Technology Closes the Gap between Students' Individual Skills and Background Differences

Elsa-Sofia Morote
Research Laboratory of Electronics, Massachusetts Institute of Technology, USA
Department of Education, Leadership and Technology, Dowling College, USA
morotee@dowling.edu

David E. Pritchard
Research Laboratory of Electronics, Massachusetts Institute of Technology, USA
dpritch@mit.edu

Abstract: We examine the correlation of 12 background variables determined from a student survey with assessment instruments including paper-testing instruments (final exam and weekly tests) and an electronic assessment from the results of a Web-based homework tutorial. Several of the initial background variables showed correlation with hand-graded weekly or final paper tests. Level of previous math and physics courses taken correlated with positive results on both. However, none of the background variables correlated with the student's performance on the more reliable Web-based tutorial homework score. On this basis we suggest that the use of un-timed Web-based tutorials can make an important contribution to assess student performance, without bias due to students' background differences.

Purpose

Assessment is a key component of the President's plan in the No Child Left Behind Act (2002) to ensure that all students receive an adequate education. Preferably, such assessment should be embedded in a normal course of instruction rather than be administered sporadically, as are examinations in most courses. In principle, the intelligent tutoring community can use tutoring data to determine students' levels of knowledge. Even better, a good assessment should let us know not only which students are falling behind, but also the major stumbling blocks to each student's progress (Sclafani, 2004). Any electronic tutor can be programmed to provide a detailed record of each student's interactions while working through the tutorial, and such data can be processed to provide a highly reliable assessment (Pritchard & Morote, 2002). This information can be used by the electronic tutorial to provide Just-in-Time feedback, which is the essence of Just-in-Time Teaching pedagogy, and to help students achieve mastery (Novak et .al., 1999; Gavrin, 2003).

The ultimate goal of Web-based homework tutorials is to teach scientific knowledge, a solid understanding of scientific concepts, and problem-solving skills that help students master a required level of knowledge regardless of their background. This paper concerns embedded assessment in a Socratic tutorial environment, specifically a Web-based homework tutorial that has been shown to offer a positive educational benefit (Morote & Pritchard, 2002). Similar results predicted by others' studies (Mestre, 2001, Novak et al., 1999). We contrasted tutoring data from a Web-based homework tutor with paper testing instruments such as weekly tests and a final exam. The students' responses in the Web-based homework tutor are "observed" by this existing sophisticated electronic tutor and scored by its algorithm. The principal result of this paper is that the average final scores from a tutorial environment of students with different backgrounds such as different genders and those with prior subject knowledge are equal. In contrast, paper-testing instruments such as a final exam or weekly tests are shown to differentiate based on students' initial backgrounds.

Data sources

This study was conducted using the framework of an introductory Newtonian Mechanics (8.01) course given in the spring semesters of 2001 and 2002 at the Massachusetts Institute of Technology. There were approximately 100 students in 2001 and 75 students in 2002. The pools of students in 2001 and 2002 were very similar, scoring within one-third of a point of each other on the Mechanics Baseline Pre-Test (Hestenes & Wells, 1992).

This course had been designed especially for the approximate 90% of students who had previously taken this course but failed to obtain a grade of C or better. The course utilized several instructional elements, including class participation (in recitations), hand-graded written homework, group problem-solving, weekly open tutorials for all students (those students doing poorly on the previous week's test were required to attend the tutorials), and an electronic homework tutor (myCyberTutor, EET, 2003). Paper testing included weekly tests and a final exam which evenly reviews the entire course; all were hand-graded paper tests with some partial credit given for incorrect answers.

A typical student interacted with myCyberTutor approximately 48 hours/semester, averaging 10-20 minutes on each problem. Each problem involves approximately 11 interactions between student and computer. These were logged with a time stamp, giving 11 timed reference points along the student's route to the solution. In contrast, the 12 weekly tests consumed a total of 4.5 hours, and the final exam length for both years was 3 hours.These three assessment instruments were considered in this study:

Weekly test (WT). Students were given weekly hand-graded written tests on the topics of that week. These were timed tests, but with sufficient time only about 10% did not finish early. The weekly test average for the whole term is designated as WT01 for the Spring 2001 semester and WT02 for the Spring 2002.

Final exam (FINAL). This is a timed final test. We note in passing that this particular final exam had three distinct segments. A standard test, Mechanics Baseline test (Hestenes & Wells, 1992), was used for one-third of the final. The remainder was divided into basic skill and complex problems which accentuate the performance of different students. The total test requested 47 specific responses from the students (compared with approximately 28 responses typical at MIT). Additional questions should make this final exam more reliable. Final01 designates the 2001 final and Final02 the 2002 final.

MyCyberTutor lost points algorithm (CT). The tutoring data was asssessed with an algorithm based on the "ad-hoc" generalization of a typical homework scheme in which points are awarded primarily for correct work. A small penalty was subtracted for submitting incorrect answers as well as for requesting hints to discourage students from requesting hints and subproblems without thinking first. The 2001 algorithm, CT01, depended on the number of correct responses minus solutions requested, and provided a bonus of 0.03 points for each unopened hint. In addition to correct responses, solutions requested, and hints, in the Spring 2002, the CT02 algorithm also deducted 0.03 points for incorrect responses that received useful feedback.

Previous studies show that students' backgrounds have some influence on their course performance. Therefore, myCyberTutor asks twelve questions in an initial survey (Table 1). No attempt has been made to use this data to improve students' assessment during the term, although it would probably be valuable for this.
Background variables are divided in four: level of knowledge, technological experience, demographic and others:

Level of knowledge
The intention was to learn about students' initial level of knowledge. They were asked for their level of experience in specific subject areas:
Algebra. Students chose one of these options regarding their knowledge of Algebra: a) No experience b) 1 year of high school Algebra c) 2 years of high school Algebra d) 3 years of high school Algebra
Physics. Students chose on of the following options about their experience with Physics: a) No experience b) 1 year of high school Physics c) 2 years of high school Physics, c) 3 years of high school Physics d) 1 semester of college Physics e) 2 semesters of college Physics f) More than 2 semesters of college Physics
Calculus. a) No experience b) High School Calculus c) High School AP Calculus d) College Differential Calculus e) College Integral Calculus f) Multivariate Calculus
TriGeo. Students were asked for their level of Geometry/Trigonometry, they had to choose one of these options: a) No experience b) High School Geometry c) High School Trigonometry d) High School Analytic/ Solid Geometry
Math. Students answered the following question: "Which math course are you taking this semester?" Their answer was scored from 1 (minimum level in class) to 5 (maximum level in class) depending of the course mentioned.
Technological experience
Access. We wanted to know if access to a computer will influence performance. On a scale of 1-5, rate your level of access to a computer (1=easy access, 5=difficult access).

OtherE. We were interested in learning if the students already had experience with Web-based learning systems. They chose one of the following options: a) no experience b) some experience c) extensive experience
Comfort. This variable is intended to learn about the reaction of the students to Web-based systems. Students answered “How comfortable would you say you are with the World Wide Web?” they chose one of these options: a) very comfortable b) comfortable c) not comfortable

Demographic
Demographic data were included such as: “***age***” *and* “***gender***”

Other
Level. Students were asked for their current grade level : a) college freshman b) college sophomore c) college junior d) college senior e) high school freshman or sophomore f) high school junior g) high school senior
Major. Students were asked for their current or intended major.

Variables	**Description**
Level of knowledge	
Algebra	previous level of Algebra
Physics	previous level of Physics courses taken
Calculus	previous level of Calculus
TriGeo	previous level of Trigonometry and Geometry
Math	MIT Math course currently enrolled in
Technological experience	
OtherE	other experiences with Web-based educational systems
Access	access to computer
Comfort	comfort with Web-based systems
Demographic	
Age	Age
Gender	female or male
Others	
Grade level	i.e., college freshman, junior; or high school freshman, sophomore
Major	planned major

Table 1. Background Variables

Methods
Data transformation
The initial survey yielded two types of data, nominal and ordinal.
Ordinal data. To measure the level of experience in a particular subject we collected ordinal data such as the level of Algebra, Physics, Calculus, Trigeo and Math. For example, we asked students for their level of experience in Calculus; a student can choose one of the following options a) no experience, b) high school Calculus, c) high school AP Calculus, d) college Differential Calculus, e) college Integral Calculus, and f) Multivariate Calculus. We coded these responses from 1 (no experience) to 6 (Multivariate Calculus), then renormalized them on a 0 to 1 scale. Additionally, the survey provides interval data presented in a scale that allows us to place the response options in order. Variables such as access, otherE and comfort level are presented in this way. For instance, we asked for the access level to a computer; a student can choose an option from an interval of 1 (easy access) to 5 (very difficult access).
Nominal data. Gender and major are representatives of nominal data. The coding is completely arbitrary.

Determining relationship
The relationship between background variables (from Table 1) and student performance on more standard assessments (weekly test and final exam) was determined in this study as well as the relationship between these background variables and performance on a Web-based homework tutor, myCyberTutor. To measure the

relationship, we performed an analysis of variance (in the case of nominal data such a major and gender) and correlation studies:

Correlation analysis. To Correlation is a statistical technique which can show whether and how strongly pairs of variables are related. To measure the correlation of the continuous variables with our assessment instruments, we used Pearson's formula that tests for a linear relationship. The significance of the correlations was evaluated using Student's t-test, which allows us to calculate a probability-value - the probability that the observed correlation arises solely due to random fluctuations.

Analysis of the variance. A different analysis was used to find the relationship between the testing and background variables, such as planned major and gender. An analysis of the variance (ANOVA) test (comparing the means (μ)) was used to examine connections between major and gender, and the assessment instruments. We want to know if means of the groups differ. The null hypothesis in the case of gender is H_o: $\mu_{female} = \mu_{male}$. In the case of major H_o: $\mu_{engineering} = \mu_{biology} = \mu_{physics} = \mu_{mathematics}$... The F ratio (Wright, 1997, p. 123-125) and the p-value were carried out to test the relationship's significance. The significance of F depends on the degrees of freedom of the model and of the residuals.

Combining the results for several years. Some correlations showed significance in one year but not the other, but we can see a tendency if we calculate a p-value which combines both years using the Fisher (1950) concept. He showed that -2 ln(p) is distributed as a χ with 2 degrees of freedom. Then, we can add these χ together to produce a combined result. Since we are combining 2 years, each with 2 degrees of freedom for the χ^2, we have to tabulate the critical value for a χ^2 with 4 degrees of freedom at the desired level (e.g. 1%, 5% or 10%) (See Steel, Torrie and Dickey, 1996).

Results

Table 2 shows the correlation results in both years; r is the correlation coefficient calculated for assessment instruments (final exam, weekly tests, and myCyberTutor algorithm) against the background variables. A t-statistic was calculated for r, and its corresponding probability value (p) is presented in the table to measure significance. The suffixes 01 and 02 after r and p represent the years of calculations 2002 and 2002. To calculate the probability (p-value) for both years of significant correlation between assessment instruments and background variables, we add -2ln(p01) to -2ln(p02) (Table 2). We only show those variables that had significant correlation in any of the years.

Regarding the initial level of knowledge, the variables Algebra (algebra) and Trigonometry and Geometry (TriGeo) did not show significant correlation with any of the assessment instruments. However, the initial level of knowledge of Physics, Calculus, and Math as well as the paper-testing instruments (final exam and weekly tests) presented a significant correlation in at least one of the years. For example, the previous level of Physics is significantly correlated with the final exam in 2001 and 2002 (r01 = 0.27, r02 = 0.33) (Table 2), a result that does not usually emerge from other studies at MIT. In the same way, previous knowledge of Calculus and Math also helped with paper testing instruments. All of the level of knowledge variables seem irrelevant to the performance on the electronic homework measured by the CT algorithm in both years. Considering both years, a significant correlation was found between the variables "Physics" and "Math" and weekly tests and final exam.

Technology experience variables such as "OtherE" and "Comfort" showed no correlation with assessment instruments. However, having access to a computer (Access) and the weekly tests showed a positive correlation. This significant correlation could be because the students who have easy access to a computer can use myCyberTutor more for practice and study purposes and then slightly improve their performance on the weekly tests. Similar to the case of knowledge variables, no technology variables showed correlation with CT.

Demographic variables such as "Gender" did not show a significant relationship with the assessment instruments. In the case of age, it showed correlation only in one year, in 2002 (Table 2). Nevertheless, considering the p-values of both years, it is concluded that age is not a defining influence variable. In the same way, correlations between CT and "Age" were not significant in both years.

The variable "Grade level" did not show correlation with any of the assessment instruments. An ANOVA test was done to analyze the relationship between planned major and the testing instruments. As shown in Table 2, we found no indication of a relationship between "major" and the weekly tests. Nevertheless, the ANOVA test determined a suggestive indication of a relationship in the case of final exam and planned major in both years. It is important to note that there is no difference between performances of different majors and the score obtained on myCyberTutor.

Variables	Final Exam Final01			Final02			Both years	
	r01	p01	Sig.	r02	p02	Sig.	-2ln(p01/p02)	Sig.[1]
Physics	0.27	0.01	**	0.33	0.01	**	18.42	**
Calculus	0.22	0.03	*	0.01	0.92		7.18	
Math	0.22	0.03	*	0.16	0.2		10.23	*
Access	0.04	0.7		-0.03	0.82		1.11	
Age	-0.03	0.74		0.25	0.04	*	7.04	
	F	p		*F*	p			
Major	1.99	0.08	*¹	2.1	0.08	*¹	10.10	*
	Weekly Tests WT01			WT02			Both years	
	r01	p01	Sig.	r02	p02	Sig.	-2ln(p01/p02)	Sig.
Physics	0.16	0.12		0.18	0.15		8.03	*¹
Calculus	0.21	0.04	*	0.03	0.78		6.93	
Math	0.31	0.00	**	0.02	0.87		13.03	*
Access	0.21	0.04	*	0.19	0.12		10.68	*
Age	0.03	0.77		0.24	0.05	*	6.51	
	F	p		*F*	p			
Major	1.29	0.28		1.9	0.38		4.48	
	MyCyberTutor CT01			CT02			Both years	
	r01	p01	Sig.	r02	p02	Sig.	-2ln(p01/p02)	Sig.
Physics	0.1	0.33		0.06	0.61		3.21	
Calculus	0.06	0.56		0.14	0.26		3.85	
Math	0.06	0.55		0.14	0.25		3.97	
Access	-0.14	0.18		0.01	0.96		3.51	
Age	-0.07	0.48		0.06	0.64		2.36	
	F	p		*F*	p			
major	0.28	0.92		0.99	0.44		1.81	

** Correlation is significant at the 0.01 level
* Correlation is significant at the 0.05 level
*¹ Correlation is significant at the 0.10 level

Table 2. Correlations between the Background Variables and the Final Exam, Weekly Tests and myCyberTutor algorithm (Spring 2001 and 2002) *(only those that showed significance in any of the years)*

Conclusion and Discussion

In conclusion, the only variables that were consistently significant correlated both years with the paper testing instruments (final exam and weekly tests), were "Physics" and "Math". Students with a different planned major showed a difference on the final exam but not the weekly tests, and having access to a computer correlates with better performance on the weekly tests, but not on the final exam. Significantly, discrimination of students due to their initial background is not shown on electronic homework since none of the background variables correlated with CT.

None of the background variables have a connection with performance on the Web-based tutorial. This finding is particularly exciting because the ultimate goal of a Web-based tutorial is to teach scientific knowledge - a solid understanding of scientific concepts, and problem-solving skills - regardless of students' backgrounds. This could

[1] The critical, tabulated value for a χ^2 with 4 degrees of freedom at the 1%, 5% and 10% level is 13.2, 9.5 and 7.8, respectively.

be one more argument against high-stake examinations, where problems such as gender discrimination and ethnic background discrimination (Wilson, 2004) are known to be correlated with test results. Thus, if we were to rely solely on the standard criteria for e.g., college admission, many children from disadvantaged backgrounds would be denied admission in favor of students with strong background knowledge who are not affected by economic, social, and gender differences, and who, therefore, tend to score higher on these conventional tests.

It is important to note that both final exam and weekly tests are “timed tests”. Timed tests usually assess how fast students can go against how much they really know. However, when you remove the speed demand on the test-taker, the test can truly measure a student’s ability and knowledge (Radichel, 2003). Thus, time restriction on a test could contribute to differentiating students’ backgrounds. A Web-based tutorial homework, such as MyCyberTutor is a powerful tool that provides insight about students’ knowledge and allows them to learn at their own pace. This may be a contributing factor that contributes to closing the gap between students with different prior knowledge and demographic background.

Bibliography

Atherton, J. (2001) Direct Observation: Assessment [on-line] URL: http://www.staff.dmu.ac.uk/~jamesa/teaching/assessment_direct.htm (8/15/2001)

Effective Education Technologies (2002-2004). myCyberTutor. [on-line] http:// www.mycybertutor.com (01/10/2004)

Fisher (1950).*Contributions to Mathematical Statistics.* John Wiley, New York

Gavrin, M. (1999-2003). Just-in-Time-Teaching. [on-line] http://webphysics.iupui.edu/jitt/jitt.html (retrieved 01/10/2004)

Hestenes, D., & Wells, M. (1992, March). A mechanics baseline test. *The Physics Teacher, 30,* 159-166.

Mestre, J., Dufrense, R., Hart, D., & Rath, K. (2000). The effect of web-based homework on test performance in large enrollment introductory physics courses. *Journal of Computers in Mathematics and Science Teaching*

Morote, E.S. & Pritchard, D. (2002). What Course Elements Correlate with Improvement on Tests in Introductory Newtonian Mechanics? *Educational Resources Information Center (ERIC). ED463979.* http://www.eric.ed.gov/index.html

Novak, G., Patterson, E., & Gavrin, A. (1999). *Just-in-Time Teaching: Blending active learning with web technology.* Prentice Hall Series in Educational Innovation.

Pritchard, D., & Morote, E.S. (2002). Reliable Assessment with Cybertutor, a Web-Based Homework Tutor. *Proceedings of World Conference on E-learning in Corporate, Government, HealthCare, & Higher Education,* E-Learn 2002, Vol. 2002, Issue. 1, 2002, pp. 785-791. http://www.aace.org/DL/index.cfm/fuseaction/default

Radichel, M. (2003) *Taking time for types of test terms!* AGS Publishing Development. [on-line] URL: http://www.slpforum.com/cafe/10.asp (01/07/2004)

Sclafani, S. (2002). "No child left behind: assessment is a key component of the president's plan to ensure that all students receive an adequate education." *Issues in Science and Technology Winter 2002,* 43.

Steel, R., Torrie, J. & Dickey , D. (1996). Principles and Procedures of Statistics: A Biometrical Approach. McGraw Hill.

Wilson, W. (1991). "Affirming Opportunity." *The American Prospect Sept. 1999*, 61.

Wright, D. (1997). *Understanding statistics: An introduction for the Social Sciences*, SAGE publications, London , pp 71-72.