Case-Based Grading: A Conceptual Introduction

Allen Briscoe-Smith[1]
Dept. of Management Information Science, California State University, Sacramento
Sacramento, CA 95819-6088, USA

and

Nicholas Evangelopoulos[2]
Dept. of Business Computer Information Systems, University of North Texas
Denton, TX 76203, USA

Abstract

This paper examines how intelligent grading of Web-based examinations can be achieved using a Case-Based framework. Multiple-choice and essay questions are common examination formats widely used by the academic community. A survey of about 2,000 University faculty, conducted in the fall of 2001, revealed that preferences are evenly split between the two formats. In this paper we propose the Multiple-Choice with Free Text justifications (MCFT) format. The goal is, on one hand, to combine the benefits of multiple-choice and essay questions and, on the other hand, to allow for computer-aided grading based on a mechanism similar to Case-Based Reasoning. The specific properties of Case-Based Grading (CBG), namely case representation, retrieval, reuse, revision, and learning, are examined. Finally, a two-phase CBG algorithm is proposed.

Keywords: Case-based reasoning, MCFT format, online examinations, partial credit, intelligent exam grading

1. INTRODUCTION

Into the new millennium, higher education is forced to operate amidst a remarkable new environment. As the 1,200-year-old paradigm of universities being a self-contained village (the college campus) becomes shattered by new technologies and new societal values (Jennings 2001), Web-based education and on-line examinations enjoy an increasing popularity. While on-campus college enrollment numbers continue to increase, advances in Information Technology and globalization trends create new dynamics and redefine the concept and the scope of the academic classroom. Instructors are asked to serve audiences that are geographically dispersed, interact with the class in an asynchronous mode, and have significantly larger sizes. To address these issues, implementations of on-line education tools are being examined by several universities, such as Western Governors University and Southern Regional Education Board's Electronic Campus (Carnevale 2000). On-line examination administration entails several advantages: electronic storage of student submissions, shorter examination administration cycle, instantaneous scoring and reporting, and greater test reliability (Bennett et al. 1997). It seems certain that while computers have had a profound impact on the development of testing, they can be expected to have an even greater influence on the future of automated testing (Lippey 1972; Stewart 1990). However, within this environment of the Global Virtual Classroom, traditional grading practices involving human graders are fast approaching their limits. While the Internet provides the means by which examinations can easily be administered to an almost unlimited number of students, the problem of how to grade the examinations quickly, accurately, and intelligently, persists.

Many traditional test question formats have been adopted by test administrators, including multiple-choice (MC), true-false, fill-in-the-blank, essay/free text (FT), short-answer, matching, and so on. Each question format exhibits advantages and disadvantages unique to each format. Multiple-choice and essay type questions today remain among the most popular testing formats (Geiger 1996; Martinez 1999; Steele 1997). While MC

[1] icede@cs.ucr.edu
[2] evangeln@unt.edu

examinations are easy to grade by a computer, this question format is commonly criticized because it allows students to blindly guess at the correct answer. Essay questions, on the other hand, can reveal the depth and breadth of students' knowledge, as well as erroneous conclusions that are drawn (Steele 1997), but are much more difficult to grade via computer.

Fitzpatrick et al. (1996) discusses the grading practices of multiple-choice examinations where a constructed response (CR) item allows the examinee to construct a written response that is scored by raters using a rating scale having two or more score points. In some testing programs, such as Alabama’s end-of-course tests in algebra and geometry, California’s Learning Assessment System, Indiana’s Performance Assessment for School Success, and Michigan’s High School Proficiency Test, multiple-choice and CR format items are used in parallel to assess the skills of interest (Fitzpatrick et al. 1996).

The grading of examinations comprised of MC questions via a computerized system is a relatively trivial task. A computer simply checks to see if the student’s recorded answers match the answers of the key, and marks the examination accordingly. From the cognitive point-of-view, multiple-choice and essay type questions measure significantly different aspects of knowledge. MC questions ask from a student to recognize the correct answer from a finite pool of choices, whereas essay type questions require the student to recall the knowledge and express it in writing. The hybrid question format we introduce in this study brings multiple-choice closer to essay, thus better addressing the issue of knowledge recall.

Despite the inhibitions towards MC questions, many professors today use this question format due to its ability to make grading efficient and practical, especially for large sized classes. In order to reassess the situation, in the fall of 2001 we conducted a survey across the Unites States. A random sample of 10,000 University faculty members from all academic disciplines was invited to participate in a brief on-line survey. The goal of the survey was to examine how ready and how open would university faculty be in the proposed hybrid question format. Responses from 2,063 professors were received in a five-week period. One of the questions was: “Refer to a course you taught recently. Have you used the multiple-choice testing format on exams, tests, quizzes, or assignments?” The results reveal that the professors’ opinions on the use of multiple-choice questions are split almost evenly. The percentage of professors using the MC question format was 53%, while a 47% did not use them. Most of the participants (1,134 out of 2,063) also submitted optional comments, where the opponents of the MC format expressed their opposition quite passionately. Another question addressed the issue of partial credit. Hilton (1993) explains that some instructors are against the awarding of partial credit because partial credit rewards a student that has not displayed a complete understanding to a specific problem. In our survey, 88% of the participants said that they have awarded partial credit for exam answers that were neither entirely correct nor entirely wrong. These findings set the stage for our proposed Multiple-Choice question with Free Text justifications (MCFT) format: faculty seem to appreciate the practicality of MC questions, recognize the pedagogical superiority of the FT questions, and embrace the practice of awarding partial credit. One would, therefore, expect them to receive positively a new question format which would be almost as fast to grade as MC, would offer almost the same pedagogical value as FT, and would still allow for partial credit. In fact, our fall 2001 survey also included the following question: “If (in the future) a well-established Web-based examination paradigm combines the advantages of multiple-choice and partial credit, would you be willing to use it for your course?” A percentage of 36% of the participants responded “probably yes”, “yes”, or “definitely yes”.

The next challenge to address is the actual grading of this new test format. Contrary to the trivial nature of grading multiple-choice questions, computer grading of essay questions can be a very difficult task. The rest of this paper will explore the problems and issues of creating a computer based examination system that will employ case-based reasoning to intelligently grade the multiple-choice/free text submissions from students. The intelligent grading of this new question format using a Case-Based Reasoning (CBR) framework will be called Case-Based Grading (CBG).

2. CASE-BASED REASONING

Case-based reasoning is a problem-solving paradigm that in many respects is fundamentally different from other major AI approaches. Instead of relying exclusively on general knowledge based upon a specific knowledge domain, CBR is capable of applying specific knowledge from previous, concrete problem experiences (Aamodt and Plaza 1994). The fundamental principles of case-based reasoning have been developed, and many applications have demonstrated CBR’s role as a useful technology (Leake 1996). To better understand CBR, it is important to understand what “reasoning” entails. Systems based on reasoning attempt to apply generalized rules onto new situations, starting from scratch with each case. In systems based on CBR, the primary focus of the knowledge base is centered not on generalized rules, but on a memory of stored cases recorded from prior experiences. In CBR, new solutions are generated through the retrieval of the most relevant cases from memory, and thus adapting them to the current case. CBR is based solely on remembering, and not through redundant application of previous cases.

Case-based reasoning was motivated by the function of remembering in human reasoning. Over the last few decades many studies of CBR have been employed, making CBR the mature field it is today. The study of CBR is driven by two primary motivations (Leake 1996). The first motivation is from cognitive science in the desire to effectively model human reasoning and learning. The second motivation for studying CBR is from artificial intelligence, in the desire to develop technologies making AI systems more effective.

The first motivation, cognitive sciences, is what this paper is primarily interested in, as cognitive reasoning is present in every test-taker's mind while taking an examination. A great deal of research has been conducted in the area of cognitive model constructions using CBR. For example, previous research has supported the importance of remembering prior examples in learning how to use a computer text editor (Ross 1984), learning computer programming (Pirolli and Anderson 1985), mathematical problem solving, diagnosis by automobile mechanics, diagnosis by physicians, explanation of anomalous events, and decision-making under time pressure. To fully understand these processes requires the development and testing of models that effectively mimic how humans store, retrieve, and apply prior cases (Leake 1996).

3. THE MCFT FORMAT

In order to intelligently grade on-line examinations using a case-based framework, we must first seek an appropriate examination question format. A hybrid question format that combines the advantages of multiple-choice and essay format while accommodating the specific functional needs of CBR, is introduced.

Throughout formal education, one of the most widely used devices to evaluate progress and performance has been the multiple-choice test, or some variation of this objective format (Geiger 1996). Most professors and students are quite familiar with the Multiple-Choice (MC) and essay (Full Text, FT) question format types. While MC questions are relatively easy to create and even easier to grade, this test question format makes it also easier for students to either cheat on exams, or to simply guess at the correct answer. In a typical MC examination with four choices per question, a student would have a 25 percent chance of randomly guessing the correct answer. Research has found that the greater the number of options per question, the less likely it is that a student will select the correct answer through guessing (Hopkins and Stanley 1981; Wang and Calhoun 1997).

The technique of granting a student the opportunity to justify their reasons for choosing an answer on a MC examination has been recommended periodically (Burrill 1976; Gorrell and Cramond 1988), but never put into commercial use. One of the drawbacks of MC questions is the fact that a student is not given the chance to explain their reasons for selecting the answer. A test question format is needed which will allow a student to not only select the answer they feel is correct from a set of possible answers, but also explain why they feel their choice is correct. Allowing a student to justify their answer selection would be highly desired, especially in classes that are covering topics beyond simple knowledge (Gorrell and Cramond 1988). It would also allow students to gain partial credit for responses that would otherwise be considered completely incorrect.

Figure 1 shows the hybrid result of combining the MC format with the FT question format. We call this the Multiple-Choice with Free Text Justification (MCFT) question format. When a student is to be presented with this testing format, it may look a little odd to them at first, however, as Gorrell and Cramond (1988) report, students responded positively to being presented with the ability to justify their answers on examinations.

Who discovered gravity?
a) Einstein
b) Newton
c) Beethoven
d) Wang

Please justify your answer:

Figure 1. Multiple-choice with Free Text Justification Test Format.

The ultimate goal of any test is to reliably measure a student's understanding of the topic being tested. If the testing medium interferes with a student's performance on the examination, then the examination is not reliably measuring the student's understanding. One advantage of the MCFT question format is the fact that students will potentially feel like they don't need to change answers, as they will be graded not only on the answer they select, but also the text justification they supply. Another potential benefit of MCFT is the fact that because students will feel that they have a better chance at receiving either full credit or partial credit, students will spend less time worrying if they chose the correct answer, and more time conveying their knowledge in the text justification boxes.

By combining the MC and FT question format types into one, we are also combining the advantages, and disadvantages, of both question types together. Table 1 outlines the advantages of these two question types, along with the advantages of the MCFT question format.

Advantages		
Advantages	**FT**	**MCFT**
Easy to create	Easy to create	Easy to create
Easy to grade (human or computer)	Assesses knowledge recollection	Assesses knowledge recollection
Easy to represent on a computer-based exam	Student cannot easily cheat or guess	Easy to represent on a computer-based exam
Widely accepted across academic industry and familiar format to professors	Widely accepted across academic industry and familiar format to professors	Student cannot easily cheat or guess
		Student able to justify answers

Table 1. Advantages of Multiple-Choice (MC), Essay (FT), and MCFT question types

Disadvantages		
MC	**FT**	**MCFT**
Open to guessing and cheating	Difficult to grade consistently	Time consuming to grade by human grader
Does not accurately measure student's cognitive understanding of topic	Can become cumbersome if too many students to grade	Unfamiliar testing format to most professors
Student not able to justify answers (all or nothing).	Time consuming to grade by human grader	

Table 2. Disadvantages of MC, Essay, and MCFT question types

Table 2 outlines the disadvantages of these two question types, along with the disadvantages of the MCFT question format. As can be seen by the two tables above, the advantages of the MCFT question format are significant. Not only is the test format as easy to create as a multiple-choice question, but also the free text component of the question will allow the capture of a student's cognitive thoughts on a per-question basis. However, having the FT answers accompanying each question submission, the difficulty of grading of these examinations will increase substantially, even more so if grading is performed by a computer. In the next section we will explore how this MCFT question format can be used to intelligently grade on-line examinations using a case-based framework for the awarding of Partial Credit.

4. CASE-BASED GRADING

We now look at how the world of Artificial Intelligence can assist with the intelligently grading of examination answers captured by a Web-based examination administration system. The solution to our problem lies in case-based reasoning. CBR will assist us in intelligently grading the free text submission segments of each question while awarding partial credit according to the perceived correctness of each question.

Case-based reasoning functions through the usage of previous cases. New cases are compared to a case-base and checked for similarity. As the CBR system is utilized, new cases will be added to the case-base, building upon the general knowledge component found in the center of the CBR cycle. A common human equivalent of case-based reasoning is that of an automotive mechanic. When automotive mechanics work on cars, they naturally remember how they go about fixing different problems. When a new problem arises, the mechanic subconsciously checks his memory of previous cases to find which case is most similar to the current case. In the event that a past case from the knowledge base is found in memory, the previous case is adapted to the current case, providing a more intelligent solution to the current problem. However, if a similar past case is not found in the general knowledge of the mechanic, this too can become valuable to the CBR system. This would be considered a brand new case for the general knowledge module. Which ever way the problem is fixed (either through memory of a previous case or not), the solution to the current case is stored into the general knowledge, and can be revisited later.

Where does all of this fit in with the intelligent grading of on-line examinations? Quite simply, this will fit in with the intelligent grading of the free text submissions of every student. Say for example, a class of 60 students is taking an examination comprised of 10 MCFT-format questions. After the examination is administered, there will be 60 text justifications for each question, resulting in at most 600 free text justifications for the entire examination. When grading comes around, case-based grading will work on a question-by-question basis, grading each free text justification individually, each time referring to previously graded questions to gain a better assessment as to what grade should be awarded to the current question. Automatic comparisons between documents can be accomplished by utilizing techniques of calculation of similarity between documents. Such comparisons are common in the field of Information Retrieval, which offers the basis for document comparisons performed by search engines.

The CBG Cycle

CBG exhibits a learning cycle similar to the one we would find in a generic CBR setup. To start the CBG cycle, the initial case-base must be created. The computer will present the human grader with a few exams to be graded by hand. The initial exams graded by the human will begin the creation of the case-base for the CBG system. As the rest of the examinations are graded by the computer, CBG cycles around the following steps (Aamodt and Plaza 1994):

1. Retrieve. In the first phase of the CBG cycle, the most similar free text submission or submissions are retrieved. To begin this phase, a new free text submission from the student is introduced. That is analyzed and compared with all the existing submissions stored in the CBG database. In the worst case, no similar match is found. In the best-case scenario, a similar match is found during the first comparison. In the average-case scenario, a similar match is found.

2. Reuse. In the second phase of the CBG cycle, the information and knowledge in the retrieved case is reused to grade the new problem. If a similar match is found in the CBG database, then the matched case will be selected to help accurately grade the current free text submission. The information retrieved from the CBG database will include detailed analysis of the free text submission, as well as the numerical grade assigned to the submission. The result of this phase will be the creation of a proposed grading solution.

3. Revise. In the third phase of the CBG cycle, the proposed solution is revised to better fit the current free text problem being graded. By testing the proposed solution for success, either by human or preferably by some computer algorithm, this will allow the individualized screening of cases (previously graded free text submissions) stored in the CBG database.

4. Retain. In the fourth phase of the CBG cycle, pertinent information on the newly graded free text submission is retained in the CBG database. Typical information stored would be the free text submission itself, as well as the grade assigned to the free text submission.

As indicated in Figure 2, the CBG database plays a key role in the entire CBG cycle by supporting the CBR-like processes. The database is the heart and brain of the CBG cycle, as it is contains all previously graded free text submissions. In the CBG cycle, a new student submission is graded by applying previously graded free text submissions to the current free text submission. Aamodt and Plaza (1994) broaden the CBR cycle from a general view of sequential steps to a decomposition and description of each of the four general levels of CBR. CBG will follow the same task-oriented view where each step is represented as a task that the CBG system must achieve. Figure 2 displays the task-oriented view of CBR.

As we can see from Figure 2, the four general steps previously discussed in CBG reside at the top of the task tree. Retrieving a previous free text submission requires the steps of identifying the features or characteristics by which recognition will take place, searching for these features/characteristics, the initial match of cases in the CBG database to the current free text submission, and selection of the most appropriate case. The next step, reuse, involves the copying and adaptation of free text submissions from the database to current application. The revision phase involves evaluating the proposed solution and repairing faults as needed. The last phase, retain, requires the integration, indexing, and extraction of key characteristics in efforts to store the current free text submission into the CBG database. This addition to the CBG database will be used to assist in grading future free text submissions. We will now discuss each subtask of the CBG cycle in detail.

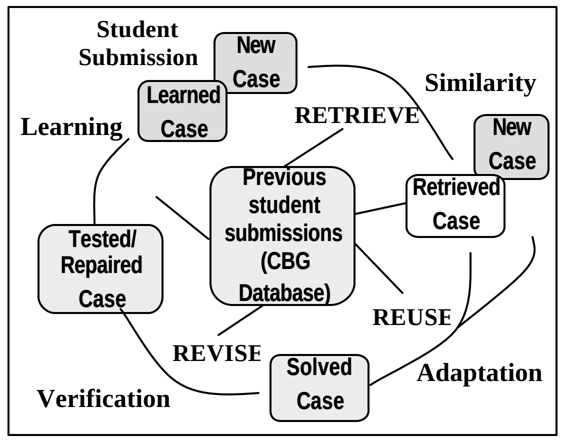

Figure 2. The Case-Based Grading Cycle (Adapted from Aamodt and Plaza, 1994).

Representation of Graded Cases

In case-based grading, the structure and content of the collection of cases is very important. The accurate and efficient storage of previously graded free-text submissions is paramount to the functionality of the CBG system. The case representation problem in CBG is primarily the problem of deciding what to store in a case, finding an appropriate structure for describing case contents, and deciding how the case memory should be organized (Aamodt and Plaza 1994). The Vector Space Model, widely used by Information Retrieval theories to support search engines, can provide a framework for free text representation that facilitates calculations of similarities between answers.

Graded Case Retrieval

The retrieval task starts with a non-graded submission, and ends with a best matching previously graded submission. As outlined in Figure 2, the subtasks for this task are executed in the following order: Identify Features, Initially Match, Search, and Select. In order for previous cases to be retrieved, certain criteria must be outlined as a means to accurately match a current free-text submission, with those existing in the CBG database. The heart of the CBG cycle revolves around the identification of the characteristics of free text submissions, and how to formally compare previous submissions with that of current submissions. Once this paradigm is decided upon, the actual matching algorithm to match previous submissions with a current submission can be set. After an initial match between the current free text submission and a stored case in the CBG database are identified, the matching submission case in the CBG database is essentially reused to effectively grade the current free text submission.

Graded Case Reuse

In the case reuse phase, the graded submission contained in the CBG database is reused to provide intelligent information to accurately grade the current submission. There are two aspects to reusing a previous case upon a new case. First, the differences between the two cases must be recognized, and second, the identification of the reusable segments of the previous case as applicable to the new case must take place. The two possible subtasks for this task are Copy and Adapt (Aamodt and Plaza 1994).

In the case of copying, a trivial copy of the graded outcome of the previous case is applied to the current case. The second choice for case reuse is Adaptation, by which current free text submissions are graded by a transformational method. In this method, knowledge of the previous case in the form of transformational operators exists, by which this knowledge can be used to accurately grade the current free text submission. In a CBG system, the free text snippet as well as the final grade for the question will be stored in the CBG database. Also stored along with the graded submission can be the actual mathematical intelligence used to generate the grade. All this knowledge can be used by the CBG system as a means to derive previous knowledge of how a gr ade was calculated, and apply it to a current free text submission. Aamodt and Plaza (1994) describe this as derivational method of case adaptation. Now that a case has been identified in the CBG database as a useable match to the current submission, the case must be revised to meet the needs of grading the current free text submission.

Graded Case Revision

In some cases, the grade obtained from the reuse phase may not be correct, as determined by a human or a computer algorithm. In this case, graded case revision must take place as a means to either learn from the failure, or revise the case to become accurate. In revision, the graded free text submission from the previous phase must be evaluated for correctness by either a human grader, or a computer algorithm. A computer algorithm or some other intelligent verification system may flag down submissions that appear to be marked incorrectly. Human graders can intervene with the CBG system as a means for quality control and to make sure the CBG system is grading problems properly. The evaluation subtask is generally outside of the CBG system if it is handled by a human, and not a computer algorithm.

If a graded submission is deemed correct, it will flow through to the last phase, graded case retainment. Otherwise, the case must be repaired appropriately either by human or computer algorithm. This involves the detection of the errors in the proposed solution, and repairing the errors. In the case of CBG, the only error that could occur is a drastically incorrect score to a problem. For example, if the average score on a problem was 90, and the CBG system was awarding 10 points to a student, either the graded submission will be flagged for a human to review the graded submission, or a computer algorithm may intervene to find out what the issue was. Every successfully graded submission can be inserted into the CBG database as a new case to be learned by.

Graded Case Retainment (Learning)

The input to this last phase is a successfully graded free text submission. Because the submission (case) has successfully passed through almost all the phases of CBG, it will be added to the CBG database. Once added to the CBG database, the case will become an example for future free text submissions to be graded. As described in the first phase of CBG, the extraction of pertinent information to be stored in the CBG database is paramount. The information stored in the CBG database must be accurate and plentiful for the CBG system to correctly match cases in the CBG database to that of future free text submissions. The subtasks in this phase are: Extract, Index, and Integrate.

In these three subtasks, information about the free text submission is extracted. For example, key words in a submission can be stored, along with the final score for the problem. Indexing involves the choice of indexing methodology to store the information in the CBG database. The last subtask, Integrate, involves integrating the newly graded free text submission into the CBG system. Successful storage of the graded submission into the CBG database is the result of the integration.

5. IMPLEMENTATION OF CASE-BASED GRADING

To ensure increased accuracy, we propose that the CBG cycle be a two-phase process when grading examinations:

Phase 1: In the first phase of grading the examinations, the computer will make a pass through every student's examination and create a unique case for each free text submission. During this phase, the computer can analyze all the examinations and present a small number of examination submissions (in the neighborhood of three to twenty submissions) that are deemed different by computer algorithm to a human grader, asking them to accurately grade each examination. The graded examinations will be stored in the CBG database along with the scores associated with each problem. This is the initial creation of the CBG database (a.k.a. case-base).

Phase 2: In the second phase of grading the examinations, the computer will iterate through each examination submission, implementing the CBG cycle on each submission. The human grader is out of the picture now, and the computer takes over grading every question on every examination.

Compared to the fast and accurate turnaround of grading examinations via a computer, a human grader cannot parallel the timeliness and accuracy in grading by hand. In the example above, the professor would have to look through 600 text justifications in order to award the correct number of points per question, all while awarding partial credit consistently. If it took the professor approximately one minute per question, it would take 10 hours to completely grade all examinations. This assumes the professor will grade straight through with no break, and will not experience fatigue. In practice, of course, fatigue does occur, and the grader's memory does fade, resulting in unreliability and inconsistency, hence the tendency of different instructors to assign different grades to the same paper (Williams et al. 1991). Unlike human graders, computers do not experience fatigue, can instantly check back corrected papers to ensure consistency, and do not experience temptations of bias.

When grading essay papers, it would be ideal not to have any sources of error, so that differences in students' grades would reflect actual differences in their performances. Some authors suggest that the reliability of grading essays can be improved when the human graders are supplied with explicit scoring criteria, detailed instructions on grading rules and paradigms, and adequate training and monitoring of grading activities (Bell 1980). While these suggestions were made over twenty years ago, today's technologies can supply the answer for these problems. The creation of a grading standard by a human grader is, very often, an iterative process: as the grader discovers new variations of correct or wrong answers, the grading standard is continuously updated. It is very rare for a human grader to be able to anticipate the entire range of submitted answers. It, therefore, seems only natural for the human grader to seek the assistance of a computer system. This CBG system will discover those submissions that offer a good coverage of the spectrum of the variability in submissions, based on an analysis of the submitted answer. Grading of such representative student submissions by a human grader would create an optimum grading standard, not in the form of a set of grading rules, but in the form of a graded set of submissions, which is what we call here the *case base*.

When using CBG, a human would only need to grade three to twenty examinations, and would then let the computer grade the rest of the examinations automatically. This paradigm is highly opportunistic, as the human grader only needs to grade a hand full of examinations. The energy of the human grader can be spent on grading these few examinations, while letting the computer accurately grade the other examinations. In the future of academia, when on-line universities could potentially have thousands of students per class, CBG of on-line examinations is expected to save copious amounts of time and money for academic institutions.

In order to implement this system, any programming language that can read data from a database, analyze strings, and output back to the database, may be employed. Examples of languages to be used are Visual Basic (stand-alone executable), and Cold Fusion (web-based grading module).

During the 2002-2003 academic year, experimental implementation of a CBG system is underway by the authors and their associates. The CBG system is being implemented in a multi-module system approach, where the CBG grading component is a stand-alone program separate from any software used to administer the examination and capture student submissions. The student submissions are stored in a database, where by the CBG program will read in the student submissions, pre-process them, compile an optimal set that maximizes the variability in answer similarity, and present it to the human grader. While the human grader is still working on the initial training set, the computer keeps calculating measures of possible inconsistency in the awarded grades, by comparing them to internal measures of exam submission similarity, such as multiple-choice answers, occurrence of course-specific keywords in the free text answers, location of question-specific keywords, and overall document similarity. The human grader may address the inconsistencies by adding new terms to the dictionary of the course-related keywords, adding new terms to the set of the answer-specific keywords, or revising the grades. At the end of this interactive work, the initial case base will be built. The computer will then start following the steps of the CBG cycle, as they were described in section 4. The authors expect to present some preliminary results from this implementation around the end of the current academic year.

6. CONCLUSIONS

The proposed CBG methodology introduces new exam grading technology that utilizes human-computer interaction for the generation of an initial set of graded cases, and subsequent automatic grading that follows the case-based reasoning paradigm. The new MCFT question format is not only a practical solution to the implementation of the proposed CBG system, it is also a hybrid format that combines essay type and multiple choice questions. IT education, as a field that combines the requirements of testing students on problem-solving capabilities, as well as conceptual understanding, appears to be well-suited for such a hybrid format. We expect the proposed methodology to help IS/IT instructional delivery when it takes the specific form of online student evaluation of distance education courses, an avenue that seems to be an inevitable mainstream in the near future.

7. REFERENCES

Aamodt, Agnar; Enric Plaza, 1994, “Case-based reasoning; Foundational issues, methodological variations, and system approaches,” AI Communications, Vol. 7, No.1, pp. 39-59

Bell RC, 1980, “Problems in improving the reliability of essay marks,” Assessment in Higher Education, Vol. 5, pp. 254-263

Bennett, R.E.; M. Steffen; M. Singley; M. Morley , 1997, “Evaluating an automatically scorable, open-ended response type for measuring mathematical reasoning in computer-adaptive tests,” J. Educ. Measurement, Vol. 34, No. 2, pp. 162-179

Burrill, R. “Modified objective questions for tests in introductory geography,” The Journal of Geography, Vol. 75, pp. 338-342

Carnevale, Dan, 2000, “Two models for collaboration in distance education,” The Chronicle of Higher Education, May 19, 2000, Vol. 46, No. 37, p. A53

Fitzpatrick, Anne, 1996, “Scaling performance assessments: A comparison of one-parameter and two-parameter Partial Credit” Journal of Educational Measurement, Vol. 33, Issue 3

Geiger, Marshall A., 1996, “On the benefit of changing multiple-choice answers: student perception and performance,” Education, Vol. 117, No. 1, p. 108

Gorrell, Jeffrey; Bonnie Cramond, 1988, “Students' attitudes toward use of written justifications for multiple-choice answers,” Educational and Psychological Measurement, Vol. 48, No. 4, p. 935

Hilton, Peter, 1993, "The tyranny of tests," Chemtech, Vol. 23, Issue 8

Hopkins, K.D; J.C. Stanley, 1981, Educational and Psychological Measurement and Evaluation (6th edition). Englewood Cliffs, N.J., Prentice Hall

Jennings, Lane, 2001, “Alternative Visions for the Future University,” The Futurist, Vol. 35 (1), p. 58

Leake, David B., 1996, Case-based reasoning: experiences, lessons & future directions, Menlo Park, Calif.: AAAI Press; Cambridge, Mass.

Lippey, Gerald, 1972, "A Comprehensive Question Retrieval Application to Serve Classroom Teachers". In Proceedings of the 1972 Fall Joint Computer Conference.

Martinez, Michael E., 1999, “Cognition and the question of test item format,” Educational Psychologist, Vol. 34, Issue 4, p. 207

Pirolli, P.; J. Anderson, 1985, “The Role of Learning from Examples in the Acquisition of Recursive Programming Skills,” Canadian Journal of Psychology, Vol. 39, pp. 240-272

Ross, B., 1984, “Remindings and Their Effects in Learning and Cognitive Skill,” Cognitive Psychology, Vol. 16, pp. 371-416

Steele, Craig W., 1997, “Essays - well worth the effort,” College Teaching, Vol. 45, No. 4, p. 150

Stewart, Deby, 1990, “Testing & the computer,” International Journal of Instructional Media, Vol. 17, Issue 4, p. 309

Wang, Jianjuni; George Calhoun, 1997, “A useful function for assessing the effect of guessing on true-false and multiple-choice tests,” Educational and Psychological Measurment, Vol. 57(1), p. 179

Williams, Julie, P. Sanford, W. Stratford; A. Newman, 1991, “Grading written essays: a reliability study,” Physical Therapy, Vol. 71(9), p. 679