Measuring Engineering Students' Intellectual Development Using Neural Network and Expert System Technology

Ronald L. Miller, Barbara M. Olds, Michael J. Pavelich
Colorado School of Mines

Summary

Students completing an undergraduate engineering degree are expected to develop intellectually in addition to acquiring specific engineering knowledge and skills. However, effectively measuring intellectual development involves a time-consuming and expensive interview conducted and evaluated by trained human experts. In order to develop a quick and inexpensive alternative method for making these measurements, we are writing a software package based on neural network and expert system technology to emulate the interview and evaluation process. If successful, the software will allow engineering programs to rapidly and reliably measure the intellectual development of their students as a formative and summative assessment tool. This paper describes our progress on the project and remaining research questions under investigation.

Introduction and Background

Most engineering programs expect that their students will develop intellectually in addition to acquiring knowledge and skills in a specific engineering discipline. However, nearly all measures of student achievement are focused on content knowledge, process ability (e.g. design), or communication skills; students are assumed to be developing intellectually, especially in their ability to think critically, but rarely are meaningful data collected and reported which support such an assumption.

Using the techniques presently available to us, measuring intellectual development is difficult, time-consuming, and expensive. However, the recent movement towards outcomes assessment now requires reliable measures of students' abilities to make reasoned decisions as they solve complex problems. For example, ABET requires institutions to develop assessment processes which can demonstrate "that the outcomes important to the mission of the institution and the objectives of the program are being measured." [1]

Numerous pencil-and-paper test instruments including the Watson-Glaser Critical Thinking Appraisal [2] and California Critical Thinking Skills Test [3] purport to measure some aspect of intellectual development or ability to think critically. These types of tests are typically inexpensive and easy to administer, but their validity in measuring true intellectual development and thinking ability is questionable because pencil-and-paper instruments rely on close-ended questions with one right answer; no information is collected describing how or why the student chose a particular answer and no mechanism exists to adapt exam questions based on prior responses from the student.

Perhaps the most recognized and valid method to quantify maturation of college students' intellectual abilities relies on developmental process models such as Perry's model of intellectual and ethical development [4] and King and Kitchener's Reflective Judgment model [5]. These models measure students' positions along a hierarchical construct of stages representing increasingly more sophisticated ways of understanding and solving complex problems. A student's position on the Perry or Reflective Judgment model scales is measured using one of three techniques: 1) a videotaped or audiotaped interactive interview conducted by a trained expert, and evaluated by a second trained expert, 2) a written essay exam scored by a trained expert, or 3) a multiple choice examination. Experts [5,6] generally agree that the interactive interview is the most reliable measure of position on the Perry or Reflective Judgment model scales and, despite significant work by many researchers, no acceptable pencil-and-paper examination has been developed which provides an educationally useful statistical correlation to interview results (correlation coefficients typically do not exceed 0.4). [7] Since conducting reliable interviews is time-consuming (about three hours for the interview plus scoring) and expensive (about $50-$150 per student), the process is not consistently used as an institutional or programmatic assessment tool.

The success of the interview method relies on the ability of the interviewer to probe for evidence of a student's thought processes and decision-making strategies. No static pencil-and-paper test instrument can search for such evidence, but we believe that neural network and expert system computer technology may be used to develop software which replaces the role of the expert interviewer and evaluator. In this paper, we briefly describe two of the most prominent intellectual development models used with engineering students and how the interview process is used to evaluate a student's position on the model scales. We then describe the characteristics of neural networks and expert systems which we are employing to develop software for measuring intellectual development.

The Perry and Reflective Judgment Models of Intellectual Development

William G. Perry, Jr. developed his model from clinical studies of Harvard students in the 1970's. [4] As he interviewed a group of students at the end of each academic year, probing their views of their university experiences, he observed patterns of thinking that were hierarchical and chronological. These patterns described an intellectual development path that all students seemed to follow and that Perry translated into a nine-stage model of development that he validated by a second, more extensive, longitudinal study.

The model, a portion of which is summarized in Table 1, describes the stages students pass through as they mature in their understanding of the nature of knowledge, use of evidence, and open-ended problem solving. For example, students at Perry position 2 believe that all questions have single right answers and, thus, no problem is "open-ended." They often view professionals who admit to not knowing an answer as incompetent. Students at position 4 understand that there are legitimate unknowns and uncertainties, even in science and engineering, and they do use evidence well. However, they feel that there are no legitimate ways to weigh alternative possibilities, and, thus, all solutions are equally valid and "everything is relative." Therefore, students at position 4 see no reason to explore alternatives before reaching a decision because one well-argued possibility is sufficient. At position 6, students understand the need to use evidence and explore alternatives when solving an open-ended problem, the need for judgments based on personal and articulated standards, and the need to be open to changing circumstances.

The Reflective Judgment (RJ) model developed in the early 1980's by Patricia M. King and Karen S. Kitchener resembles the Perry model in most respects. [5] In fact, the models are nearly identical through position 4. At position 5 and above, the two models focus on slightly different aspects of complex thinking: the Perry model searches for commitment to action based on articulated values, while the RJ model searches for integration of reasoning between disparate domains of thought. The RJ model has the advantage of a more substantial research history and more precisely articulated and documented interview/rating protocols. Both models are helpful frameworks within which to develop software to measure intellectual development of engineering students.

Table 1 - Summary of Perry Model Positions 2, 4, and 6

Position 2	dualist--ideas are seen as right or wrong; authority has all the answers; use of evidence is not understood; ambiguity in knowledge is a shortcoming or a game played to get THE answer
Position 4	ambiguity is legitimate, but vexing; uses evidence, but without trust; no need to consider alternatives; "all opinions are equally valid"
Position 6	ambiguity is common to most questions; evidence is used to explore alternatives; seeks the better answer in context

<u>Measuring intellectual development with the Perry and RJ models</u>. Currently, the only universally accepted measure of a student's position on the Perry or RJ developmental scales is an extensive interview of the student by a trained interviewer and an evaluation of that interview, using transcripts of videotapes or audiotapes, by another trained professional. Results from pencil-and-paper tests designed to measure intellectual development have been disappointing, showing correlation coefficients of only about 0.4 with interview results. [7-9] To illustrate the complexity of the interview process, consider the following quotations from actual engineering student interviews. Given a scenario in which a hypothetical mountain town could gain economic health at the cost of polluting the town's stream, P., a sophomore, reasons:

> I would vote to not let them build the plant because it might affect the beauty of the town, but I see [the other] point too. It is not really a question of right or wrong. It's what you think is best for the community and that comes down to your personal preference.

P. is comfortable with multiple possible answers, but views the solution as something of a "coin flip," a personal preference. After further probing about this scenario and others, her thinking was judged to be at approximately position 4 on Perry's scale.

Student R., in contrast, was judged to be near Perry position 6 based on explanations such as the following:

> A good decision makes sense technically, but also makes sense politically, economically and socially. A bad decision considers only the technical aspect or only the political aspect. You don't just build a technical project and it doesn't just function in a technical world. It functions in a broader world of political and social values. If you concentrate on any single [aspect], it is not as good a decision as if you concentrate on all aspects and try to come to some agreement where all your needs are met to the best of your ability.

An expert interviewer elicits how a person is thinking and why he/she reached a particular decision. These insights allow the rater to determine the developmental level at which the participant thinks, unlike paper-and-pencil tasks which invariably are only able to determine what decision is reached. We believe that neural network and expert system software can be developed which will more closely emulate the expert interview by leading a student through a series of scenarios and questions while making elementary decisions about the participant's thought and problem-solving processes.

Using Neural Networks and Expert Systems to Measure Intellectual Development

As the brief excerpts in the previous section demonstrate, an intellectual development interview consists of rich and complex responses to questions from the interviewer, who must make reasoned decisions about how and where to probe for additional explanation and elaboration of the student's thoughts on each scenario posed during the session. The evaluation expert must then search the interview transcript for evidence of intellectual processes which indicate where the student is positioned along the Perry or RJ model scales. Typically, a student receives three rating scores for a one hour interview session; for example, a rating of {3,3,4} indicates a student who is generally positioned at Perry level 3 but who also demonstrated some level 4 thinking. Thus, assessing intellectual development is not an objective measurement easily adapted to close-ended questions or traditional algorithmic computer software.

Our software contains several features designed to emulate the interview and evaluation process including:

- Use of open-ended scenarios similar to those posed in Perry and RJ interviews
- Sample responses extracted from actual interviews for software users to respond to
- Use of neural network technology to analyze complex student response patterns
- Use of expert system technology to decide which follow-on scenarios should be posed based on the current status of the session and prior student responses.

An example scenario is shown in Table 2. [adapted from ref. 5] Our software presents the scenario to the student user via a Visual Basic graphics user interface along with several statements to a posed question that the user is asked to agree or disagree with using a 1-5 Likert scale (1 = no agreement, 3 = some agreement, and 5 = great amount of agreement). User responses to the posed questions provide us with a pattern which is then analyzed using a simple neural network to compute the predicted intellectual development level. Neural nets, a computerized attempt to emulate human though processes and decision-making, are particularly effective at recognizing and analyzing complex patterns with subtle features [10,11] and are working well with the scenarios we have written and tested so far.

The key to successful neural net performance is obtaining a comprehensive and valid data set consisting of responses to each scenario statement provided by students of known intellectual development level obtained using the traditional interview process. These data are then used to train the neural net so that it can recognize pertinent pattern features for each scenario and accompanying statements. For a scenario with six statements, we require response data for approximately 20 students to adequately train a neural net consisting of 6 input nodes, 5 hidden nodes, and one output node. Training on a Pentium™ personal computer requires approximately 1000 iterations of the training dataset; typical computation time is 30-40 seconds to train the net.

After the student user has responded to 3-4 introductory scenarios and the trained neural net code has computed a value of intellectual development based on each scenario, we use a simple rules-based expert system [12,13] to adapt the session by deciding which (if any) follow-on scenarios should be posed to the student. Knowledge rules for the expert system focus on answering questions such as the following: Is the standard deviation for results from the introductory scenarios larger than acceptable? What follow-on scenarios should be used to collect new data to reduce the deviation? What follow-on scenario should be selected to help refine the student's initial intellectual development placement (to differentiate between two adjacent levels, for example)? Is there any indication that the student's responses are inconsistent? Are further scenarios required to determine if the student is providing genuine responses or guesses?

The alpha version of our software is currently being written in Visual Basic and testing with student volunteers has begun. Training of the neural nets is completed as each scenario is written and tested and development of knowledge rules for the expert system is underway. We expect to have a complete version of the software ready for field tests by January, 1999.

Table 2 - Sample Intellectual Development Scenario [adapted from ref. 5]

Most historians claim that the pyramids were built as tombs for kings by the ancient Egyptians using human labor and aided by ropes, pulleys, and rollers. Others have suggested that the Egyptians could not have built such huge structures by themselves, for they had neither the mathematical knowledge, the necessary tools, nor an adequate source of power.

When people differ about an opinion like this, is it the case that one opinion is right and one is wrong? For each statement below, indicate how well it conveys your response to this question.

Greatly	Much	Some	Little	None	Response
					Yes, because I believe that there are some other intelligent societies in the universe and it is possible that they came here in the past and helped build the pyramids.
					No, because some people have different information than others. If I had their information I might change my mind, although I think there is only a 30% chance that aliens did it.
					Yes, the experts have proven that you can use pulleys and such to make them. I don’t think you need mathematics.
					Yes, because we might find some eyewitness accounts that can prove it one way or the other.
					Yes, in the sense that something did happen back then. But “no” in the sense that we may never have enough information to be certain.
					No, I would not say as much right or wrong, but that one opinion has more basis and better evidence than the other.

The Future

As we continue to develop and test our software, several significant research questions remain to be answered including:

- Can we write scenarios and response statements which capture the complexity and dynamics of the traditional interview process?
- Will the neural network be able to identify valid student response patterns for each scenario?
- Will the expert system be able to adequately decide how to adapt the user session?
- Will intellectual development measurements using the software agree with interview results?
- Can the computer interface be designed to maintain student interest and avoid boredom?

If these questions can be addressed successfully, a valuable assessment tool will be readily available to engineering educators for monitoring the intellectual development of their students.

Acknowledgment

We wish to acknowledge the Fund for the Improvement of Postsecondary Education of the U.S. Department of Education for financial support of this project.

References Cited

[1] "Criteria for Accrediting Programs in Engineering," Accreditation Board for Engineering and Technology, Baltimore, MA, 1998 (available on ABET WWW homepage: www.abet.org).

[2] Watson, G., and E.M. Glaser, *Watson-Glaser Critical Thinking Appraisal, Forms A and B*, Psychological Corporation, New York, 1980.

[3] Facione, P.A., and N.C. Facione, *The California Critical Thinking Disposition Inventory*, The California Academic Press, Millbrae, California, 1992.

[4] Perry, W.G., Jr., *Forms of Intellectual and Ethical Development in the College Years*, Holt, Rinehart and Winston, Inc., New York, 1970.

[5] King, P.M. and K.S. Kitchener, *Developing Reflective Judgment*, Jossey-Bass Publishers, San Francisco, 1994.

[6] Pavelich, M.J., and W.S. Moore, “Measuring the Effect of Experiential Education Using the Perry Model,” *Journal of Engineering Education*, vol. 85, pp. 287-292, 1996.

[7] Baxter-Magolda, M.B., “Comparing Open-Ended Interview and Standardized Measures of Intellectual Development,” *Journal of College Student Personnel*, vol. 28, pp. 443-448, 1987.

[8] Pavelich, M.J. and P. Fitch, “Measuring Students’ Development Using the Perry Model,” *Proceedings of the American Society for Engineering Education Annual Conference*, Washington, DC, pp. 668-672, 1988.

[9] Culver, R.S., P. Cox, J. Sharp, and A. Fitzgibbon, “Student Learning Profiles in Two Innovative Honours Degree Engineering Programmes,” *International Journal of Technology and Design Education*, vol. 4, pp. 257-287, 1994.

[10] Mehrotra, K., C.K. Mohan, and S. Ranka, *Elements of Artificial Neural Networks*, MIT Press, Cambridge, Massachusetts, 1997.

[11] Eberhart, R., P. Simpson, and R. Dobbins, *Computational Intelligence PC Tools*, Academic Press, Inc., Boston, 1996.

[12] Durkin, J., *Expert Systems Design and Development*, Macmillan Publishing Company, New York, 1994.

[13] Hart, A., *Knowledge Acquisition for Expert Systems*, 2nd ed., McGraw-Hill, Inc., New York, 1992.

Biographical Information

RONALD L. MILLER is associate professor of chemical engineering at the Colorado School of Mines. His research interests include applications of constructivist mental modes in engineering education and authentic assessment of student learning.

BARBARA M. OLDS is professor of liberal arts and international studies and principal tutor of the McBride Honors Program in Public Affairs for Engineers at the Colorado School of Mines. Her research interests include methods for student outcomes assessment and integrated uses of writing in engineering education.

MICHAEL J. PAVELICH is professor of chemistry and Director of the Office of Teaching Effectiveness at the Colorado School of Mines. His research interests include intellectual development measurements as an assessment tool and use of computer technology in science education.