How We're Changing Computer Science Education and How You Can Help

Jeremy Straub, Scott Kerlin and Tom Stokke
Department of Computer Science
University of North Dakota
Grand Forks, ND 58202-9015
jeremy.straub@my.und.edu, skerlin@cs.und.edu, tstokke@cs.und.edu

Abstract

During 2013-2014, a study is ongoing in the Computer Science Department at the University of North Dakota. We're testing a technique of empirical objective evaluation of student learning in core courses. This initial work has focused on three entry-level courses (two of which are substantively similar, but use a different instruction language). We have created pre- and post-tests to quantify student performance relative to the ACM Model Curriculum areas that course instructors have selected. The questions include multiple formats (multiple-choice, true/false, short answer, etc.), are language neutral and have been selected (mostly, some had to be created or adapted to fill the requisite counts) from common test and quiz banks. Most importantly, the questions are not known to the instructors of the particular course, meaning that they cannot be taught to. We also asked students questions about their preparation, study habits and performance expectations.

This paper presents an overview of the initial results of this exploratory study and begins the process of assessing their meaning. We also discuss the larger study that this work is a part of, which will next include a North Dakota-wide assessment, using similar techniques to our single-institution exploratory study. The use of common courses between institutions in North Dakota facilitates the inter-institutional characterization required for further analysis. We conclude the paper by presenting an overview of the potential long-term value of this assessment work. Most notably, we assert that it will enable other types of research, by making quantitative comparison of year-to-year and local-versus-national performance easier, eliminating (or at least reducing) the need for complex mechanisms to try to establish experimental control. We also explain how the participation of other regional schools can greatly aid our effort to further this work.

1 Introduction

Most educational assessment work is complicated by the need to be able to control for the impact of the experiment in some way. The need for student choice, combined with year-to-year (or even semester-to-semester) student capabilities differences and other confounding factors makes conducting a valid educational experiment problematic. Some have responded by creating elaborate multi-school controlled experiments, while others have opted to report on their own work and leave the potential extrapolation of this work as a subject for others to investigate. However, if a well-known set of standards existed to compare performance against (for example using a pre-test, post-test approach which would allow a performance difference to be ascertained), conducting assessment on educational innovations (both large and minute) would be much easier.

The same information that can enable experimental work in educational assessment can facilitate formative assessment as well. In too many cases, the fundamental question of what (if anything) did the students learn and how much is replaced with surrogate questions about student perceptions of instruction and other (possibly correlating in some cases) topics. Being able to assess students' starting and ending points would enable a more effective assessment of what techniques, styles and other inputs into the learning process work well and which do not work as well.

We're working to develop these standards for the Computer Science discipline. A pilot project which was run during the Spring, 2014 semester at the University of North Dakota is conducting pre- and post-test based assessment of three 100-level computer science courses and one 300-level computer science course. We plan to expand this during the Fall, 2014 semester to incorporate other schools in North Dakota and beyond. This work seeks to generate the basics of a standard that can be used on a national or international scale.

2 Background

Evaluating the performance of courses, education approaches and educators is a subject that provokes no shortage of problems. Significant disagreement exists regarding how to achieve the best results for students, or even what results should be generated and assessed [1-3]. Others fear that determining an evaluative criteria may allow an administration to 'clean house' of those not subscribing to an approach. O'Mahony and Garavan [4] contend that this "managerialism" perception, the notion that university leadership is uses systems to manage the school in a business-like way and seeks to "advance strategic objectives" for some can be problematic. However, a robust approach, which considers knowledge, skill and experience attainment may identify numerous benefits and the trades that must be made to get each. This may include benefits beyond what are typically assessed, such as enhanced creativity [5], motivation and self-image [6] and job placement benefits [7].

Current policy makers perceive an ever growing cost of higher education [8] with generally positive results, but which suffers from a difficulty of de-confounding the selection effect (of who seeks to attend and is admitted to colleges) from the impact of the college's educational services [9]. Baum, Kurose and McPherson [9] proffer that value is being created; however, its characterization in the specific is elusive – even though student earnings differentials [9, 10] demonstrate the presence of significant value. Many metrics show U.S. education systems trailing behind other countries, across all levels (e.g., [11, 12]). However, these measures may exclude metrics (such as the hands on experience generated by project-based [13] and other experiential education techniques [14, 15]) under which the U.S. may perform more favorably. Alston, et al. [16] indicate that many of these other skills are key indicators of students ability to succeed in the workplace. Quite pragmatically, if the educational community doesn't take the lead in developing metrics for higher education institutional success, others may (as was the case with K-12 [17]) do so instead.

3 What Has Been Done

To-date, work in several areas has been performed. A set of surveys for multiple courses have been created and approved by the University of North Dakota's (UND's) institutional review board (IRB) for human subject research. Courses for which surveys have been crated include: (1) UND's CSCI-160 (Computer Science one), (2) UND's CSCI-161 (Computer Science two) and (3) UND's CSCI-130, a course for non-majors which covers similar topics to CSCI-160 using an alternate language. We have administered pre-tests in all sections of these courses during the Spring, 2014 and are in the process of administering the post-tests. We have also created a test for an introductory AI course which, while not material-specific, will facilitate comparison between the attitudes and beliefs of students in upper- and lower-division courses.

These surveys have been created by asking instructors of the courses to identify the areas of the ACM model curriculum which they cover. Questions were selected from several common test banks in most cases and a limited number were adapted or written from scratch, when suitable questions could not be readily identified. All questions were language agnostic. This allowed us to use largely the same pre- and post-tests for CSCI 160 and CSCI 130, despite the fact that they are taught in different languages. The primary difference between the two was the addition of a limited number of additional questions to the CSCI 130 test due to the course covering an additional area of the ACM model curriculum. Instructors were not shown the pre- and post-tests to prevent intentional or inadvertent focus of these particular question topics.

A combination of questions from two previously-used surveys about factors which might affect performance was also given. These questions request that students disclose various characteristics which can be correlated with performance on each individual (pre/post) test or longitudinally.

We have performed limited analysis on the pre-test results without considering the data collected from the limited set of post-tests given. This data has provided limited insight, but detailed analysis is ongoing.

4 Limited Initial Results

The data that has been analyzed, thus far, comes from the introductory artificial intelligence course. This data is presented in [18].

First, we assessed the correlation between age and expectations (Table 1).

Table 1. Correlation between age and expectations [18].

Age	Expected Difficulty	Writing Computer Programs	Working at Computer	Level of Understanding	Participation in Class	Participation in Lab	Office Hours
18-22	3.00	3.50	4.25	3.50	2.75	3.25	3.00
22-30	3.00	3.60	3.60	3.00	3.20	3.40	3.40
P-Val	0.50	0.40	0.14	0.09	0.18	0.38	0.24

We then assessed correlation between the academic year of students and their expectations (Table 2).

Table 2. Correlation between academic year and expectations [18].

Year	Expected Difficulty	Writing Computer Programs	Working at Computer	Level of Understanding	Participation in Class	Participation in Lab	Office Hours
Junior	3.00	3.60	4.00	3.40	2.80	3.40	3.00
Senior	3.00	4.00	4.00	3.25	3.50	3.50	3.25
P-Val	0.50	0.22	0.50	0.34	0.05	0.40	0.36

We then performed a similar correlation with the students' reported GPA (Table 3) and whether they were self-taught programmers or not (Table 4).

Table 3. Correlation between GPA and expectations [18].

GPA	Expected Difficulty	Writing Computer Programs	Working at Computer	Level of Understanding	Participation in Class	Participation in Lab	Office Hours
2.0-2.99	3.00	4.00	4.00	3.00	4.00	4.00	3.50
3.0-3.99	3.00	3.50	4.00	3.33	2.67	3.17	3.17
P-Val	0.50	0.04	0.50	0.09	0.00	0.02	0.43

Table 4. Correlation between self-learning programming and expected class activities [18].

Self-learned	Expected Difficulty	Writing Computer Programs	Working at Computer	Level of Understanding	Participation in Class	Participation in Lab	Office Hours
N	2.75	3.50	4.00	3.00	3.00	3.25	3.00
Y	3.17	3.83	4.00	3.50	3.00	3.33	3.33
P-Val	0.11	0.23	0.50	0.04	0.50	0.44	0.27

As is indicated by the p-value in the tables, a number of statistically significant results have been identified. Those significant at $p<0.05$ are shaded in green, while those only significant a $p<0.1$ and $p<0.15$ have been shaded in yellow and orange, respectively. There five significant correlations at $p<0.05$, two that are significant at $p<0.10$ and one that is only significant at $p<0.15$.

5 Next Steps

We are currently working to develop a similar examination for CSCI 242 (Computer Science three) which we plan to give at UND for the first time during the Fall, 2014 semester. We are also working to identifying and coordinate with other colleges and universities around the State of North Dakota to participate in an expanded trial during the Fall, 2014 semester, as well. The existence of commonly-defined courses in North Dakota enables this extension. We will still, however, ask each instructor to define what areas of the ACM Model Curriculum he or she is covering to ensure that only relevant areas are tested.

We are seeking participants from other areas around the country, as well. For those that opt to participate, we will create a customized examination based on the areas of the ACM Model Curriculum indicated as covered. We are slowly working to expand our question bank to include questions from all areas identified in the Model Curriculum.

6 Conclusion

This paper has presented an overview of work to-date on the development of a standardized assessment tool for Computer Science education. It has described current progress, presented limited results and described the planned next steps.

This nascent effort would greatly benefit from the participation of instructors of Computer Science courses everywhere. Through these increased numbers we can build a question set and performance dataset that can serve to enable future research, formative and evaluative assessment in computer science. This revised assessment model will be based on quantitative data about students' performance, based on a standardized set of criteria. The ability to quickly compare local performance, using this standard, to national and regional averages should facilitate expedient analysis and enable more and more expedient work in this area.

References

[1] A. Korzh. What are we educating our youth for? *European Education 45(1),* pp. 50-73. 2013.
[2] E. C. Lagemann and H. Lewis. *What is College for?: The Public Purpose of Higher Education* 2012.
[3] M. Tomlinson. Graduate employability: A review of conceptual and empirical themes. *Higher Education Policy 25(4),* pp. 407-431. 2012.

[4] K. O'Mahony and T. N. Garavan. Implementing a quality management framework in a higher education organisation: A case study. *Quality Assurance in Education 20(2),* pp. 184-200. 2012.
[5] A. Ayob, R. A. Majid, A. Hussain and M. M. Mustaffa. Creativity enhancement through experiential learning. *Advances in Natural and Applied Science 6(2),* pp. 94-99. 2012.
[6] Y. Doppelt. Implementation and assessment of project-based learning in a flexible environment. *International Journal of Technology and Design Education 13(3),* pp. 255-272. 2003.
[7] N. Hotaling, B. B. Fasse, L. F. Bost, C. D. Hermann and C. R. Forest. A quantitative analysis of the effects of a multidisciplinary engineering capstone design course. *J Eng Educ 101(4),* pp. 630-656. 2012.
[8] G. L. Brown. Dissolving the iron triangle: Increasing access and quality at reduced cost in public higher education. *Masters Thesis, George Mason University* 2012. Available: http://digilib.gmu.edu/dspace/bitstream/1920/7911/1/BrownG_thesis_2012.pdf. DOI: http://www.refworks.com/refworks2/?r=references|MainLayout::init#.
[9] S. Baum, C. Kurose and M. McPherson. An overview of american higher education. *The Future of Children 23(1),* pp. 17-39. 2013.
[10] P. Oreopoulos and U. Petronijevic. Making college worth it: A review of the returns to higher education. *The Future of Children 23(1),* pp. 41-65. 2013.
[11] J. Vigdor. Solving america's math problem. *Education Next 13(1),* pp. n1. 2013.
[12] E. A. Hanushek, L. Woessmann and P. E. Peterson. Is the US catching up?. *Education Next 12(4),* pp. n4. 2012.
[13] L. Young and T. Papinczak. Strategies for sustaining quality in PBL facilitation for large student cohorts. *Advances in Health Sciences Education* pp. 1-9. 2012.
[14] M. Horn, "Game Changer," *Education Next,* vol. 12, pp. 93-94, Fall 2012, 2012.
[15] M. Horn, "The Transformational Potential of Flipped Classrooms," *Education Next,* vol. 13, pp. 78-79, Summer 2013, 2013.
[16] A. J. Alston, D. Cromartie, D. Wakefield and C. W. English. The importance of employability skills as perceived by the employers of united states' land-grant college and university graduates. *Journal of Southern Agricultural Education Research 59(1),* pp. 59-72. 2009.
[17] B. Fuller, J. Wright, K. Gesicki and E. Kang. Gauging growth: How to judge no child left behind? *Educational Researcher 36(5),* pp. 268-278. 2007.
[18] J. Straub, S. Kerlin and E. Kim, "Analysis of Student Characteristics and Feeling of Efficacy in a First Undergraduate Artificial Intelligence Course," *Submitted to the Journal of Educational Computing Research,* .