MEASURING THE INTELLECTUAL DEVELOPMENT OF STUDENTS USING INTELLIGENT ASSESSMENT SOFTWARE

Barbara M. Olds[1], Ronald L. Miller[2] and Michael J. Pavelich[3]

Abstract *¾ Most methods currently available to measure intellectual development in college students are either unreliable or prohibitively expensive and time consuming. To circumvent these difficulties, we have developed Cogito, a software package which uses a neural network to find patterns in "noisy" paper-and-pencil data and relate them to the Perry or Reflective Judgment models of intellectual development. If we are successful, Cogito will provide an inexpensive and reliable alternative to current interview or paper-and-pencil tests. Our work to date shows a high correlation between interview and Cogito results.*

Keywords ¾ intellectual development, assessment, neural network, Cogito©

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, in the U. S. the Accreditation Board for Engineering and Technology (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 (up to $150 per student), the process is not consistently used as an institutional or programmatic assessment tool.

[1] Barbara M. Olds, Colorado School of Mines, Liberal Arts and International Studies Dept, bolds@mines.edu
[2] Ronald L. Miller, Colorado School of Mines, Chemical Engineering Dept, rlmiller@mines.edu
[3] Michael J. Pavelich, Colorado School of Mines, Chemistry Dept, mpavelic@mines.edu

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

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 networks may be used to find patterns in "noisy" data and thus predict levels of intellectual development in subjects. In this paper, we briefly describe the Perry and Reflective Judgment models of intellectual development. We then describe the software, Cogito, which we have created 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 1960'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.

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 no greater than 0.4 with interview results[7-9].

USING SOFTWARE TO MEASURE INTELLECTUAL DEVELOPMENT

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 per interview scenario; for example, a rating of {3,3,4} indicates a student who is generally positioned at Perry or RJ 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, Cogito, 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

Our software presents the scenario to the student user via a graphics user interface along with several statements to a posed question that the student 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). Student 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 2-7 on the Perry or RJ scales. Neural nets, a computerized attempt to emulate human thought 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 5-6 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 3000 iterations of the training data set; typical computation time is 1-3 minutes to train each net.

SOFTWARE DEVELOPMENT AND VALIDATION RESULTS TO DATE

Cogito uses open-ended scenarios based on controversial topics as the mechanism for collecting student responses for scoring by trained neural networks. The four scenarios listed in Table 2 have been written and tested in the first version of Cogito. In each case, the topic focuses on an open-ended dilemma or controversy with posed questions for which test subjects provide responses during an interactive software session. Just as with the traditional interview method, we are not interested in a student's solution to the posed dilemma, but in <u>how</u> he or she has developed a solution and what evidence (if any) is used to support the solution. Response fields in the computerized scenarios have been carefully written to differentiate this type of information.

Response data using Cogito have been and continue to be collected from high school students, college undergraduates, graduate students, and college faculty (results for 40 subjects have been collected and analyzed so far and we will have collected an additional 40-50 data points by July, 2000). The "correct" intellectual development level of each subject is also determined using the traditional interview process. These results are then randomly subdivided into two groups: 1) data to train neural networks for each scenario, and 2) data to verify predictions from the trained software.

Figure 1 shows an example of the results we collect with Cogito. We are able to successfully train neural nets over the entire range of intellectual development studied (training means that the neural net algorithm can successfully find patterns in student response data when statistically compared to known student intellectual development levels measured by traditional interviews). The correlation coefficient for the training set shown in Figure 1 exceeds 0.80. More importantly, the trained net is able to predict verification data points with a correlation coefficient of 0.75 for data below level 5 and provides predictions within the precision of the interview method itself (approximately 0.3 levels). However, above level 5, our predicted results show more scatter than expected. We have seen a similar trend with each scenario created and tested so far.

Table 2. Scenario topics included in first version of Cogito software

Topic	Dilemma or Controversy
Overpopulation	Is overpopulation a significant problem in the world?
College education	Describe what a college education should do for a student – educate for life or train for a job?
tax rebates	Who should get tax cuts – rich or poor people?
nitrate contamination	How might nitrates in groundwater be controlled?

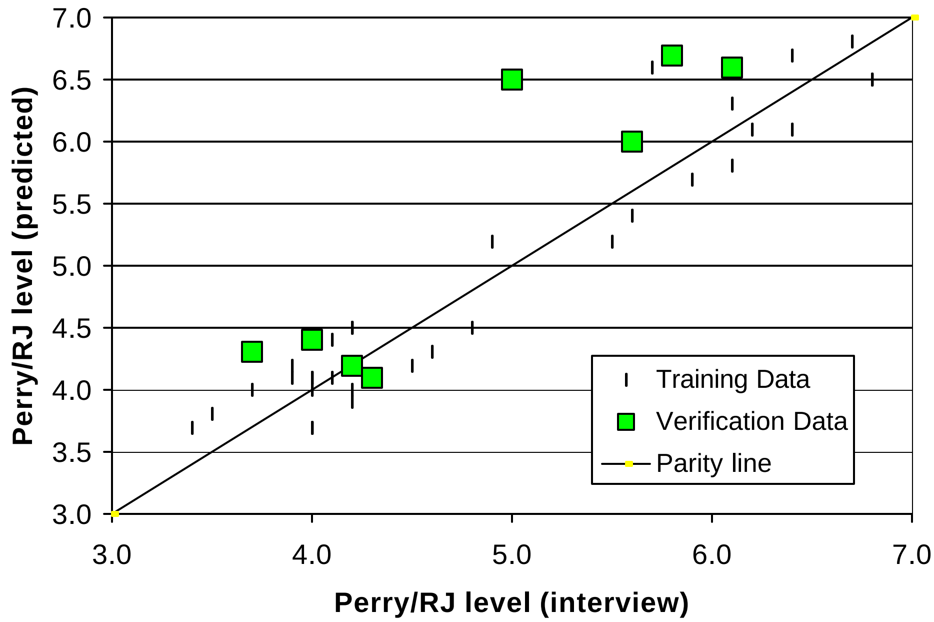

Fig. 1. Comparison of computer-predicted and interview intellectual development results (overpopulation scenario, training set 1.1)

Recently we have developed a new statistical technique to help analyze and correlate the data we've collected by providing more rigorous predictions of intellectual development levels from Cogito data. The technique involves randomly sorting the overall data set for each scenario into 4-6 training data/verification combinations, each of which is used to train a neural net. Because different combinations of data are included in each training set, we obtain several slightly different trained neural nets that provide slightly different correlations of student response data. Thus, we obtain not one estimate of intellectual development from each subject's response to a scenario, but as many as six, and we are then able to use standard statistical techniques to estimate the upper and lower bound on a student's intellectual development level rather than computing and reporting only one numerical value.

An example of this technique is shown in Table 3 in which a student was judged to be at an intellectual development level of 4.2 based on a traditional interview. The student responses to three scenarios posed by Cogito (overpopulation, education, and nitrate contamination) were used by a total of 11 trained neural nets to produce a range of estimated intellectual development level. Two of the estimates (5.6, 6.0) were identified as statistical outliers and eliminated. We then used the remaining 9 data points to compute a 95% confidence interval of 4.1 to 4.5 indicating that with less than a 5% chance of being wrong, the student's intellectual development level ranged from a low of 4.1 to a high of 4.5 which agrees very well with the interview result.

NEXT STEPS

As we continue to develop our Cogito software, we have several tasks planned to improve our process and product.

- We re-train the neural net on Cogito using 80-90 data points collected to see if fits are maintained
- We will develop new scenario topics using proven designs for pre/post-testing purposes.
- We will conduct interviews with 60 additional subjects (40-50% at the 5+ level) to acquire data for neural net training and validation of new scenarios and to demonstrate that pre-test and post-test scenarios provide statistically equivalent results.
- We will update Cogito's graphical interface based on feedback from the alpha version.
- We will continue to disseminate results of the project and continue to form partnerships with institutions such as Rose-Hulman Institute of Technology, the University of Washington, and the

University of Pittsburgh who are pilot testing the software for institutional and program assessment purposes.

Table 3. Summary of statistical prediction of intellectual development level based on response data from overpopulation, education, and nitrate scenarios

subject:	C99-2				
Interview result:	4.2				
Overpopulation scenario results:		5.6	4.1	6.0	
Education scenario results:		4.1	4.1	4.1	4.3
Nitrate scenario results:		4.6	4.8	4.6	4.0
Statistical outlier points (t-test, 95% confidence):			5.6	6.0	
Computed 95% confidence interval excluding outlier points:				[4.1, 4.5]	

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.