
3 AIRFLOW HOSE NO AIRFLOW
4 OPEN INLET VALVE CLOSED
5 OPEN OUTLET VALVE CLOSED
6 UP PISTON DOWN
CAUSES
CONTRIBUTES
PRIOR TO
© 1997 GRAPHIC CONCEPTS
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9


ON HEADLIGHTS OFF
ON LIGHT SWITCH OFF
MORNING TIME OF DAY EVENING
CAUSES
CONTRIBUTES
PRIOR TO
6
1 OXYGEN IN AIR SACS CARBON DIOXIDE OUT
2 INCREASED OXYGEN BLOODSTREAM DECREASED CARBON DIOXIDE
3 EXPANDED CHEST CAVITY REDUCED
4 UP DIAPHRAGM DOWN
5 EXPANDED LUNGS CONTRACTED
6 RELAXED RIB MUSCLES CONTRACTED
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9

Appendix F

Main Study
Mapping Task Instruction Sheets

2-State Bicycle Pump Task
2-State Human Respiratory System Task

Knowledge Mapping

Task: Using the ovals and arrows provided on the following sheet, your task is to construct a knowledge map containing the relationships of the human respiratory system.

Task Instructions:

1. Look at the mapping page and correctly identify each part of the system.
2. After you are sure you have correctly identified each part of the system, remove the colored ovals and place them on the mapping page. Remember, you **CANNOT** attempt to remove them if you make a mistake**.**
3. Construct relationships for pairs of ovals you believe are related.
   - Draw the first part of the line from the **edge of the first oval.**
   - Place the **back end of the arrow on that line.**
   - Draw the rest of the line from the **arrowhead to the edge of the second oval.**

   You should only put **1 arrow per line**. Also be sure **the arrowhead is going in the proper direction.**

Mistakes:

1. If you make a relationship that you don't want, you can erase the line, however, **do NOT remove any labels**.
2. If you want to change a relationship, you can place the **new arrow** so that it **COMPLETELY** covers the first one.

Knowledge Mapping

Task: Using the ovals and arrows provided on the following sheet, your task is to construct a knowledge map containing the relationship of the operation of a bicycle pump.

Task Instructions:

1. Look at the mapping page and correctly identify each part of the system.
2. After you are sure you have correctly identified each part of the system, remove the colored ovals and place them on the mapping page. Remember, you **CANNOT** attempt to remove them if you make a mistake**.**
3. Construct relationships for pairs of ovals you believe are related.
   - Draw the first part of the line from the **edge of the first oval.**
   - Place the **back end of the arrow on that line.**
   - Draw the rest of the line from the **arrowhead to the edge of the second oval.**

   You should only put **1 arrow per line**. Also be sure **the arrowhead is going in the proper direction.**

Mistakes:

1. If you make a relationship that you don't want, you can erase the line, however, **do NOT remove any labels**.
2. If you want to change a relationship, you can place the **new arrow** so that it **COMPLETELY** covers the first one.

Scoring:

For each correct relationship you make, you will receive 1 point. The total number of points for this task is 12. You should draw a maximum of 12 lines on your map. If you draw more than 12, only 12 will be counted.

Knowledge Mapping

Task: Using the ovals and arrows provided on the following sheet, your task is to construct a knowledge map containing the relationships of the human respiratory system.

Task Instructions:

1. Look at the mapping page and correctly identify each part of the system.
2. After you are sure you have correctly identified each part of the system, remove the colored ovals and place them on the mapping page. Remember, you **CANNOT** attempt to remove them if you make a mistake**.**
3. Construct relationships for pairs of ovals you believe are related.
    - Draw the first part of the line from the **edge of the first oval.**
    - Place the **back end of the arrow on that line.**
    - Draw the rest of the line from the **arrowhead to the edge of the second oval.**

    You should only put **1 arrow per line**. Also be sure **the arrowhead is going in the proper direction.**

Mistakes:

1. If you make a relationship that you don't want, you can erase the line, however, **do NOT remove any labels**.
2. If you want to change a relationship, you can place the **new arrow** so that it **COMPLETELY** covers the first one.

Scoring:

For each correct relationship you make, you will receive 1 point. The total number of points for this task is 10. You should draw a maximum of 10 lines on your map. If you draw more than 10, only 10 will be counted. There is no penalty for guessing.

Knowledge Mapping

Task: Using the ovals and arrows provided on the following sheet, your task is to construct a knowledge map containing the relationships describing the operation of a bicycle pump.

Task Instructions:

1. Look at the mapping page and correctly identify each part of the system.
2. After you are sure you have correctly identified each part of the system, remove the colored ovals and place them on the mapping page. Remember, you **CANNOT** attempt to remove them if you make a mistake**.**
3. Construct relationships for pairs of ovals you believe are related.
   - Draw the first part of the line from the **edge of the first oval.**
   - Place the **back end of the arrow on that line.**
   - Draw the rest of the line from the **arrowhead to the edge of the second oval.**

   You should only put **1 arrow per line**. Also be sure **the arrowhead is going in the proper direction.**

Mistakes:

1. If you make a relationship that you don't want, you can erase the line, however, **do NOT remove any labels**.
2. If you want to change a relationship, you can place the **new arrow** so that it **COMPLETELY** covers the first one.

Knowledge Mapping

Task: Using the ovals and arrows provided on the following sheet, your task is to construct a knowledge map containing the relationships describing the operation of a bicycle pump.

Task Instructions:

1. Look at the sticker sheet and review the conceptual information contained in the ovals.
2. After you have reviewed the information in the ovals, remove the colored ovals and place them on the mapping page. Remember, you **CANNOT** attempt to remove them if you make a mistake**.**
3. Construct relationships for pairs of ovals you believe are related.
   - Draw the first part of the line from the **edge of the first oval.**
   - Place the **back end of the arrow on that line.**
   - Draw the rest of the line from the **arrowhead to the edge of the second oval.**

   You should only put **1 arrow per line**. Also be sure **the arrowhead is going in the proper direction.**

Mistakes:

1. If you make a relationship that you don't want, you can erase the line, however, **do NOT remove any labels**.
2. If you want to change a relationship, you can place the **new arrow** so that it **COMPLETELY** covers the first one.

Knowledge Mapping

Task: Using the ovals and arrows provided on the following sheet, your task is to construct a knowledge map containing the relationship of the operation of a bicycle pump.

Task Instructions:

1. Look at the sticker sheet and review the conceptual information contained in the ovals.
2. After you have reviewed the information in the ovals, remove the colored ovals and place them on the mapping page. Remember, you **CANNOT** attempt to remove them if you make a mistake**.**
3. Construct relationships for pairs of ovals you believe are related.
    - Draw the first part of the line from the **edge of the first oval.**
    - Place the **back end of the arrow on that line.**
    - Draw the rest of the line from the **arrowhead to the edge of the second oval.**

    You should only put **1 arrow per line**. Also be sure **the arrowhead is going in the proper direction.**

Mistakes:

1. If you make a relationship that you don't want, you can erase the line, however, **do NOT remove any labels**.
2. If you want to change a relationship, you can place the **new arrow** so that it **COMPLETELY** covers the first one.

Scoring:

For each correct relationship you make, you will receive 1 point. The total number of points for this task is 12. You should draw a maximum of 12 lines on your map. If you draw more than 12, only 12 will be counted.

Knowledge Mapping

Task: Using the ovals and arrows provided on the following sheet, your task is to construct a knowledge map containing the relationships of the human respiratory system.

Task Instructions:

1. Look at the sticker sheet and review the conceptual information contained in the ovals.
2. After you have reviewed the information in the ovals, remove the colored ovals and place them on the mapping page. Remember, you **CANNOT** attempt to remove them if you make a mistake**.**
3. Construct relationships for pairs of ovals you believe are related.
   - Draw the first part of the line from the **edge of the first oval.**
   - Place the **back end of the arrow on that line.**
   - Draw the rest of the line from the **arrowhead to the edge of the second oval.**

   You should only put **1 arrow per line**. Also be sure **the arrowhead is going in the proper direction.**

Mistakes:

1. If you make a relationship that you don't want, you can erase the line, however, **do NOT remove any labels**.
2. If you want to change a relationship, you can place the **new arrow** so that it **COMPLETELY** covers the first one.

Knowledge Mapping

Task: Using the ovals and arrows provided on the following sheet, your task is to construct a knowledge map containing the relationships of the human respiratory system.

Task Instructions:

1. Look at the sticker sheet and review the conceptual information contained in the ovals.
2. After you have reviewed the information in the ovals, remove the colored ovals and place them on the mapping page. Remember, you **CANNOT** attempt to remove them if you make a mistake**.**
3. Construct relationships for pairs of ovals you believe are related.
    - Draw the first part of the line from the **edge of the first oval.**
    - Place the **back end of the arrow on that line.**
    - Draw the rest of the line from the **arrowhead to the edge of the second oval.**

    You should only put **1 arrow per line**. Also be sure **the arrowhead is going in the proper direction.**

Mistakes:

1. If you make a relationship that you don't want, you can erase the line, however, **do NOT remove any labels**.
2. If you want to change a relationship, you can place the **new arrow** so that it **COMPLETELY** covers the first one.

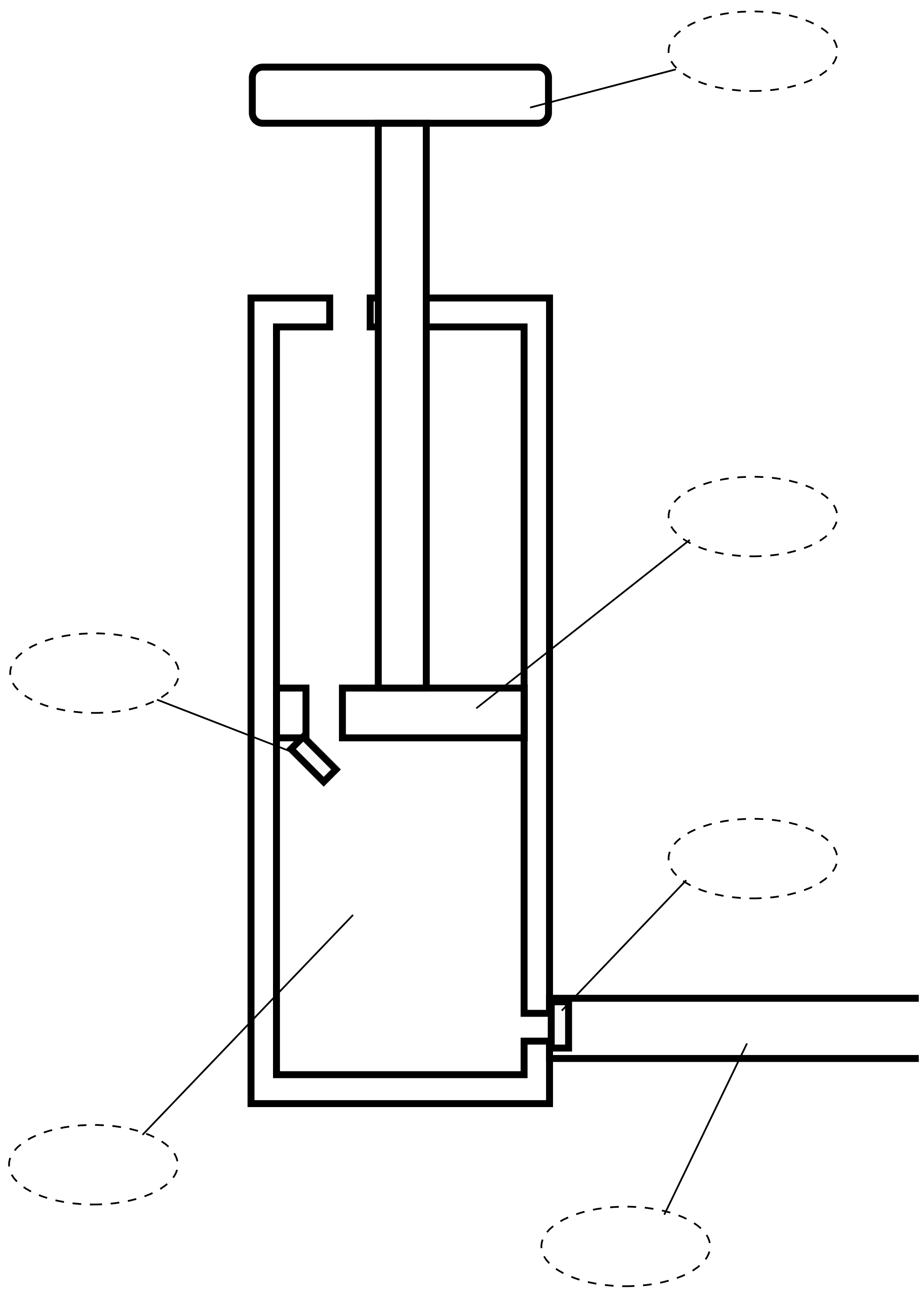
Scoring:

For each correct relationship you make, you will receive 1 point. The total number of points for this task is 10. You should draw a maximum of 10 lines on your map. If you draw more than 10, only 10 will be counted. There is no penalty for guessing.

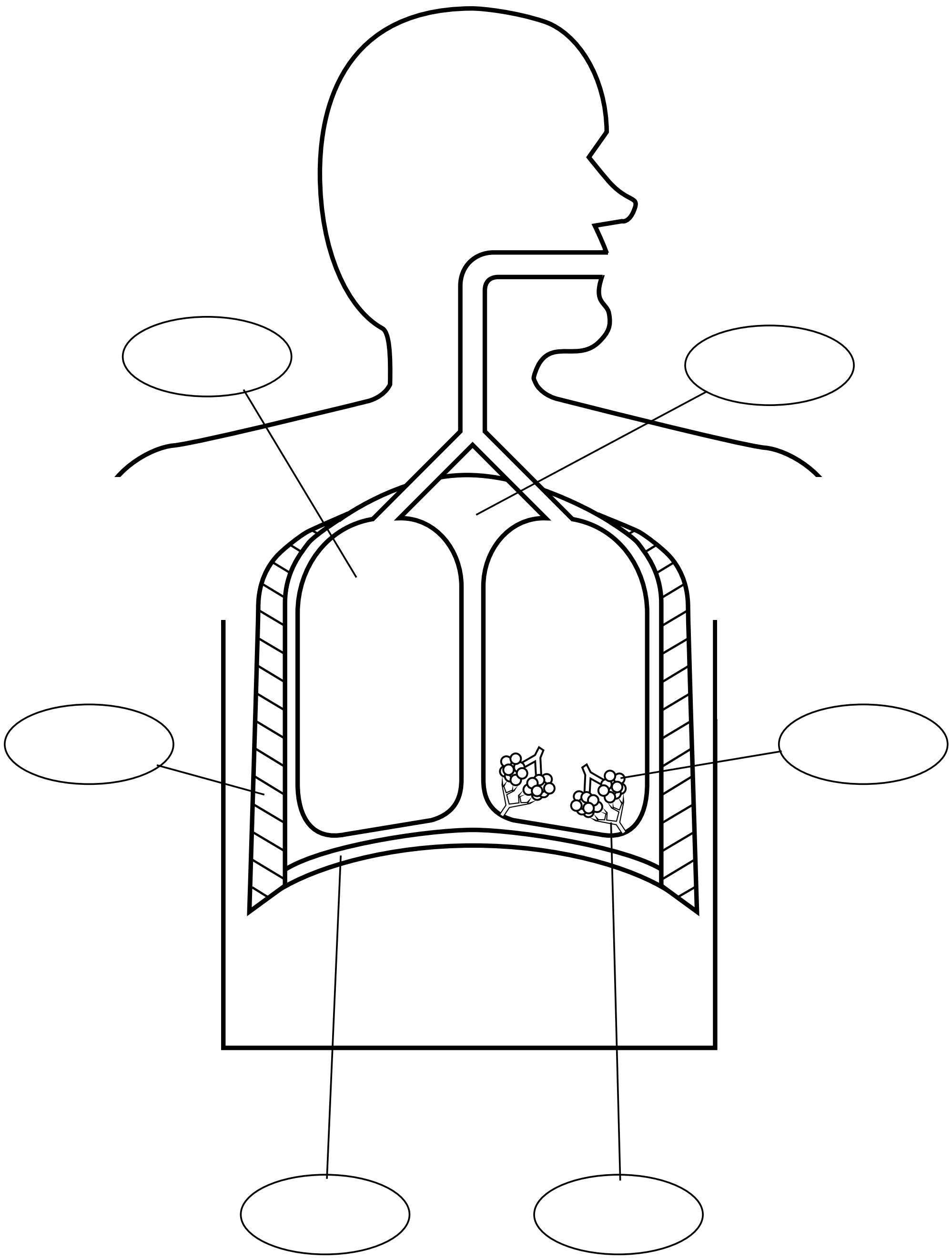
Appendix G

Knowledge Mapping Task Diagrams

2-State Bicycle Tire Pump Task



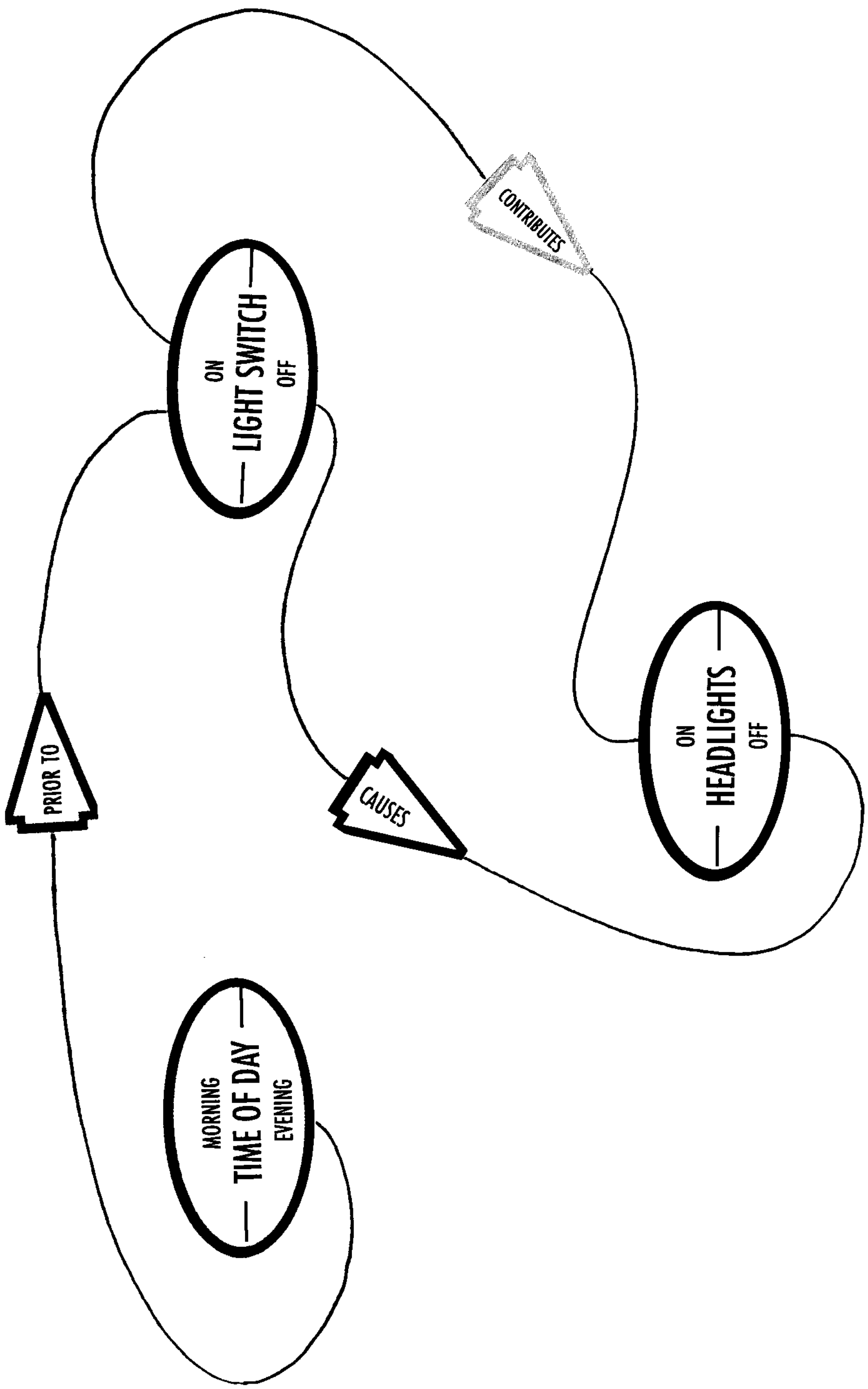
2-State Human Respiratory System Task



Appendix H

Knowledge Mapping Task
Map Construction Training Session

Cause and Effect Model

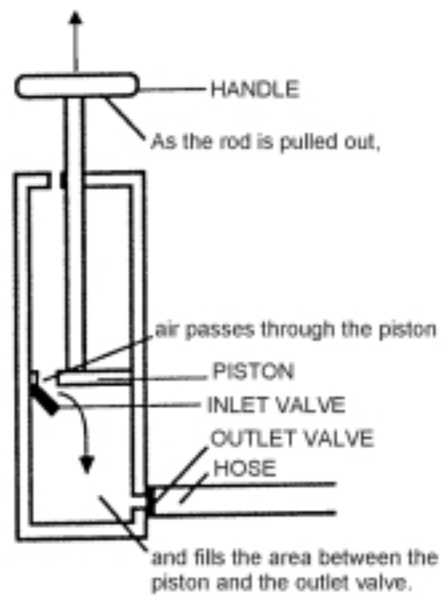


Appendix I

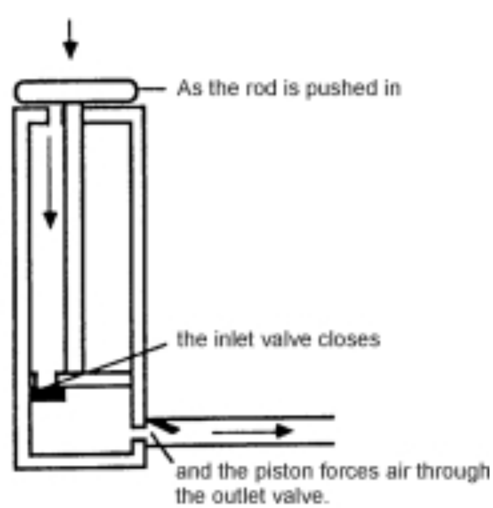
Problem Solving Strategy Task

Diagram of Bicycle Tire Pump

Bicycle Tire Pump



(1)



(2)

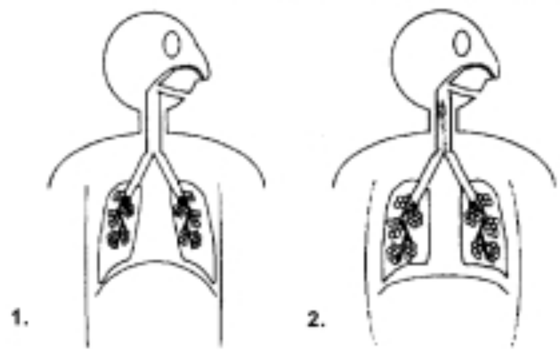
Appendix J

Problem Solving Strategy Task

Diagram of Human Respiratory System

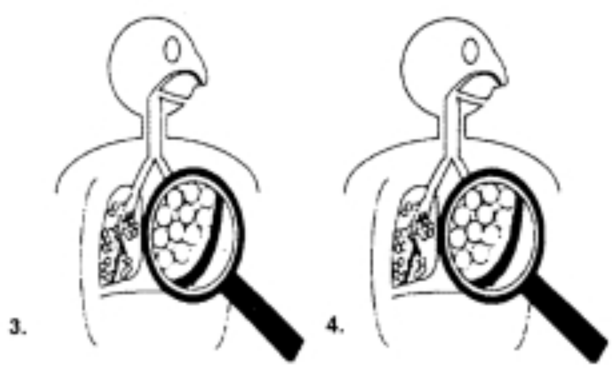
Respiratory System

There are three phases in respiration--inhaling,exchanging, and exhaling.



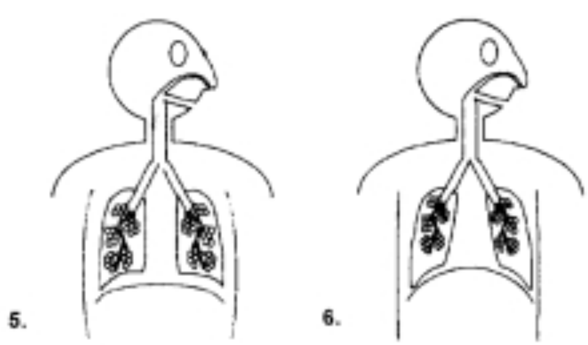
Inhaling

During inhaling, the diaphragm moves down creating more space for the lungs, air enters through the nose or mouth, moves down through the throat and bronchial tubes to tiny air sacs in the lungs.



Exchanging

During exchange, oxygen moves from the air sacs to the blood stream running nearby, and carbon dioxide moves from the bloodstream to the air sacs.



Exhaling

During exhaling, the diaphragm moves up, creating less room for the lungs, air travels through the bronchial tubes and throat to the nose and mouth, where it leaves the body.

Appendix K

Problem Solving Strategy Task: Bicycle Tire Pump Task

Troubleshooting and Design Question Prompts

Bicycle Pump*

1. Suppose you push down and pull up on the handle several times but no air comes out of the bicycle pump. What could be wrong? List as many reasons as you can think of.

Bicycle Pump*

2. How can you increase the efficiency of the bicycle pump? That is, how can you move more air through the pump? List as many reasons as you can think of.

Bicycle Pump*

3. Suppose that when you begin to push down on the handle of a bicycle pump it won't go down easily and little air is coming out of the hose. As you continue to push down on the handle, it suddenly becomes easier and the air is expelled as it would normally. What could be causing this situation to happen? List as many reasons as you can think of.

* In the research setting, each prompt was presented on a single page.

Appendix L

Problem Solving Strategy Task: Human Respiratory System Task

Troubleshooting and design question prompts

Respiratory*

1. Not enough oxygen is getting to the brain, and a person is about to faint. What could be wrong with the respiratory system? List as many reasons as you can think of.

Respiratory*

2. Suppose you are a scientist trying to improve the human respiratory system. How could you get more oxygen into the bloodstream faster? List as many reasons as you can think of.

Respiratory*

3. Suppose a person is hyperventilating (over-breathing). What is happening to the person's respiratory system? List as many reasons as you can think of.

* In the research setting, each prompt was presented on a single page.

Appendix M

Scoring key

Self-Report Trait Self-Regulation Questionnaire

Scales	Items
Planning	1, 5, 9, 13, 17, 21, 25, 29
Self-Checking	2, 6, 10, 14, 18, 22, 26, 30
Effort	3, 7, 11, 15, 19, 23, 27, 31
Self-efficacy	4, 8, 12, 16, 20, 24, 28, 32

PLANNING

1. I determine how to solve a task before I begin.
5. I carefully plan my course of action.
9. I try to understand tasks before I attempt to solve them.
13. I try to understand the goal of a task before I attempt to answer.
17. I figure out my goals and what I need to do to accomplish them.
21. I imagine the parts of a task I have to complete.
25. I make sure I understand just what has to be done and how to do it.
29. I try to determine what the task requires.

SELF-CHECKING

2. I check how well I am doing when I solve a task.
6. I ask myself questions to stay on track as I do a task.
10. I check my work while I am doing it.
14. I almost always know how much of a task I have to complete.
18. I judge the correctness of my work.
22. I correct my errors.
26. I check my accuracy as I progress through a task.
30. I ask myself, how well am I doing, as I proceed through tasks.

EFFORT

3. I work hard to do well even if I don't like a task.

7. I put forth my best effort on tasks.

11. I work as hard as possible on tasks.

15. I am willing to do extra work on tasks to improve my knowledge.

19. I concentrate as hard as I can when doing a task.

23. I work hard on a task even if it does not count.

27. A task is useful to check my knowledge.

31. Practice makes perfect.

SELF-EFFICACY

4. I believe I will receive an excellent grade in this course.

8. I'm certain I can understand the most difficult material presented in the readings for this course.

12. I'm confident I can understand the basic concepts taught in this course.

16. I'm confident I can understand the most complex material presented by the teacher in this course.

20. I'm confident I can do an excellent job on the assignments and tests in this course.

24. I expect to do well in this course.

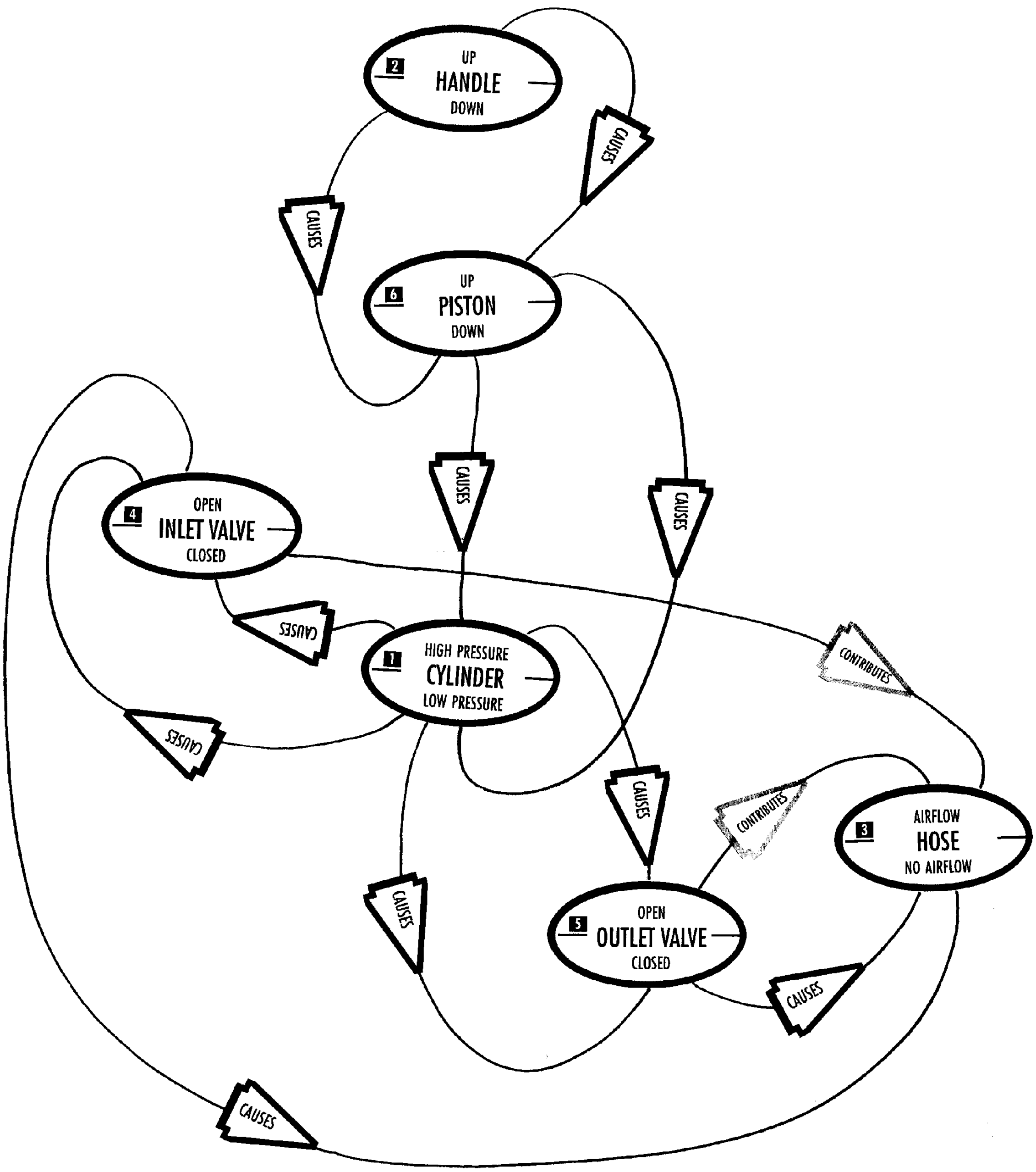
28. I'm certain I can master the skills being taught in this course.

32. Considering the difficulty of this course, the teacher, and my skills, I think I will do well in this course.

Appendix N

Knowledge Mapping Task—Bicycle Tire Pump Task

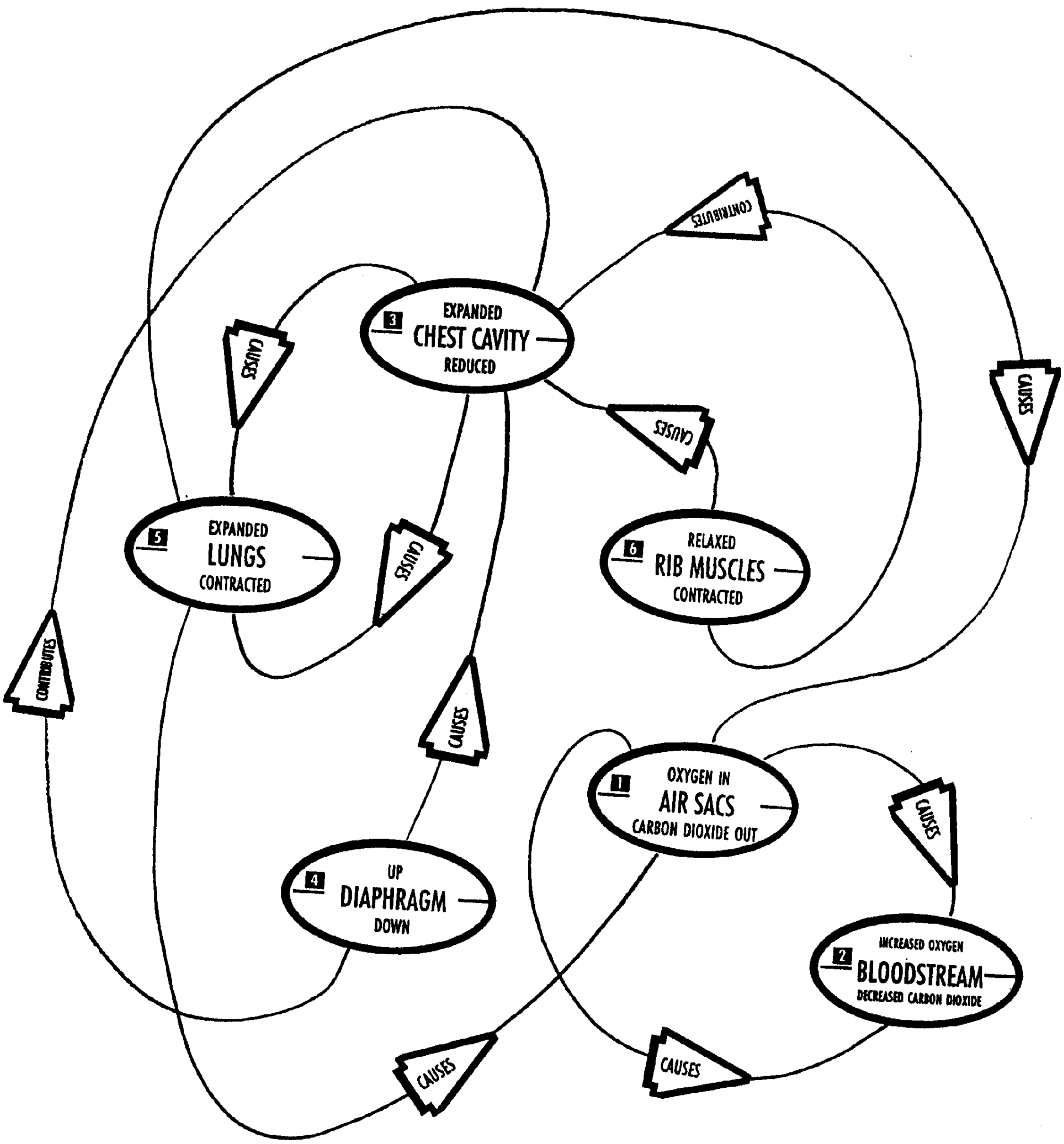
Scoring Model Map



Appendix O

Knowledge Mapping Task—Human Respiratory System Task

Scoring Model Map



Appendix P

Problem Solving Strategies Task

Bicycle Tire Pump
Human Respiratory System

Rubrics for the Problem Solving Questions

Bicycle Tire Pump

Question 1. Suppose you push down and pull up on the handle several times but no air comes out of the pump. What could be wrong? List as many reasons as you can think of.
Leak in the seal around inlet valve-Leak in the seal around outlet valve

Acceptable:

1. Leak between piston and cylinder
2. Hole in hose
3. Outlet valve stuck open or closed (inlet valve did close; outlet valve did not open; outlet valve fell off)
4. nlet valve stuck open or closed (inlet valve did close; inlet valve did not open; inlet valve fell off)
5. Handle disconnected from rod (handle stuck; handle did not move)
6. There is no air going into cylinder (atmosphere lacks air)
7. Did not push and pull hard enough
8. Hole in cylinder
9. Piston stuck up or down
10. Inlet valve related answers
11. Outlet valve related answers
12. Piston related answers
13. Handle related answers
14. Cylinder related answers
15. Hose related answers
16. Valve(s) stuck open / closed
17. Moving piston too fast, air can't go out

Not acceptable:

1. Hose is not connected to tire
2. Hose is turned off
3. Piston is broken
4 Inlet valve is broken
5. Outlet valve is broken
6. Cylinder broken
7. Leaks (no specific location)

Bicycle Tire Pump

Question 2. How can you increase the efficiency of a pump? That is, how can you move more air through the pump? List as many reasons as you can think of.

Acceptable:

1. Move the handle up and down faster; move the handle higher
2. Increase the length, the width, or the volume of the cylinder
3. Increase the size of the piston
4. Increase the size of the valve(s)
5. Increase the size of the hose
6. Improve efficiency of seal, decrease leakage

Not acceptable:

1. Increase the length of the handle
2. Connect pump to a motor (without explanation)
3. Increase pressure, suck more air in
4 Add lubricant, clean up the dirt and the rust

Bicycle Tire Pump

Question 3. Suppose that when you begin to push down on the handle of a bicycle pump it won't go down easily and little air is coming out of the hose. As you continue to push down on the handle, it suddenly becomes easier and the air is expelled as it would normally. What could be causing this situation to happen? List as many reasons as you can think of.

Possible answers:

1. The cylinder is rusty (Cylinder is clogged with dirt, is sticky). As the handle is pushed down, rust falls off.
2. The piston is rusty (Piston clogged with dirt; is sticky; piston is stuck). As the handle is pushed down, rust falls off.
3. Outlet valve is blocked by foreign object (Outlet valve is stuck; is dirty or sticky). As the handle is pushed down, high air pressure dislodges it.
4. Hose is blocked by foreign object (Hose is stuck; hose partially closed). As the handle is pushed down, high air pressure dislodges it.

The Human Respiratory System

Question 1. Not enough oxygen is getting to the brain, and a person is about to faint. What could be wrong with the respiratory system? List as many reasons as you can think of.

Acceptable:

1. Not enough oxygen in the air (air is too polluted; altitude is too high)
2. Person is not breathing enough
3. Diaphragm is stuck (diaphragm does not move down enough; no vacuum is created)
4. Nose or mouth is blocked (not enough oxygen is going into the nose or mouth)
5. Throat or bronchial tubes are blocked (not enough oxygen is going down the throat or bronchial tubes)
6. Air sacs do not receive enough oxygen (air leaks out on way to air sacs; hole in air passage)
7. Oxygen does not move from air sacs to bloodstream (no blood vessels near air sacs; blood vessels unable to absorb oxygen; blood is too full of carbon dioxide)
8. Blood vessels do not travel to brain (blockage in vessels to brain; heart does not pump hard enough)
9. Not sufficient red blood cells
10. Hole in the lung, unable to expand
11. Ribs broken, pierce through the lungs
12. Inter-costa muscles damaged, can't expand lungs/chest

Not acceptable:

1. Person is having a stroke, has lung cancer, has heart disease
2. Too much cholesterol
3. Too much smoking
4. Use smelling salts (slap person in face)

The Human Respiratory System

Question 2. Suppose a person cannot breathe, what could be wrong with his/her respiratory system? List as many reasons as you can think of.

Acceptable:

1. Breathe in smoke, not enough oxygen in the environment
2. Diaphragm is stuck (diaphragm is damaged; diaphragm does not function normally)
3. Nose or mouth is blocked
4. Throat or bronchial tubes are blocked
5. Hole/leaks in air sacs (air leaks out on way to air sacs)
6. Hole in the lung, oxygen can't go to the lungs
7. Chest cannot expand (with any internal or external reasons)

Not acceptable:

1. Tuberculosis, asthma, or other lung disease

The Human Respiratory System

Question 2b. Suppose you are a scientist trying to improve the human respiratory system. How could you get more oxygen into the bloodstream faster? List as many reasons as you can think of.

Acceptable:

1. Increase number of air sacs, thus increasing surface area and rate of diffusion
2. Increase the volume of air sacs, thus increasing the alveolar pressure so more air can enter the lungs
3. Increase the chest cavity, thus increasing the volume for more air to enter
4. Increase breathing rate (e.g., increasing the contraction and relaxation of the diaphragm muscle)
5. Increase amount of hemoglobin the blood (oxygen-carrying molecule)
6. Breathe cleaner air (which will increase breathing efficiency and thus take in more oxygen)
7. Take deeper breaths
8. Increase the blood flow to lung (either by increasing heart rate or dilation of blood vessels)
9. Transfusion of oxygenated blood
10. Reduce carbon dioxide level, thus blood pH will decrease and cause hemoglobin molecules to have greater affinity for oxygen. Then, more oxygen can be uptake into the blood
11. Breathe oxygen-rich gas (e.g., oxygen tank / mask)
12. Stronger diaphragm will increase breathing efficiency, thus more oxygen is breathe in and faster

The Human Respiratory System

Question 3. Suppose a person is hyperventilating (over-breathing). What is happening to the respiratory system?

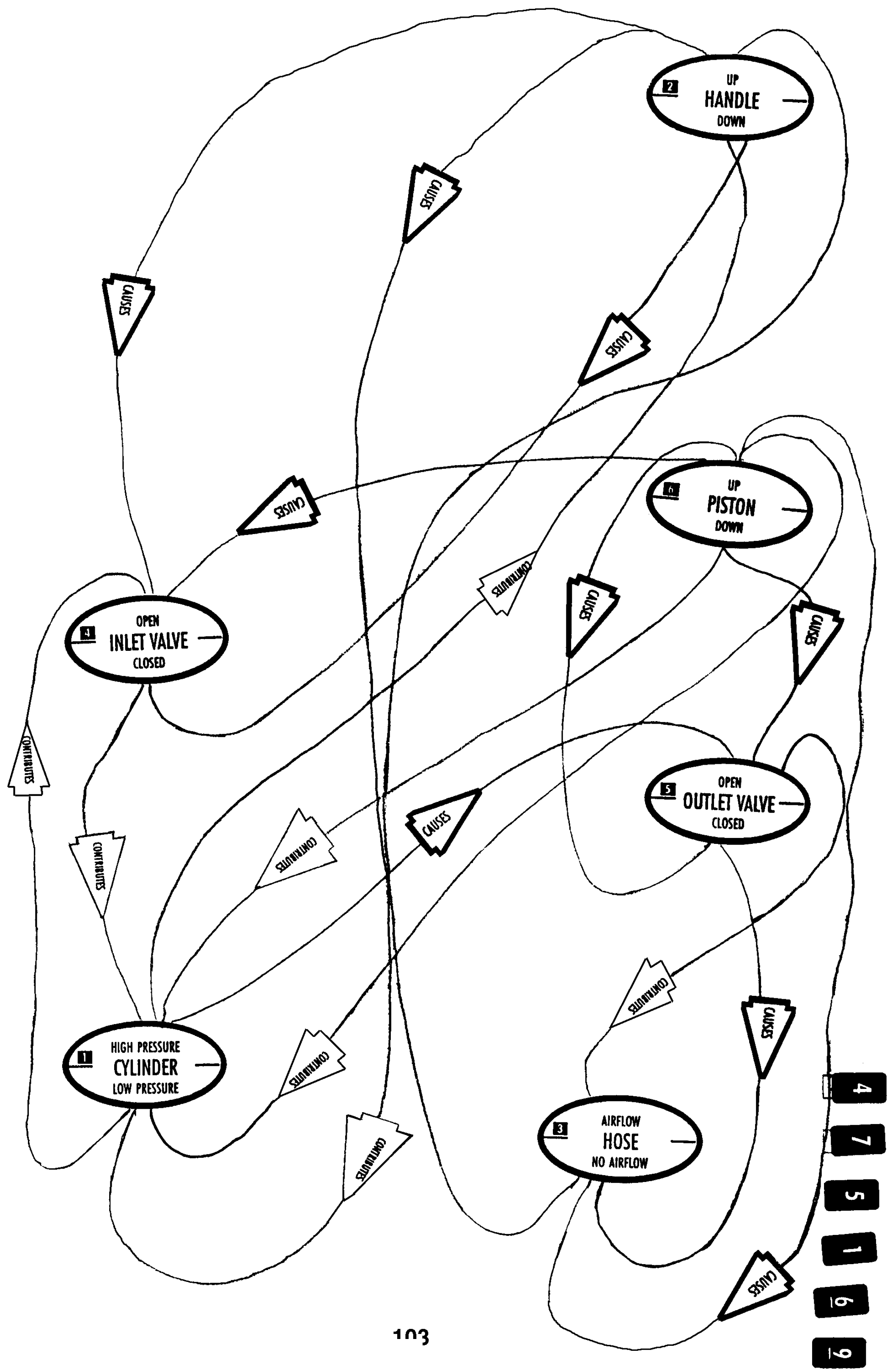
Acceptable:

1. Increase in respiratory depth (After physical activity)
2. Increase O_2 consumption by the body / tissue (O_2 consumption is higher than O_2 entry) (After physical activity)
3. Normal O_2 consumption by the body / tissue (respiratory problem)
4. Low level of oxygen in bloodstream because of high consumption by body / tissue (After physical activity)
5. High level of oxygen in bloodstream because of high O_2 entry at lung (Respiratory problem)
6. Increase in CO_2 production by body / tissue (After physical activity)
7. Normal CO_2 production by body / tissue (Respiratory problem).
8. High level CO_2 in bloodstream (level of CO_2 is higher than CO_2 excretion from lung) (After physical activity)
9. Low CO_2 in bloodstream because CO_2 is removed faster (Respiratory problem) (After physical activity)
10. Vasodilatation of blood vessel (to increase diffusion of gases at lung) (After physical activity)
12. Vasoconstriction of blood vessel (to keep more O_2 from coming in and to decrease the amount of CO_2 leaving too fast) (Respiratory problem)
13. Lung ventilation is higher
14. Decrease in respiratory depth (Respiratory problem)

Appendix Q

Knowledge Mapping Task: Bicycle Tire Pump Task

High-Scoring Knowledge Map



Appendix R

Knowledge Mapping Task: Human Respiratory Task

High-Scoring Knowledge Map