

Structural Assessment to Support Engineering Education

Jennifer Turns, Alex Kirlik
Georgia Institute of Technology

Abstract: The ABET 2000 changes in accreditation standards of engineering programs will soon make it necessary for engineering programs to demonstrate that their students have the knowledge and abilities necessary for performing professional engineering activities. Structural Assessment, a method for assessing students knowledge of the relationships among concepts, methodologies, and problems in a domain, may represent a valuable assessment resource. Past research shows that this assessment methodology can be valuable, but does not focus on how to make the methodology feasible for classroom use. In this paper, we describe how we created a suit of tools that support structural knowledge assessment for classroom use and a pilot evaluation of these tools that demonstrates the feasibility of such a method.

1. DEVELOPING ASSESSMENT PRACTICES FOR ENGINEERING EDUCATION

Assessment, in the context of education, is the process of measuring what a student knows [1]. The reasons to perform assessment can be quite varied. Such reasons include a need or desire to: (1) determine whether students have learned the material that is being taught, (2) determine whether students have mastered knowledge that is a prerequisite for a class, (3) diagnose why a student is having difficulty in a class, (4) direct students attention at a particular facet of the material to be learned, and (5) trace students development of knowledge and abilities over time. At a school level, the new ABET 2000 accreditation standards will soon make it necessary for schools to be able to demonstrate what students have learned (outcomes based assessment) rather than simply demonstrating that they are teaching an acceptable complement of courses. This has created a need to develop strategies, and possibly even new assessment instruments, for assessing engineering students knowledge and abilities as they move through the curriculum.

Assessment is, and has always been, very difficult. Some of the difficulties associated with assessment include the following.

- *Deciding what to assess:* Within engineering it can be difficult to determine what exactly students should be accountable for knowing and being able to do. For example, in an engineering design class, do we want to assess students' *knowledge* of the steps of a particular method, their *ability to evaluate* how well a method was applied, their *ability to execute* the method, and/or their *ability to integrate* the results of the method into an emerging designed artifact.
- *Deciding how to assess:* Given an assessment goal, another problem is developing an instrument that is capable of capturing levels of knowledge relative to the goal. A frequent problem with assessment is having, at hand, assessment instruments that can support the instructor with his/her goals and can be easily adapted to his/her situation. Ongoing debates explore the relative merits of standard testing procedures, authentic performance based assessments, and alternative assessments such as the use of concepts maps.
- *Designing and Executing the Assessment:* Designing and executing an assessment, even when the basic goals and methods have been identified, is typically very time consuming. Anyone who has ever made up a set of test questions and then graded students responses is aware of this phenomenon.

Given a decision on what to assess (e.g., ability to apply theories to case study problems) and how to assess (e.g., short answer questions), the tasks of designing the instrument, collecting the responses from the students, keeping track of the information, determining what to count as correct, and correcting the responses can take a lot of time.

- *Creatively Balancing Instruction and Assessment Time.* Time spent performing assessment, particularly collecting assessment information from students, is often seen as antithetical to more learning oriented activities such as lecture and project work. Assessment *can* be part of a good learning experience, but such benefits do not come without effort. Having students answer large numbers of multiple choice questions in order to determine their level of understanding may not help them improve their understanding, particularly if they do not get sufficient feedback on their mistakes.
- *Fully Using the Assessment Results.* Finding time to process assessment results, above and beyond the individual student level, can be difficult. This includes the activities of synthesizing individual student assessment results into a characterization of the class in order to (1) make teaching decisions (e.g., whether to move to another subject, or explore the current subject in more depth), (2) assess the assessment tool (e.g., determining if the assessment instrument captures known differences among students, discriminates at a level which is useful, etc.), and (3) refine the assessment methodology (e.g., improve portions of an instrument that were ambiguous to students and therefore not very useful).

The remainder of this paper explores how a theoretically motivated and laboratory validated assessment process, Structural Assessment, can be adapted to a classroom environment. Structural Assessment is a method targeted at structural knowledge, knowledge of the relationships among concepts in a domain, a method whose properties are being studied in the human factors and educational psychology communities, and a method which is simple to define. By combining the basic properties of Structural Assessment with a set of tools to design, execute and synthesize the assessment results in a class setting, we believe that many of the above issues can be addressed. The resulting approach would then offer instructors and engineering schools an approach to add to their assessment tool-kit.

2. STRUCTURAL KNOWLEDGE AND ASSESSMENT: THEORY AND PRACTICE

Research on expertise (e.g., research on chunking in expert memory) has shown that not only do experts know more than novices, but they also have significantly different ways of organizing their knowledge [2-4]. Structural knowledge is a term that refers to the structure or organization of concepts that one knows [5]. This is a form of knowledge beyond the actual understanding of individual concepts (i.e., declarative knowledge). By identifying methods for capturing and representing structural knowledge, we can 1) characterize exactly how structural knowledge differs in experts and novices and subsequently use this type of method to 2) assess whether students are growing closer to expertise.

Despite its importance, student development of structural knowledge often goes unmonitored in classes due to a lack of useable techniques for assessing this type of knowledge. Several types of techniques can be used to capture and model structural knowledge [5]. Some traditional methods include multiple choice questions focused on the relationships among concepts in a domain [6], essay questions where students must discuss the structural aspects of a domain [6], or concept mapping tasks where students draw the organization of the concepts in a domain [7].

While these methods are all useful for focusing assessment on structural knowledge, there are several limitations associated with their use:

- *The elicitation and representation of structural knowledge are blurred.* In a concept mapping task, a student is responsible both for having the knowledge and being able to express the knowledge. Many studies have shown that there are large discrepancies between what we know and what we can articulate about our knowledge [8]. While the ability to articulate knowledge may be a valuable educational goal, an inability to accurately express what one knows could interfere with the use of a structural assessment technique for diagnosis of student problems.
- *Interpreting the results can be difficult.* There is a lack of pre-specified metrics for interpreting each of these techniques. Often the products are interpreted in highly subjective and qualitative ways [7].
- *Performance of the tasks can require training.* The usefulness of essay writing and concept mapping may be limited by the student's ability to write clearly or fully understand how to construct a concept map. With training, students can learn to perform both tasks better, but sufficient time is not always available for training [7].

2.1. The Beginnings of a Formalized Structural Assessment Process

Structural Assessment is a method which has shown some promise for addressing these difficulties. Structural Assessment (SA) is a term coined by Kraiger to represent a formalized assessment methodology. SA uses graph representations of the structure of concepts based on relatedness ratings [9]. Specifically, given a set of concepts whose structure is to be assessed, the *elicitation* process involves students providing ratings of relatedness between pairs of concepts, the *representation* process involves the transformation of the relatedness ratings into a network representation of the structure of the concepts via the "Pathfinder" algorithm, and the *interpretation* process involves the comparison of the students network representation with a referent network, typically that of an expert.

The heart of the process is the Pathfinder algorithm [10, 11]. The pathfinder algorithm takes a set of ratings of relatedness (or psychological proximity) of the pairs of concepts as its input and produces a link-weighted network representation capturing local relationships among the concepts. The networks produced can capture hierarchical structure and/or clustering if such structures exist, but the procedure does not constraint output to have either format. The technique is similar to Multi-dimensional Scaling (MDS) [12] in that it transforms a set of relatedness ratings into a representation of the structure of the concepts. The techniques differ in the basis upon which they represent the structure of the knowledge. The product of the Pathfinder algorithm is the specification of a graph (i.e., the set of nodes and links between nodes) while MDS is based on an spatial metaphor in which concepts are position at points along the dimensions.

Studies in domains as diverse as Human-Computer Interaction (HCI), statistics, computer programming, and fighter pilot maneuvering have demonstrated that (1) the similarity between a student's derived Pathfinder representation and an expert's is predictive of performance differences [6, 13-15] and (2) this similarity measure is more predictive of performance than similarity measures based on MDS representations or simple correlations of the ratings of the expert and the student [16]. In these studies, many measures have been used to operationalize the construct of similarity between two concept networks with one particular measure, "C", showing the most explanatory power. This measure is based on the comparison of the

“neighborhood” of links connected to any node in each of the two graphs being compared. The values on this measure range from no similarity, 0, to perfect similarity, 1.

Given these findings, Kraiger et al. [9] have begun to formalize an assessment process, Structural Assessment, based on the use of the similarity value “C” as the assessment metric. They are attempting to characterize the properties of “C” as an assessment metric for learning situations, above and beyond its ability to discriminate levels of expertise. In a first experiment (conducted in the domain of SPSS instruction), “C” was found to improve over the course of instruction. In a second experiment, “C” was found to be sensitive to students use of an advance organizer, a technique known to improve students learning of material during instruction. Essentially, students who used the advance organizer during instruction and learning had higher “C” scores than students who did not use the advance organizer. With respect to performance, the “C” metric was found to be predictive of performance in the second experiment but not in the first. The authors concluded that the inclusion of assessment scores for several students who had an extremely poor understanding of the material were creating variance in the first experiment. When these students were removed from the analysis, the “C” metric was found to be predictive of performance. The authors conclude from this that the “C” metric may provide useful assessment information for students who generally understand the material.

2.2. *Issues in Adapting Structural Assessment to Classroom Situations*

Past research shows that this assessment methodology can be valuable, but issues of validity of and feasibility of the method for classroom uses still need to be explored. From a validity perspective, more data on the properties of “C” are needed, including data on the use of this metric in different domains, with different types of concepts, and at different levels of understanding. This paper focuses primarily, though, on the feasibility issues. Several features of the work on Structural Assessment up to this point make it difficult to put it into new contexts without some changes. First, the authors of these studies appear to have made many decisions based on their experience with this process, experience which instructors new to the process will not possess. Second, the data collection involved either students going to a particular site in order to perform the rating tasks or students receiving diskettes with the electronic versions of the rating task that they must return with a completed task. Third, the analysis appears to have been time consuming since few tools exist to perform the analysis in a batch type manner. Further, while there is software for analyzing relatedness data and creating the graph representations (i.e., Pathfinder software [17]), this software does not permit the analyses to be performed in real time so that students can get immediate feedback

3. TASK ANALYSIS: CLASSROOM ASSESSMENT USING STRUCTURAL ASSESSMENT

In order to develop tools to support the use of a primarily experimental methodology in more traditional academic contexts (i.e., classroom and curriculum assessment), it helps to understand the set of tasks required to carry out the assessment. This section presents an analysis of the tasks required to use Structural Assessment in the classroom. The task analysis is diagrammatically represented in Figure 1. The tasks themselves are described in the following subsection. More details about these tasks, including details on suggestions and advice from the literature about the specific task, can be found in Turns and Kirlik (1997). The software that supports these tasks is described in Section 4 of this paper.

3.1. Design the Assessment

An extensive amount of work is involved in the design of an assessment, before students or learners ever become involved. These design tasks are carried out by the instructors of the course.

1. *Determine the scope of the assessment.* The first decision to be made is the decision about the scope of the assessment. One must decide whether the goal is to assess the learning over a set of classroom lessons spanning a few days, the learning over an entire term, and the growth of expertise as a student moves through a curriculum. This decision pervades the remaining efforts.
2. *Identify concepts.* The second task is to determine the set of concepts whose relationships are going to be modeled. There should be between 15 and 30 concepts. Generally, fewer concepts will result in less discriminability of the resulting metric score. More than 30 concepts requires too many ratings by the subject.
3. *Create referent graphs.* The similarity calculations require the existence of a referent graphs, typically an expert's. The person(s) chosen to be the referent needs to go through the rating process so that their model can be created via the Pathfinder algorithm (essentially steps 4-5 below).

3.2. Assess Individual Students

The student's role in the assessment process is to perform the assessment activities that have been designed. In return, the student expects to receive some feedback on his/her current status. This feedback can range from simply telling the student the score that he/she has received for his/her performance to more detailed feedback attempting to help the student understand how well he/she is doing, what problems he/she seems to be having, and how he/she might want to focus his/her energies in the near future.

4. *Collect relatedness data from subjects.* The input to the Pathfinder algorithm is a set of proximities, in this case psychological proximities, between pairs of concepts taken from the set of concepts. The proximities, or relatedness ratings, for each unique pair of concepts needs to be collected.
5. *Create Pathfinder representations capturing the student's structure of the concepts.* The psychological proximities between concepts need to be transformed, through the Pathfinder algorithm, into a graph representation capturing the structure of the concepts. If possible, this process should be done in real time so that the results can be used to provide feedback to the students.
6. *Assess Expertise Relative to Referent Graph.* Expertise, in Structural Assessment, is determined by calculating "C" based on the comparison of the student's Pathfinder network with the referent network.

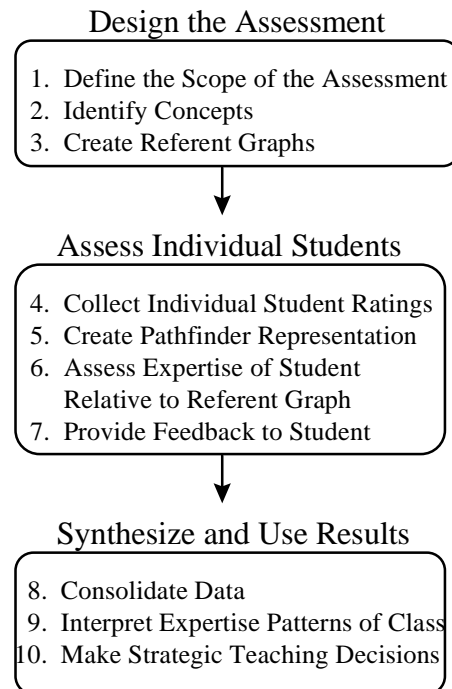


Figure 1. Task Analysis, Activities Required to Conduct Structural Assessment in a Classroom

7. *Provide individual students with feedback and debriefing information.* The Structural Assessment process is more likely to capture students' interest, and to serve as not only an assessment activity but also as a learning activity, if the results of the analysis are available as immediate feedback. Further, in order for this methodology to support instructor needs, such as enhancing feedback to students, some type of feedback needs to be made available to the student almost immediately.

3.3. *Synthesize and Use Results*

Once all students have completed the assessment activities, and all of the student work has been scored, the uses of the assessment instruments, requiring synthesis of the data, may begin. This involves the consolidation of the data for the individual students, the interpretation of the data for the entire group of students, and the use of these interpretations for teaching decisions and revisions of the assessment instrument.

8. *Consolidate data.* If the goal is to simply assess students or to provide students with more feedback, then the process would progress to this step only as a means of collecting and recording the individual assessment scores. This step will be more or less difficult depending on how the data was collected (i.e., students rated pairs on paper, on a standalone computer system, or via a networked system). If the rating process is supported over a network, then the data consolidation can be virtually automatic.
9. *Interpret expertise patterns of class.* With all of the students' results consolidated in a single place, an instructor can then explore patterns within the entire set of data. An instructor might want to perform this step in order to understand patterns of misconceptions in a class. An instructor might also want to double check that the classification of expertise based on the measure chosen in step 7 corresponds to both his/her intuition and/or other measures being used in the class.
10. *Use data to support instructional decisions.* A final step in the process would be to use the results of the assessment to make strategic instructional decisions. For example, a teacher may change the next period's lecture based on the concepts that students are not understanding in class. A departmental administrator might suggest that a student revisit course concepts before moving on to the next level of a class.

4. SOFTWARE TOOLS

Three pieces of software were developed to support the use of Structural Assessment. These three pieces of software map directly to the three classes of activity which must take place in using Structural Assessment with students. The designing of the assessment is supported by the Web-based SA-Admin software. The data collection and modeling for individual subjects (the execution of the assessment) is supported by the Web-based SA-Testing software. Finally, the analysis of the results is supported by the Excel-based SA-Analysis software. In the next three subsections, we describe the function and form of these software tools.

4.1. *Design the Assessment: SA-Admin*

The administrative software supports the tasks of defining and monitoring a Structural Assessment test. An assessment is defined by its name, the set of concepts and the sets of referent graphs (i.e., expert models) against which students will be compared. The SA-Admin software supports the administrator in creating and editing tests by allowing the instructor to create and edit the set of concepts that are the focus of the assessment, making the referent graphs available, and monitoring the identity and output of the students who have completed the

assessment. The software is located at <http://chmsr.isye.gatech.edu/cgi-bin/Pathfinder/admin.cgi>.

Main SA-Admin Screen. The screen in Figure 2 is the initial screen of the Web-based implementation of SA-Admin. At this screen, an instructor (the administrator) can choose to edit an existing assessment (on the left of the figure) or create a new assessment (on the right figure). To edit an existing assessment, the instructor simply selects the test by choosing it from the pull down list showing all assessments that exist on the server and then pressing the “Click here to edit” button. To create a new assessment, the instructor enters the name that will be associated with the new assessment and presses the “Click Here to Create” button.

Concept Editing. Initially the concepts are entered in a “freeform” manner, using the interface shown in Figure 3. The concepts are entered sequentially with commas to separate them. Each time the concepts are saved, the system updates the display showing the current set of concepts to the left of the editing window as well as in the window. This type of freeform editing can continue until a complete set of rating data is collected from a subject or referent. Since the number of questions given in the ratings task is based on the total number of concepts, the number of concepts cannot be changed after a set of ratings exists. Thus after a set of ratings exist, editing is still possible, but it must be performed using the “fixed” concept editing mode. In fixed concept editing, the wording of the concept name can be changed, but concepts can be neither added nor removed. Changing the number of concepts at this point would corrupt the



Figure 2. Main Screen in the SA-Admin Tool

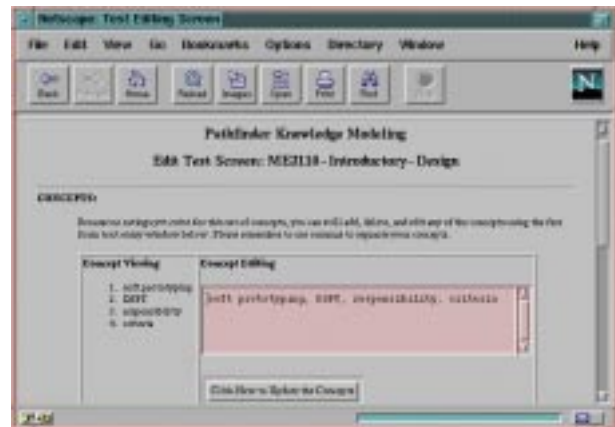


Figure 3. Freeform Concept Editing



Figure 4. Management of Referent Ratings

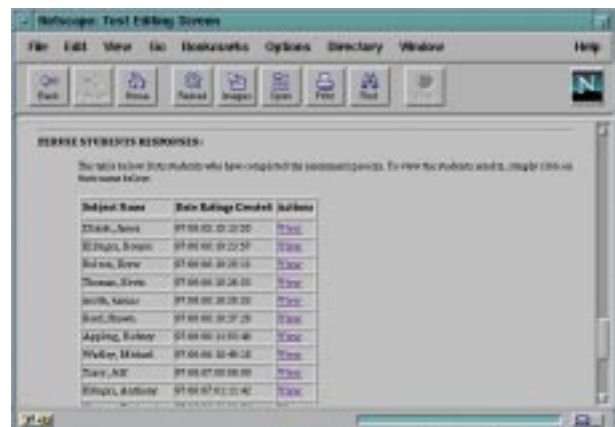


Figure 5. Monitor Student Participation

data that has already been collected.

Management of Referent Graphs. In order to compute similarity of a student graph to the referent's, there must be a referent graph available. Clicking on “add expert ratings” from the main SA-Admin page (see Figure 4) will take the instructor through the data collection and modeling process (see the next section). Once completed, the referent graphs can be viewed.

Monitor Student Participation. As students begin to complete the data collection and graphing activity, it can be useful to be able to verify who has completed the activity and also to look at their results. As depicted in Figure 5, the “Peruse Subject Responses” portion of SA-Admin provides a list of all students who have completed the data collection activity, the time at which they began the activity, and a link to their “debriefing” page where their results and the feedback they received can be viewed.

4.2. Execute the Test: SA-Assessment

Data collection and modeling is the focus of the SA-Assessment tool. The tool, located at <http://chmsr.isye.gatech.edu/cgi-bin/Pathfinder/login.cgi>, is web-based. On the next few pages, the functionality of the tool is described.

Student Registration. Students start the assessment by selecting the one they are supposed to complete from a pull-down list, entering their first and last name, and then pressing the button labeled “proceed to instructions”. Student's data will be stored based on their first and last name and the

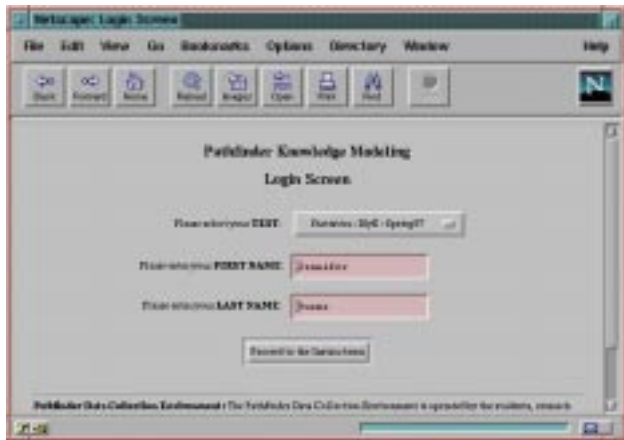


Figure 6. Subject Participation - Registration



Figure 7. Instructions



Figure 8. Warm-up Rating Task

“time-stamp” of when they began the assessment. Because the data is associated with a unique “time-stamp”, students can repeat the assessment if they wish.

Instructions. SA-Assessment provides three types of instructions - instructions on the task of assigning relatedness ratings to pairs of concepts, instructions on the meaning of the relatedness scale, and instructions about the procedure for the task. The instructions on the relatedness scale are shown in Figure 7. On a later screen, subjects are shown the set of concepts in order to give them a feel for them.

Warm-up Rating Task. Two possible problems with the relatedness rating task are getting students to become aware of the full set of concepts being graphed and getting students calibrated with respect to the relatedness scale. Our two “warm-up” tasks, the maximum and minimum relatedness tasks, were designed to address these issues. In the maximum relatedness task, the student is given the complete set of concepts and is supposed to indicate (by clicking in the check boxes), the two concepts which he/she feels to be the most related. In the minimum relatedness task the student is again given the complete set of concepts and the subject is supposed to identify the two concepts which are least related. These tasks appear on the same screen, as shown in Figure 8. The maximum and minimum warm-up tasks are followed by a debriefing screen on which the two selected pairs are displayed along with text indicating that 1) the maximum relatedness value of 7 should be associated with pairs whose relatedness is similar to the pair chosen in the maximum relatedness warm-up task and that 2) the minimum relatedness value of 1 should be associated with pairs whose relatedness is similar to the minimally related pair.

Providing Ratings. Figure 9 shows a copy of the screen that is used to collect relatedness ratings. The relatedness scale appears at the top, with the ends of the scale marked with “least related” and “most related”. The ends of the scale are also annotated with the maximally and minimally related pairs from the warm-up task, as points of reference. The pair of concepts whose relatedness is being scored appears at the center bottom of the screen, directly above the “Continue” button. The student looks at the pair of concepts, decides which rating to assign, indicates this rating by clicking in one of the radio buttons associated with the appropriate value and pressing the “Continue” button. Error checking will prevent a student from “continuing” without identifying a relatedness rating. At the very top of the screen is an indicator showing the number of the current pair and the total number of ratings that are required. The rating process continues with pairs from the set being randomly chosen until all the ratings are assigned. Because there is a distinct URL associated with each rating page, it is possible for a student to stop the test at one point, and return to it later.



Figure 9. Assignment of Ratings

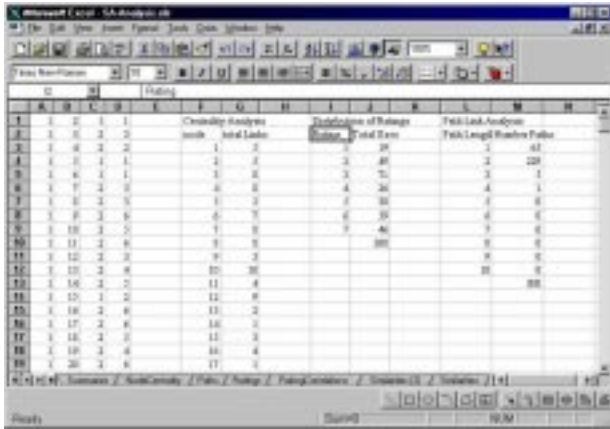


Figure 12. Subject worksheet, created by SA-Analysis.

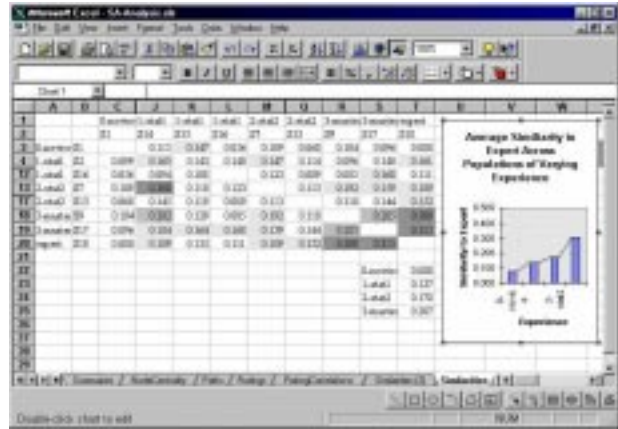


Figure 13. Summary of Inter-Subject Similarities

corresponding to the highest levels of similarity. A sample of this spreadsheet is shown in Figure 13. The primary focus on a similarity analysis would probably be on the columns where each student is compared with the referent. The data in the spreadsheet, though, would also permit one to explore how similar students are within the groups. The SA-Analysis worksheet is set up to perform three other summary analyses. One analysis captures the number of times each student/referent used each relatedness rating value. A second analysis tabulates the number of direct links, two link paths, three link paths, etc. in the graphs of each individual. The third analysis summarizes the correlations of actual ratings across all participants' graphs, including the correlations of each student with each referent. These analyses produce the "ratings", "paths", and "ratingCorrelations" spreadsheets respectively.

5. PILOT STUDY

In order to evaluate the potential usefulness of the method and the tools, we piloted the process in the domain of industrial engineering statistics. The overall goal of the pilot study was to evaluate the potential usefulness of Structural Assessment methodology, supported by the three software tools described in the previous section. Specifically, we wanted to explore two issues – **validity** of the method and the **feasibility** of the method. In terms of validity, we wanted to verify that the similarity measure, "C", would capture known expertise differences among student populations. In terms of feasibility, we wanted to explore the ease with which tests could be defined, administered and analyzed with the tools as well as the time involved.

We initially worked with a faculty member in the statistics division of the ISyE department (hereafter know as the expert) to reduce a set of potential concepts, derived from a set of departmental statistics standards and the expert's class materials, into a set of 25 statistics concepts associated with differing levels of statistics knowledge. SA-Admin was used to configure a pilot test based on these 25 concepts and the ratings of our expert. Eighteen student subjects performed the relatedness rating task and received similarity and diagnostic feedback using SA-Assessment. These subjects were representative of four discrete levels of prior statistics instruction including no previous experience, completion of undergraduate statistics I or II or advanced instruction at the Masters level. We performed analyses using SA-Analysis. Statistical results show that "C" captured a statistically significant difference between the three groups. From a feasibility perspective, the software did support the entire process, but the

evaluation provided direction in knowing how to make the process even more feasible. In this section, we describe this pilot experiment results in greater detail.

5.1. Validity

The graph in figure 14 represents the average similarity of students to the expert within the four populations. The four populations are defined by the extent of previous statistics instruction. Clearly, there is a *trend* of increasing similarity to the expert with increasing experience. Using a one-factor ANOVA, the factor of experience is significant in explaining the difference in similarity values ($p < 0.0001$).

Within the stat-1 group, there are three subjects with similarity values that are of a magnitude consistent with subjects from the stat-2 population (i.e., Similarity for subject2=0.161, subject5=0.160, and subject14=0.190 while the average for the remaining subjects in the stat-1 group is 0.115 and the average for the stat-2 group is 0.178). Two of these three subjects are juniors enrolled in statistics and expecting an A. The third subject is a freshman enrolled in statistics and expecting a B. Perhaps this student is a student with a natural tendency toward understanding statistics. Are these students (particularly the freshman) particularly good statistics students. It would be interesting to track them in Stat-2 and observe their performance.

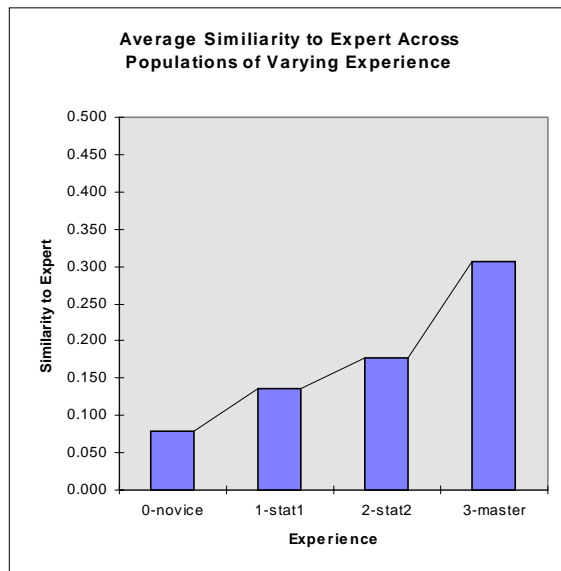


Figure 14. Similarity of Population Groups (based on experience) to the Expert

5.2. Feasibility

Overall, the pilot study demonstrated the feasibility of carrying out a Structural Assessment with the use of the software. Once we had identified the set of concepts, the setting up of the test took only the amount of time required to enter the concepts and was done using the SA-Admin tool. The expert remotely entered her ratings via the SA-Assessment tool. All students, except three, completed the test remotely using the SA-Assessment tool, without any intervention from the authors of this report. All analysis was performed via the SA-Analysis tool. During the experiment, we made the following additional observations about the feasibility.

- *Task Time for Subjects.* Participants reported that the task of providing the ratings for the 25 concepts via the web-based SA- Assessment software required between 30 and 45 minutes. These subjects reported that they spent a great portion of this time waiting for the system to accept their response and present a new rating. While the software only required about 2 seconds to process each response, this delay summed over 300 questions accounts for approximately 10 minutes of the total task time. The three subjects that performed a paper based-version of the task reported that the task required around 30 minutes.
- *Subject Interest and Engagement:* Twenty-seven students started the electronic task, but only 14 completed the task electronically. Although we never were able to ask the other 13 students why they did not complete the task, the response would probably have been similar to the response of many of the subjects when asked about their impressions once they were finished. Students generally commented that the task was straightforward and that they felt comfortable making the

relatedness judgments. Some comments did focus on the tediousness of the activity -- “it was long and boring” and “it was a little abstract”. The students completing the task at the beginning of the experiment did not ever see a graphical representation of their graph, since the code to produce this representation was not available until the end of the experiment. These earlier subjects saw only the matrix representation of the graph, which they reported to not understand. Further, these students did not seem interested in understanding the matrix. The students, though, who had seen the graphical representation in their debriefing appeared to be much more interested in the output.

- *Subject Reactions to Interface:* Several subjects reported that the connection between the warm-up task and the pairs of concepts displayed below the terms “most related” and “least related” was not clear. These subject further reported that once they determined what these reference pairs were supposed to represent, they continued to ignore them. Other subjects reported that the debriefing screen was rather overwhelming. They had not known what to expect and they were not certain what information was important. Another subject reported that the debriefing information was interesting, but that the description of the similarity value was not sufficient for him to fully understand it. In particular, this subject had had one of the highest similarity values ($C=0.30$), but had thought the value exceptionally low.
- *Capturing all of the important information.* The system currently collects relatedness ratings for all pairs of concepts from a set. While this is all the information required for the algorithm to determine a network representation, additional judgments of the subject’s familiarity with the items would have been useful in trying to understand and interpret the networks.

6. LESSONS LEARNED AND FUTURE DIRECTIONS

In this paper, we have explored issues in adapting a research-based assessment technique to classroom practice. Using software we developed to address the challenges of the classroom environment, we were able to explore the validity and the feasibility of the use of Structural Assessment to support engineering design education. Without any changes, we feel that the Structural Assessment process, conducted with tools like the ones we have built, offers instructors an additional tool for their assessment toolbox. At the same time, some suggestions for enhancing our software and proposed process to even more fluidly support the Structural Assessment process include:

- *Simplify the rating process for students.* This includes allowing students to indicate, and thus remove from the rating process, unfamiliar concepts and allowing students to rate multiple pairs of concepts at one time or possibly even use alternative techniques such as card sorting techniques to collect ratings
- *Make the Products of the Assessment Useful to the Students:* Many students did not understand the meaning of the similarity rating, the graph representation, or the feedback. Better descriptions of these products as well as better representations of the graphs will add to the value to the entire technique.
- *Validate the process more extensively.* In particular, it will be valuable to determine whether this method can capture learning within a class and learning of design knowledge. In our study, we only explore the ability of the technique to discriminate levels of expertise of students who had completed different numbers of entire courses in statistics. We would like to know if structural assessment is capable of discriminating levels of learning of students within a single term.
- *Develop classroom activities in which the structural assessment process is embedded in the activity.* As designed, structural assessment is very time consuming for students with potentially little learning benefit. We see great potential for the use of structural assessment as an introductory class activity -

an activity designed to get students to think, in advance, about the entire set of topics to be covered in a class and how these topics are related. In follow-up assignments, students could go through the assessment process again encouraging them to continue trying to think of how the class concepts are related.

BIBLIOGRAPHY

1. Astin, A.W., *Assessment for Excellence: The Philosophy and Practice of Assessment and Evaluation in Higher Education*. 1993, Phoenix, AR: Oryx Press. 335.
2. Chase, W.G. and H.A. Simon, *Perception in Chess*. *Cognitive Psychology*, 1973. **4**: p. 55-81.
3. Chi, M.T.H., P.J. Feltovich, and R. Glaser, *Categorization and Representation of Physics Problems by Experts and Novices*. *Cognitive Science*, 1981. **5**: p. 121-152.
4. McKeithen, K.B., et al., *Knowledge Organization and Skill Differences in Computer Programmers*. *Cognitive Psychology*, 1981. **13**: p. 307-325.
5. Jonassen, D.H., K. Beissner, and M. Yacci, *Structural knowledge : techniques for representing, conveying, and acquiring structural knowledge*. 1993, Hillsdale, NJ: Lawrence Erlbaum Associates.
6. Diekhoff, G.M., *Testing through Relationship Judgments*. *Journal of Educational Psychology*, 1983. **75**(2): p. 227-233.
7. Ruiz-Primo, M.A. and R.J. Shavelson. *Concept Maps as Potential Alternative Assessments in Science*. in *AERA Annual Meeting*. 1995. San Francisco, CA.
8. Nisbett, R.E. and t.D. Wilson, *Telling More Than We Can Know: Verbal Reports on Mental Processes*. *Psychological Review*, 1977. **84**(3): p. 231-259.
9. Kraiger, K., E. Salas, and J.A. Cannon-Bowers, *Measuring Knowledge Organization as a Method for Assessing Learning during Training*. *Human Factors*, 1995. **37**(4): p. 804-816.
10. Dearholt, D.W. and R.W. Schvaneveldt, *Properties of Pathfinder Networks*, in *Pathfinder Associative Networks: Studies in Knowledge Organization*, R.W. Schvaneveldt, Editor. 1990, ABLEX: Norwood, NJ. p. 1-30.
11. Schvaneveldt, R.W., ed. *Pathfinder Associative Networks: Studies in Knowledge Organization*. . 1990, ABLEX Publishing Corporation: Norwood, NJ.
12. Krushkal, J.B., *Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis*. *Psychometrika*, 1964. **29**: p. 1-27.
13. Gillan, D.J., S.D. Breedin, and N.J. Cooke, *Network and multidimensional representations of declarative knowledge of human-computer interface design experts*. *International Journal of Man-Machine Studies*, 1992. **36**: p. 587-615.
14. Cooke, N.J. and R.W. Schvaneveldt, *Effects of computer programming experience on network representations of abstract programming concepts*. *International Journal of Man-Machine Studies*, 1988. **29**: p. 407-427.
15. Schvaneveldt, R.W., et al., *Measuring the structure of expertise*. *International Journal of Man-Machine Studies*, 1985. **23**: p. 699-728.
16. Goldsmith, T.E., P.J. Johnson, and W.H. Acton, *Assessing structural knowledge*. *Journal of Educational Psychology*, 1991. **83**: p. 88-96.
17. Schvandevedlt, R., *PCKNOT*, . 1994: Las Cruces, NM.

JENNIFER TURNS

Jennifer Turns is a research assistant in the Center for Human-Machine Systems Research in Georgia Tech's Department of Industrial and Systems Engineering. She received her B.S. and M.S. in Systems Engineering from the University of Virginia. She is broadly interested in both cognitive and process issues associated with engineering design education. She is a member of ASEE and AERA.

ALEX KIRLIK

Alex Kirlik is a assistant professor in the Center for Human-Machine Systems Research in Georgia Tech's Department of Industrial and Systems Engineering.