Computerized Analysis Can Improve Education While Taking Advantage of Economies of Scale

Henry E. Schaffer, North Carolina State University

Fascinating areas of technology and economics can cause us to forget the fundamental concern of education. Computers can help learning outcomes, especially in large science, technology, and engineering courses.

Digital Object Identifier 10.1109/MC.2022.3143718
Date of current version: 8 April 2022

The interleaved questions "What is the purpose of education?" and "What are the goals?" approach the true bottom line of education: student learning outcomes. The treatment in the following describes computer-based analysis to improve student learning outcomes in a cost-effective manner. Science, technology, engineering, and mathematics (STEM) is the scope. An important reason is that in those areas, much of the course material can be dissected into ideas that form a concept tree to guide learning. A second reason is that much of the course material can be graded using multiple-choice questions (MCQs) for which computerized scoring and analysis are easy and inexpensive. Large enrollments are found in higher education STEM courses, and that is where improvements are badly needed.

0018-9162/22©2022IEEE

EDITOR **IRENA BOJANOVA**
NIST; irena.bojanova@computer.org

CONCEPTS

A physics example is that the concept of velocity/speed requires an understanding of coordinate systems. And then the concept of acceleration requires an understanding of both the velocity/speed concept and the displacement and distance concept. This building on prior concepts continues through the course. The relationship among concepts can be described as a tree graph, which can be shown in a spreadsheet table, where each notion is a row with the concept name and the following columns give the row number(s) of the immediate prior concept(s) (IPCs) on which this one is built. Then, there is a column with the resources for learning that concept, such as textbook sections, and other references. Figure 1 presents a spreadsheet concept table for a fraction of an introductory course. Figure 2 displays a tree version of the same relationships. When a student still has trouble dealing with a concept after going through the suggested resources, attention should be shifted to the IPCs that will provide further background. To assist this, an alphabetized concept table, including the resources for studying each idea, is provided along with the graphical presentation.

From an IT perspective, the concept tree is a directed graph. For educational use, it must also be acyclic.[1] A cyclic example is that to understand concept A, one must understand concept B; to understand B, one must understand C, and to understand C, one must understand A. That is a closed loop with no way to get started. The most common way this problem arises is that concept C really includes two concepts, C′ and C″. C′ does not require an understanding of A, and it is what is needed to understand B. C″ is not required to understand concept B and does demand an understanding of A. With this modification, we have the acyclic: C′ → B → A → C″. Alternatively, A, B, and C are so closely related that they need to be studied together as though they were one (complex) concept and labeled as a single notion. They need to be studied and understood together, and then the graph is acyclic. Why this emphasis on acyclic and directed? It is to facilitate the analysis described in the following to give clear, useful guidance to students who are having trouble understanding a concept, hindering progress in a course.

ECONOMICS

When teaching a small class (up to 15 or so pupils,) it is easy to get to know each student and track individual problems via questions, body language, and facial expressions. That is not possible when a class has 250+ students. The students lack individualized guidance, and the instructor cannot detect individual problems in learning. Why are such large-enrollment courses set up? Cost. In many sectors of our economy, costs have been lowered by using computers to automate processes without lowering quality. This can sometimes be done effectively in education, as described in the following for high-enrollment STEM courses, but instructor involvement is crucial. The instructor is the subject matter expert, and without that involvement, learning may be seriously misguided and not lead to the desired outcomes.

> The interleaved questions "What is the purpose of education?" and "What are the goals?" approach the true bottom line of education: student learning outcomes.

MCQs, ESSAYS, ANALYSES, AND ARTIFICIAL INTELLIGENCE

The computerized grading of MCQ tests is easy and routine in most

	A	B	C	D	E	F
1	Concept	IPC	IPC	IPC	IPC	Resources—Text Pages and More
6	Trig					10–14
7	Coordinate Systems					26–31
8	Velocity and Speed	7				35–42
9	Acceleration	8				42–45
10	Displacement and Distance					18–21
11	Motion Diagrams	6	8	9	10	52–57

FIGURE 1. A concept tree fragment in the form of a spreadsheet table.

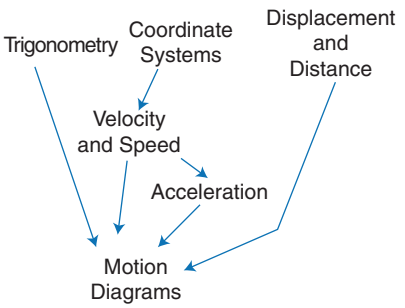

FIGURE 2. A graphical representation of a concept tree fragment.

environments that use online click-the-selected-answer or the ancient bubble sheets. Importantly, while reducing the grading workload is good, the results are also available in digital form and so can be used as computer input with essentially zero added cost. Are MCQs a good choice for assessing student progress in learning? They are often done poorly but can be done well if effort is put into constructing them. Bloom's Taxonomy of Cognitive Objectives needs to be discussed.[2] It describes remembering and rote recall as the necessary basic step in learning a subject and then, importantly, are the higher level aspects.

The taxonomy is often portrayed in the form of a pyramid with the levels remember, understand, apply, analyze, evaluate/synthesize, and create, which apply to every course, with emphasis shifted upward in advanced classes. When MCQs are oriented to the bottom steps, passing a course will not be good preparation for higher level material and a career. It takes more effort to write higher level MCQs, and that increases the cost. That appears to be the reason for the emphasis on lower level MCQs in vendors' question banks, including those that come with textbooks. The many Bloom's levels are a bit confusing and so, for FormAssess, condensing them into three levels (L, M, and H) works well (see Figure 3 in Young et al.[10]). While many courses also involve projects, presentations, and reports, they are outside this methodology's scope.

In computer science, it is easy to test whether a student can write a short program to print out the numbers from one to 10. This is essentially recall and so is not a good measure to evaluate the student's understanding and ability with respect to real programming challenges. As a result, industry interviewers have developed the Fizz Buzz Coding Challenge "to help filter out the 99.5% of programming job candidates who can't seem to program their way out of a wet paper bag."[5] We need to include the equivalent-level MCQs, and there are many examples of how to do this.[8] Why not use artificial intelligence to grade essay questions, complicated analyses, and more, which certainly would be good but which takes much too much instructor effort to grade. It might be a workable option someday, but that is not the case today and so will not be discussed further.

It takes an instructor about two half days working with the class notes and detailed table of contents of the textbook to write a concept tree for the usual three- or four-credit course.

FORMATIVE ASSESSMENT

Assessment means *testing* in educational jargon. *Summative assessment* means the final test for a course or an area that is used to judge if a class has been satisfactorily completed. This is extremely common, but while it checks whether a learning outcome has been achieved, it is of no help to a student in reaching that goal. Enter "formative assessment," which determines where the student is having difficulties and helps him or her over the encountered hurdles. That supports reaching the true course goal.

Not all students have the ability to achieve the required learning outcomes or perhaps are not willing to expend sufficient effort. That is unfortunate and cannot be overcome, but it should not divert us from assisting the students who could succeed if given help. The problem, then, is determining what assistance is needed and being able to afford to provide it. (There are many possible reasons why a student may not be willing to work hard in a course, and sometimes it may require more effort from the instructor to make the material more relevant. The methodology discussed in the following can free up instructor time, which can be devoted to this.)

The method of using an MCQ test for formative assessment harnesses the concept tree discussed previously. Our experience is that it takes an instructor about two half days working with the class notes and detailed table of contents of the textbook to write a concept tree for the usual three- or four-credit course. In subsequent semesters, that is typically reduced to 2–3 h to incorporate any changes in content and any experiences that might help student learning. Instructors are encouraged to provide their concept trees as open source, which will greatly reduce the initial upfront effort. This is the first example of "economies of scale," an important concept in the computer field. Next, each question on a test needs to be "tagged" with the concept(s) that it evaluates. We also suggest tagging each question with the Bloom's Taxonomy level. Our experience is that most questions test one or two concepts, more rarely three or four.

FORMASSESS

The analysis, an example of what can be done, is called FormAssess. It is straightforward, deals only with concepts, and does not deal with individual question quality.[6,7,10] (Item analysis is extremely useful regarding the quality of questions.[4]) FormAssess looks at each student's performance individually. It checks each student's missed questions for which concepts are tested and then sums up the whole test for each student, including which concepts were missed and how often. So, unlike the test grade, which just rates overall performance, FormAssess gives a student guidance on what to study and where to start. Even if the

student is given back the test, it is not easy to review missed questions and figure out what the basis of the lack of understanding is. This is especially true for a student who is having trouble with the course.

The student could, instead, go to office hours and review the test with the instructor. We are back to economics and students' emotional reactions. An instructor one-on-one with each student is not feasible with large courses, and experience has shown that students having trouble in a course tend to avoid discussion with the teacher. Sending the FormAssess report to each student, via automated means, gives students a private report. Our experience is that then students usually go to office hours to ask for additional help on concepts, rather than argue about individual questions.

The instructor receives all the individual reports and an overall class report showing the percentage of times each concept was correctly answered as well as the Bloom's levels. This gives feedback to the instructor to see where class presentations and assignments may need to be improved. As for the computer software and the cost of running it, once again there are economies of scale. The software[3] required a moderate development effort, but running it on a normal desktop computer takes about 1 s for a 400-student, 40-question test. The cost of this computer time is US$0 when rounded to the nearest cent.

ALTERNATIVES

Are there commercial services that will provide this with less effort? Looking at what is available shows very little that fully provides this. Most textbook publishers provide MCQ question banks. The quality of the questions is uncertain, particularly with respect to Bloom's levels, and the grading is simply correct/incorrect for each question. Sometimes there are hints to help with incorrectly answered online questions, but that focuses on questions, not understanding concepts. One other factor is that a student is required to purchase the textbook to get access to the testing service. There are also services which, in addition to test banks, enable the instructor to provide questions.[9] Once again, there is an associated cost, and the reporting remains just correct/incorrect. So, at this time, there is no service that is fully comparable with the FormAssess methodology.

Every student who fails to pass a course is a failed investment for the university, the public, and the learner. Future employers also lose. What is needed for a win–win–win is to assist all able and willing students over the learning hurdles they encounter so that they do not fail courses they could pass. And this needs to be done in an affordable manner to assist students from all backgrounds and economic circumstances. The FormAssess computational methodology takes a significant and affordable step in the desired direction.

REFERENCES

1. A. V. Aho, J. E. Hopcroft, and J. D. Ullman, *Data Structures and Algorithms*. Reading, MA, USA: Addison-Wesley, 1983.
2. B. S. Bloom, Ed., *Taxonomy of Educational Objectives: The Classification of Educational Goals: Handbook I: Cognitive Domain*. New York, NY, USA: David McKay Company, 1956.
3. "Formassess," GitHub. https://github.com/hes8/formassess (Accessed: Mar. 1, 2022).
4. G. F. Kuder and M. W. Richardson, "The theory of the estimation of test reliability," *Psychometrika*, vol. 2, no. 3, pp. 151–160, 1937, doi: 10.1007/BF02288391.
5. A. North, "Fizz buzz test," Wiki. https://wiki.c2.com/?FizzBuzzTest (Accesses: Mar. 1, 2022).
6. L. Petrovich, H. E. Schaffer, and K. R. Young, "Shrinking a large STEM class: Providing small class benefits in a large introductory chemistry class via automated formative assessment," unpublished.
7. H. E. Schaffer, K. R. Young, E. W. Ligon, and D. D. Chapman, "Automating individualized formative feedback in large classes based on a directed concept graph," *Frontiers Psychol.*, vol. 28, p. 260, Feb. 2017, doi: 10.3389/fpsyg.2017.00260.
8. "Testing and evaluation in the biological sciences: Commission on undergraduate education in the biological sciences: Report of the panel on evaluation and testing," CUEBS, Washington, DC, USA, CUEBS Publication 20, Nov. 1987. http://ofd.ncsu.edu/wordpress/wp-content/uploads/2013/09/testing-and-evaluation-in-the-biological-sciences.pdf
9. WebAssign. https://webassign.com (Accessed: Mar. 1, 2022).
10. K. R. Young, H. E. Schaffer, J. B. James, and M. T. Gallardo-Williams, "Tired of failing students? Improving student learning using detailed and automated individualized feedback in a large introductory science course," *Innov. High Educ.*, vol. 46, no. 2, pp. 133–151, 2021, doi: 10.1007/s10755-020-09527-5.

HENRY E. SCHAFFER is a professor emeritus of genetics and biomathematics in the Department of Biological Sciences, North Carolina State University, Raleigh, North Carolina, 27695, USA, where he also coordinates special IT projects and faculty collaboration in the Office of Information Technology. Contact him at hes@ncsu.edu.