Implementation of an Automated Grading System with an Adaptive Learning Component to Affect Student Feedback and Response Time

Kevin Matthews
Thomas Janicki
Information Systems Operations Management Department
University of North Carolina Wilmington
601 South College Road
Wilmington, NC 28403

Ling He
Department of Accounting
Saginaw Valley State University
7400 Bay Road University Center
Saginaw, MI 48710

Laurie Patterson
Computer Science Department
University of North Carolina Wilmington
601 South College Road
Wilmington, NC 28403

ABSTRACT

This research focuses on the development and implementation of an adaptive learning and grading system with a goal to increase the effectiveness and quality of feedback to students. By utilizing various concepts from established learning theories, the goal of this research is to improve the quantity, quality, and speed of feedback as it pertains specifically to the grading of computer skills with a focus on personal productivity software. Feedback has been identified as a key component of successful learning among students. This research builds upon the previous knowledge from the cognitive, behavioral, and resource-based views of learning as well as upon the establishment of grading rubrics. An automated grading system was developed that allows instructors to quickly grade multiple complex computer literacy assignments. Key to the success of the system is the ability of the system to "learn" the correct and incorrect responses and store them for future use. To understand the impact of the system on feedback, three hypotheses were created and experiments were developed to test them. The system was shown to positively affect the quantity of feedback and reduce the time required for grading assignments. No effect on the quality of the feedback comments was shown and may be a subject of further study.

Keywords: Rubrics, Pedagogy

1. INTRODUCTION

The National Research Council (NRC) and National Science Foundation (NSF) have defined basic requirements that today's students need to "Be Fluent in Information Technology" (BeFIT) (National Research Council, 1999). These concepts revolve around increased IT skills, concepts, and capabilities of all citizens. Many universities, colleges and two-year institutions require computer literacy for STEM (Science, Technology, Engineering and Mathematics) majors as well as for business majors. Computer literacy centers primarily on the use of personal productivity software applications, such as word processors, spreadsheets, databases and presentation applications.

Many educational institutions offer computer literacy courses to students and assist the learning process by assigning a certain number of computer projects. According to various learning theories, providing meaningful and timely feedback on assignments has been identified as a key component of successful learning among students. However, it is very time consuming and sometimes impractical to provide extensive and qualified feedback on numerous computer projects. This research reports the development and implementation of an adaptive learning and grading system with the goal to expedite and improve the feedback provided to students for their personal productivity software (i.e. spreadsheet and database) assignments. This research builds upon previous knowledge from the cognitive, behavioral, and resource-based views of learning as well as the establishment of the appropriate grading rubrics.

Computer-assisted assessments (Conole and Warburton, 2005) or automated grading systems are becoming more popular in higher education institutions because they can significantly enhance the learning process. In our study, an automated grading system, also known as the Adaptive Grading/Learning System (AGLS), was developed to allow instructors to quickly grade multiple and complex computer literacy assignments while providing meaningful feedback to students in order to stimulate an efficient learning process. The system provides for a consistent grading rubric for each assignment. A unique feature of the system is the ability of the system to "learn" the correct and incorrect responses and add them to the rubric. It is unique and different from what is currently provided by book publishers as it enables instructors to build more complex assignments and also share this enhanced grading rubric with other instructors.

This research investigated how 'auto grading' with an adaptive learning component might be used to affect the quality, quantity and the speed of feedback. Hypotheses were developed and evaluated using data collected by the existing gradebook reporting systems.

2. LITERATURE REVIEW

A student's overall success is largely influenced by the ability of the educator to present new information in creative and meaningful ways while at the same time evaluating a student's understanding of this information. This process requires students to learn the material covered by the educator. A brief overview of three learning theories is discussed in this section with particular attention to feedback theories and concepts.

2.1 Cognitive Learning Theory

Robert Gagne (1965, 1985, 1988; Gagne, Briggs and Wager, 1992) proposed a list of nine elements that should be present in any lesson in order for learning to occur. These nine elements form the framework for cognitive learning theory, where each element leads to the next, higher level element. They are: Gaining attention ("reception"), Informing learners of the objective ("expectancy"), Stimulating recall of prior learning ("retrieval"), Presenting the stimulus ("selective perception"), Providing learning guidance ("semantic encoding"), Eliciting performance ("responding"), Providing feedback ("reinforcement"), Assessing performance ("retrieval"), and Enhancing retention and transfer (generalization").

Of the nine "conditions for learning" that Gagne et al. (1992) provide, other research shows that eliciting performance and practice from the student ("responding") and providing adequate feedback ("reinforcement") are the events most directly associated with student success (Martin, Klein and Sullivan, 2007).

"Responding" is required from learners after they have been given sufficient material to comprehend a given objective (Gagne, 1985). For example, in a database lesson, "responding" might require a student to create a query that will count the number of records in a table in order to demonstrate comprehension of this newly introduced concept. The presence of "responding" enables students to reinforce their understanding. Effective practice should parallel the assessments that will be used to test the skills and knowledge reflected in an objective (Reiser and Dick, 1996). Another positive result of learning through practice is the motivation achieved through active participation and increased confidence in the objective tested (Dewald, 1999).

2.2 Behavioral Learning Theory

Behavioral learning theory includes several characteristics that should be present in an effective instructional design. These principles are contiguity, repetition, and feedback (Gagne et al., 1992). Contiguity is achieved when the response elicited from students follows the presentation of material as closely as possible. Students should be expected to perform this "responding" activity immediately after a learning objective is covered. Repetition increases the likelihood that students will retain information presented during a lesson. This can be achieved by an increase in the number of assignments that allow students to respond to many similar questions or tasks. Feedback occurs when an assignment is analyzed and answers are identified as correct or incorrect. Not only is it important to identify answers as correct, but an explanation of the incorrect answer and supporting rationale are essential (Debuse, Lawley and Shibl, 2007). Explaining both the correct answer and the faults of an incorrect answer are helpful when learners answer incorrectly (Kulhavy, 1977). Phillips, Hannafin and Tripp (1988) note that adequate feedback decreases the repetition of incorrect answers in the future and increases the probability of repeating correct responses. Schiller (2009) discusses that assessment and feedback are important in learner-centered teaching and there is a need for more of these formative feedback mechanisms.

2.3 Resource-based Views of Learning Theory

Rakes (1996) recommends increasing students' success through the addition of practice and feedback through a shift from the traditional theories of learning (cognitive and behavioral) to a resource-based view of learning. The resource view of learning involves the role of an instructor changing from an expert dispensing knowledge to a "guide" providing resources. As more online or web-enhanced courses become available, the need for this theory of instruction increases. Table 1 provides a comparison of the traditional and resources view of learning. The resource-based view of learning requires an increase in the number of problems, assignments, and exercises (Rakes, 1996). Finally, Yadin and Or-Bach (2010) discuss the continuing need for self-assessment and multiple individual exercises in an environment of collaborative learning.

Traditional Learning	**Resource-based Learning**
Teacher as an expert model	Teacher as a facilitator/guide
Textbook as primary source	Variety of sources/media
Facts as primary	Questions as primary
Information is packaged	Information is discovered
Emphasis on product	Emphasis on process
Assessment is quantitative	Assessment is qualitative/quantitative

Table 1: Traditional versus Resource Based Learning (Rakes, 1996)

In summary, a study of three different learning theories report that students' success is enhanced when they are given challenging, real-world practice assignments with rapid meaningful feedback. Following their guidelines the following key concepts were used in the development of the AGLS.

- Students' responses should immediately follow instruction to be effective.
- Multiple assignments of a similar nature should be presented repetitively to reinforce new material presented during a lesson.
- Immediate and customized feedback allows students to identify both correct answers and errors in incorrect answers.

3. NEED AND BENEFIT OF THE AGLS

Murray (1998) reports that increasing the number of exercises, problems, and assignments completed by students positively impacts content retention. Increased assignments, however, coupled with increasing enrollment in computer literacy courses implies that instructors are contending with a large amount of student work to grade. Detailed feedback on student work is beneficial-albeit critical-to their learning, but it is becoming increasingly difficult to provide adequate, timely feedback to students (Heinrich, Milne, Ramsay and Morrison, 2009). The types of assignments for introductory courses require a substantial amount of time to grade (Kay, 1998). Likewise, Tan (2009) and Janicki and Steinberg (2003) report that with increasing class size there is a real benefit to be gained by moving to automated grading. Heinrich et al. (2009) demonstrated how e-tools can be used to increase the efficiency and quality of assignment making.

The increased burden of larger classes on instructors has resulted in assessment and grading processes that do not support the previously discussed learning theory, specifically – personalized feedback. Therefore, in order to reduce the amount of time required to grade assignments, many professors may choose to give more easily graded assignments that do not adequately challenge the students.

3.1 Feedback Time and Quantity of Feedback Comments

The amount of feedback provided to students – measured by an average character count as well as time for feedback to be returned – was examined prior to introduction of a grading system. In a sampling of data taken from course management software used in the Fall of 2007 term, 130 of the 429 graded assignments (30.3%) had *no comments* from the instructor. This sampling excluded assignments that received a perfect score. Data from the same sample reveals that the average time between an assignment's due date and date graded was *28 days.*

It may also be the case that graduate assistants have graded assignments in conjunction with the professors. This fact can lead to inconsistent grading and confusing feedback (Ahonjemi and Karavirta, 2009). Each graduate assistant may have a different standard for grading and/or weights for a given assignment. For example, one grader may deduct three points for an incorrect formula in an Excel spreadsheet and provide no feedback while another may deduct only two points and comment that the student should have used an absolute reference instead of a relative reference. Whether there is one grader per assignment or several graders working together, there is an increased chance of human error. Kay (1998) states that "to preserve consistency it is best for a single individual to grade every student's response to a given question".

Grading errors can often go unnoticed and may lead students to believe their answer to be correct when it is actually incorrect. As described in the learning theory review, identification and correction of errors has a direct correlation to an increase in students' learning. Anglin, Anglin, Schumman and Kalinski (2008) reported that a 'grading rubric' is beneficial to student satisfaction. In their study, grading rubrics were 200% faster than traditional hand grading and lead to increased student satisfaction.

Because of the decrease in time professors have available per student, both the lack of adequate feedback and the failure of the current grading processes to allow practice and promote success, students' attitudes and motivations may suffer (Martin et al., 2007). This decrease in interest can result in an increase in the occurrences of plagiarism that often go unnoticed by assignment graders (Dodrill, Lidtke, Brown, Shamos and Fosberg, 1981).

Few assignments per learning objective
Less challenging assignments
Slow feedback
Generic (or no) feedback
Many different graders
Inconsistent grading
Inconsistent feedback
Grading errors
Increased plagiarism

Table 2: Summary of Current Problems in Introductory Computer Courses

The problems summarized in Table 2 do not promote effective learning. They are, in fact, contradictory to the requirements that promote effectual learning. In order to achieve student success, the resource-based learning approach states that there must be an introduction of more exercises that present a "real-world challenge" to students (Rakes, 1996).

3.2 Alternative Systems

There are alternative grading systems currently available to instructors. Each has its own limitations while all support the capability for increased assignments. It should be noted that prior research (Tan, 2009, Anglin et al., 2008, Debuse et al., 2007) discusses that the use of automated rubrics facilitate faster and increased feedback and that the use of all grading systems may be an advantage to instructors.

The first alternative system is *Case-Based Auto Graders*, such as CASEGRADER by Thomson Course Technology. Instructors are provided with a set of cases that can be quickly graded by the CASEGRADER system. These systems offer challenging, multi-step, realistic problems. Feedback is instantaneous and based on incorrect responses. Students are informed of their grade and feedback is given immediately following their submission of an assignment. A major limitation to this system is the inability of instructors to create their own cases (Crews and Murphy, 2008). CASEGRADER currently offers twelve (12) cases for the Office 2007 release.

Another grading system is *Procedural-Based Grading*, systems such as SAM2007 (2007) by Thomson Course Technology or SNAP by EMC Paradigm Publishing (2007). These alternative systems are applications that grade student responses (key strokes) based on the procedure used to reach the answer. The application may either be a web system or a software application that simulates the environment of Microsoft Office programs in order to provide a hands-on experience for the students. These systems usually incorporate smaller problems that attempt to reinforce a procedure to be remembered. A drawback is that these programs do not always include all methods of answering a problem and do not allow instant changes by the instructor like the AGLS does.

Additionally, simple *test-bank systems* normally exist within other systems such as Blackboard. These systems provide the instructor with the ability to create multiple choice, fill-in-the-blank, or paragraph/open-ended questions. They also do not allow skill-based assignments or assessments.

3.3 Comparison of Recommended System versus Current Offerings

In summary, the currently available grading systems provided by textbook and other publishers are not meeting the desired features and needs of faculty. Thus, there exists a need for a grading system that is adaptive to changing faculty needs and customized projects. This system should provide a challenging learning experience while relieving the time pressure from increased enrollment and time-intensive grading. The previously mentioned problems call for a system with a level of automation that allows instructors to quickly grade multiple complex assignments and provide quality feedback. If these needs can be met, then students can be presented with the necessary increased practice and rapid feedback required to promote effective learning.

In order to meet the need of students and instructors, a system known as the AGLS was developed. The AGLS consists of modules that provide automated grading of Microsoft Excel and Access assignments with personalized rapid feedback, shared assignment libraries, and plagiarism detection. The AGLS was developed to solve the problems identified in Table 2 while allowing for the complexity and quantity of exercises to be increased.

Table 3 provides a comparison of overall features and benefits between the proposed AGLS and the features supported by each of the commonly available automated grading systems. Table 4 details some of the limitations of the currently available grading systems that are not present in the AGLS

Benefits/Features	AGLS	Case-based	Procedural	Test-Bank
Challenging, real-world problems	■	■		
Automated grading	■	■	■	■
Consistent grading	■	■	■	■
Instant feedback	■	■	■	■
Customized feedback	■			
Web interface/portal	■	■	■	■
Multiple skills assessed concurrently	■	■		
Hands-on experience	■	■	■	
Smaller one-skill problems	■		■	■
Question/assignment library	■		■	■
Reduced preparation/paperwork time for instructor	■	■	■	■
Availability of student reporting	■	■	■	■
Expandable answer banks	■			
Repository for file submissions	■	■		
Plagiarism detection	■	■		
Instructor created exercises	■			■

Table 3: Comparison of the AGLS and alternative system benefits

Limitations	AGLS	Case-based	Procedural	Test-Bank
Answers must be exact matches		■	■	■
Limited number of cases		■	■	
Textbook/supplemental required		■	■	
Software must be installed			■	
"Simulated" environment			■	
Other purchases required		■	■	■

Table 4 Comparison of the AGLS and alternative e-system drawbacks

4. SYSTEM DESCRIPTION

The AGLS is comprised of grading and library components. The process used for grading in the AGLS is results-based. This is a key difference between the AGLS and other available systems (see Section 3, Alternative Systems section). The system grades actual answers that students provide, not the process or mouse clicks required to achieve those answers. Each gradable item (identified by the instructor) is associated with two lists. A list of correct responses allows acceptable answers to be marked accordingly. For example, in an Excel spreadsheet, a student may multiply cells C6 and C7 together by using =C6 * C7 or using =C7 * C6; both are correct. A list of incorrect responses is also coupled with appropriate feedback for the student. Figure 1 details the various components of the AGLS; the key adaptive portions are the correct answer, incorrect answer, and customized feedback components.

Appendix A diagrams the process flow for the grading functions of the AGLS. The process of the grading component of the AGLS begins when a comparison is made between a student's answer and the list of correct responses. If no match is found, the item is compared to the list of incorrect responses. If the answer has not been previously flagged as correct or incorrect, the instructor is prompted to identify the student's response as either correct or incorrect. The answer is added to the appropriate list. Figure 2 demonstrates the adaptive learning component, giving the grader more options on the 'correct' or incorrect answers as well as prompting for customized feedback on the particular answer offered by the student.

In the same prompt, the instructor may give partial or extra credit for the response and how much credit to award. With every answer, the instructor is also prompted to provide customized feedback that can be associated with an incorrect answer versus what the student answered. Appendix B lists the current gradable tasks. Specific details about the technologies employed by the AGLS can be found in Appendix C.

4.1 Shared Assignment Library

A feature of the system is that as assignments are built instructors may indicate that assignment as 'shared'. This feature gives other instructors of the same course the capability to share not only the assignment, but share the items to be graded and share the 'adaptive' learning portion of the correct and incorrect answers and their feedback. This implies that the 'incorrect answers' from one instructor are shared with other instructors on the same assignment. Thus all instructors' rubrics are updated with correct and incorrect answers.

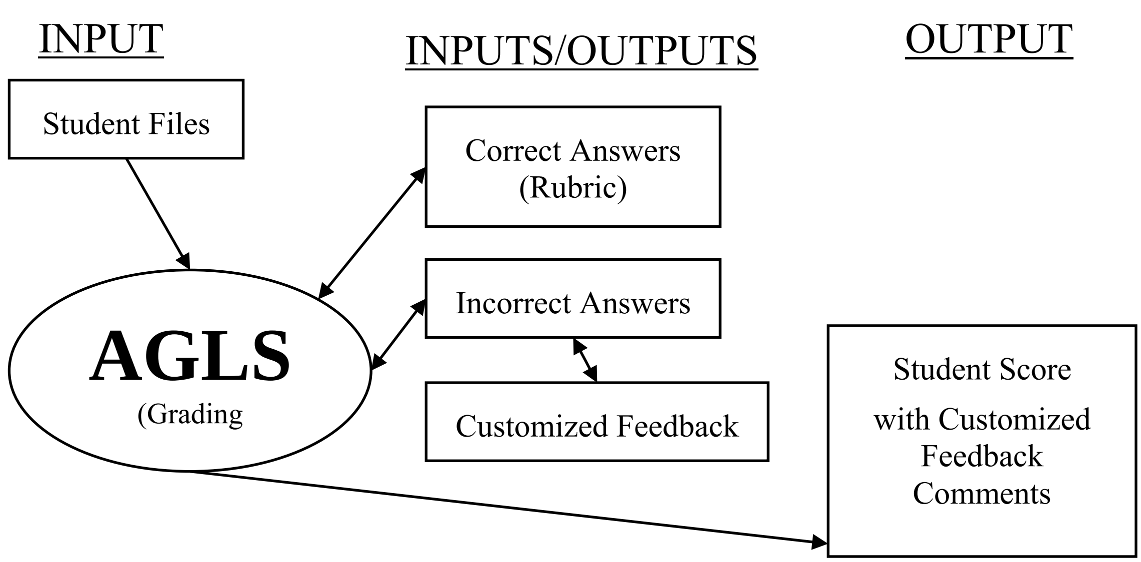


Figure 1: Input/output diagram for adaptive grading process of the AGLS

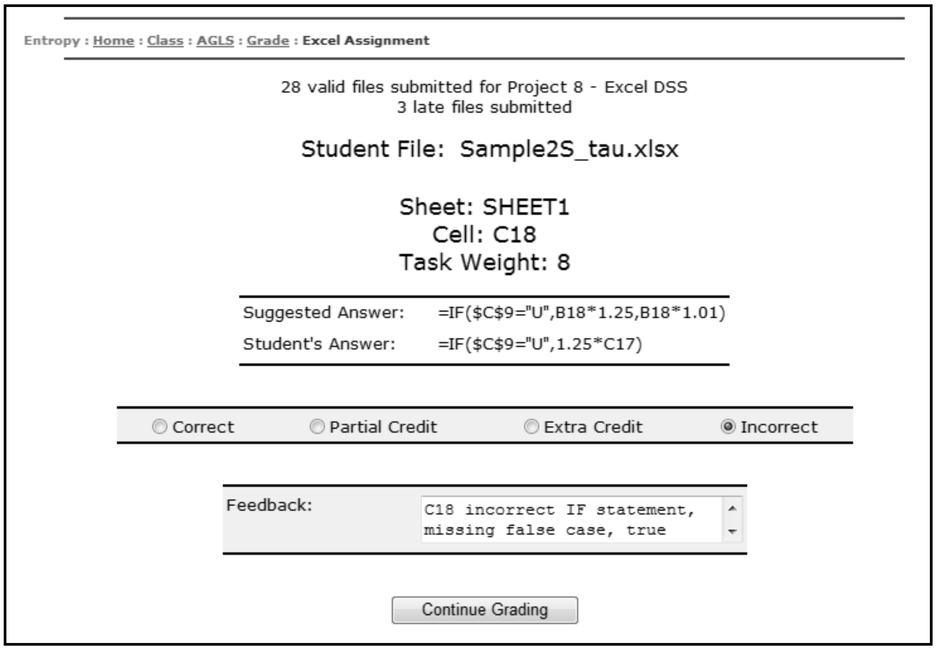


Figure 2: Instructor prompt to identify student answer as correct or incorrect and associated partial/extra credit and customized feedback

4.2 Plagiarism Detection

In addition to the grading services, instructors have access to the plagiarism detection tools provided by the AGLS for particular Office products. The methods of plagiarism detection differ based on the software product. For database assignments, the internal creation times of tables and queries are stored in the database for later comparison. A query is run on the creation times and those objects that have the same creation time are flagged by the plagiarism detection mechanism to be investigated. Comparisons may be made not only for one section but for all sections who implement the 'shared assignments' (even prior semesters). For spreadsheet assignments students download starting templates provided by the instructor. These are embedded with the student's ID in a hidden, password-protected sheet. When an assignment is uploaded, the AGLS compares the embedded downloader's ID with the ID of the student who is uploading the assignment. Thus not only is a potential 'cheater' identified, but also the individual who provided the file to them.

5. HYPOTHESIS DEVELOPMENT AND EXPERIMENT

To understand the benefits of the AGLS, an experiment was designed to test its impact on various aspects of feedback. Three specific hypotheses were developed and analyzed.

5.1 Data Sets

The setup of the experiment involved ten different sections of the same introductory MIS course. A pre-implementation set of five sections was examined and contrasted with a post-implementation set of five sections. The same instructors were involved in the pre- and post-implementation sections. The post-implementation sections used the system for all spreadsheet and database grading. The method of grading used for assignments was intended to be the only variable between the control and the experimental sets.

To limit bias that might impact their teaching or grading style, the pre- and post-implementation instructors were not made aware of the individual hypotheses under investigation. Specifically they were not aware that the authors were measuring the amount of time it took them to grade assignments, their quantity of comments, and they were not aware that there would be a review by an expert panel on the quality of the comments they generated. It should be noted that the post implementation instructors knew the grading system existed since they were involved in the actual grading of assignments.

5.2 Quantitative Data Gathering

Data for the quantitative tests came from an existing course management system used for recording grades in these sections. The fields of interest for the purposes of the experiment were student grades, feedback comments recorded, count of characters in feedback, grading response time (difference between date due and date graded), and the number of times a project may have been graded due to re-submissions. In the pre-implementation control set, student grades and feedback comments were manually entered. In the post-implementation set, the AGLS calculated student grades by starting with a perfect score of 100 and subtracting the number of points given to a gradable item when it was deemed incorrect. Feedback was manually entered into the AGLS once per answer in the post-implementation set and then duplicated for each student that gave the same answer. The count of characters in feedback, grading response time, and number of times a project was graded were all calculated queries based on data in the course management system.

To insure the reliability of the grades generated by the AGLS, they were compared to the scores generated by a set of graduate students who manually graded the projects. The results indicated that the AGLS grades were **generally lower** as it found more errors and graded more of the cells (Excel) or query entries (Access) than the manual grading. One might question if the grades were lower because the AGLS had an improper key and thus graded items incorrectly. This was found to occur at times, but it was observed that students would inform the instructor about correct answers being marked incorrect and thus the instructor had the opportunity to fix the key and regrade all assignments.

5.3 Qualitative Data Gathering

Qualitative data, such as quality of the generated feedback comments came from a panel of instructors from the instructional technology and information systems educational fields who were not involved in the delivery of the class. None of the qualitative evaluators had used the AGLS system. Draper, Brown Henderson and McAtee (1996) discuss that a panel of expert observers may be utilized to measure the effectiveness of CBT modules and their results. This panel was composed of ten instructors from the MIS department in the School of Business and from the Instructional Technology department in the School of Education.

This panel was asked to rate the quality of feedback comments. All grades with blank or no comments were eliminated. The number of assignments with zero comments provided to students in the course management system before the AGLS was implemented was 30.3%. This was reduced to only 1% for projects graded with the AGLS. For this study, the focus was determining the quality of comments with a character length greater than zero.

Fifty feedback items were randomly selected from the control set and fifty were randomly chosen from the experimental set (we eliminated assignments that received a grade of 100% as the comments for these projects were all similar to “Well Done” or “Good Job”). Panel members were instructed to assume that each assignment graded had errors, all errors were found, and the resulting feedback addressed those errors. All comments (pre and post) were compiled together so the members of the panel were unaware to which set, pre- or post-implementation, the comment belonged. The expert panel ranked the sample feedback on a scale of: Very Ineffective, Ineffective, Neutral, Effective to Very Effective.

5.4 Results and Analysis

The hypotheses developed were based on learning theory concepts that illustrate practice (“responding”) and feedback (“reinforcement”) as the events that are directly connected to student success (Gagne et al., 1992 and Martin et al., 2007). In addition, specific recommendations by other researchers are also included in this section.

5.4.1 Test 1: Effect on Quantity of Feedback

H_0: The use of the AGLS will **not affect** the quantity of feedback provided to students.

H_a: The use of the AGLS **will affect** the quantity of feedback provided to students.

Feedback quantity was measured by a count of characters used in comments given by instructors. Debuse, Lawley and Shibil (2007), Kaulhavy (1977), and Phillips, Hannafin and Trip (1988) all note that increases in the quantity of feedback assist in explaining the faults of an incorrect answer and decreases the potential for future incorrect answers.

The following pre-test conditions were considered prior to implementation of the *t*-test. The sampling distribution was considered normal or to be a near-normal distribution because the sample size is large (sample size 628 and 3138) and without outliers (Stattrek, 2010). Normal distribution can also be assumed because the Central Limit Theorem indicates that the distribution of an average will tend to be normal as the sample size increases. An F-test indicated that the data sets had unequal variances, and therefore a modified *t*-test with unequal variances was run. One other concern of the *t*-test is that it may be unreliable if the data set sizes are unequal. An additional *t*-test was run, with equal data sets (628) for both the control and experimental groups (for the experimental set every fifth data point was selected). This test indicated similar p-values approaching zero and therefore we verified that the *t*-test results were valid for the datasets.

Table 5 details the statistics from the *t*-test of character count data. This analysis supports a rejection of the null hypothesis and support of the accepting the alternative hypothesis; **the AGLS will affect the quantity of feedback provided to students** ($p= 4.11 \times 10^{-17}$, *t*-test= -8.53). In addition, the results indicate that the affect is that the AGLS will <u>increase</u> the quantity of feedback.

	Control Set	Experimental Set
Mean	45.10	71.36
Variance	4104.20	9213.33
Standard Deviation	64.06	95.99
Standard Error	2.56	1.71
Sample Size	628	3138
t Statistic	-8.53	
Two-tail p-value (unequal variance)	4.11×10^{-17}	

Table 5: t-test data related to character count of feedback

5.4.2 Test 2: Effect on Quality of Feedback

H_0: The use of the AGLS will **not affect** the quality of feedback provided to students.

H_a: The use of the AGLS **will affect** the quality of feedback provided to students.

Feedback quality was measured by the results from the survey of instructors not involved in the use of the AGLS. Rakes (1996) discusses the need for both qualitative as well as quantitative assessment techniques. The analysis shown in Table 6 supports an acceptance of the null hypothesis; **the AGLS will not affect the quality of feedback provided to students** to students (p= .261, *t*-test= -1.13). With an F-test value of 116, equal variance was assumed for this *t*-test.

	Control Set	Experimental Set
Mean	3.27	3.45
Variance	0.87	0.49
Standard Deviation	0.93	0.70
Standard Error	0.13	0.10
Sample Size	49	51
t Statistic	-1.13	
Two-tail p-value (equal variances)	0.261440502	

Table 6: t-test data related to expert panel review of feedback

This is a disappointing finding, as one of the reasons for the system was to increase the quality of feedback provided to students. However a potential reason for lack of change in the quality of the feedback may be that the same instructors who wrote the comments manually (prior to the system) into the course management system are the same individuals who entered the comments into the AGLS.

5.4.3 Test 3: Effect on Response Time

H_0: The use of the AGLS will **not affect** the amount of time for an assignment to be graded.

H_a: The use of the AGLS **will affect** the amount of time for an assignment to be graded.

	Control Set	Experimental Set
Mean	28.59	8.13
Variance	380.11	159.33
Standard Deviation	19.50	12.62
Standard Error	0.78	0.23
Sample Size	628	3138
t Statistic	25.26	
Two-tail p-value (unequal variances)	6.92×10^{-102}	

Table 7: t-test data related to response time of grading

Response time was measured by the difference between the due date of an assignment and the date a grade was issued by the instructor. The analysis shown in Table 7 supports a rejection of the null hypothesis and support of the accepting the alternative hypothesis; **the AGLS will affect the amount of time for an assignment to be graded** ($p = 6.92 \times 10^{-102}$, *t*-test= 25.26). In addition, by looking at the mean values from this *t*-test the use of the AGLS will <u>decrease</u> the amount of time for an assignment to be graded. Faculty still took an average of eight days to grade an assignment even when an automated solution was available; however, this is well below the average of 28.6 days prior to the system.

5.4.4 Effect on Number of Re-grading/ Errors/ Inconsistencies: An originally unseen benefit of the grading system was an increase in the number of times one assignment was graded by the instructors. A query of instructors indicated that they were offering 'pre-grading' on several assignments per semester. This permitted students to submit their projects early, receive a grade with feedback, and then learn from the feedback and resubmit the assignment. The gradebook system provided data on the number times a student submitted each assignment for grading.

Table 8 details the number of grade changes noted and demonstrate that 48% of the grades were re-graded when utilizing the AGLS. As noted earlier, this may be the result of offering students the option to submit their projects early for pre-grading. This process may allow students to learn during the resubmission process as they correct their assignments based on feedback. Due to the high response time and the amount of assignments, this pre-grading option was not feasible with manual grading. We could not identify a pattern or rationalization to justify why 20% of the grades were changed prior to the implementation of the AGLS. With the data available, it was unclear if instructors were changing 20% of their grades or if they were curving the results once all projects had been graded.

	Control Set	Experimental Set
Number in Grade Log	917	6067
Number of Grades Recorded	651	3133
Percent of Grades Changed	20.01%	48.36%

Table 8: Data related to the percent of grades changed

5.5 Limitations and Possible Solutions

The AGLS was tested in only one introductory to computer literacy course. The sections involved were introductory courses that teach spreadsheet and database skills and concepts. Originally, this was chosen to insure the AGLS was the only changing factor or variable in the two data sets. Similar tests across multiple disciplines and curriculums could further strengthen the impact of the AGLS on effective learning.

To further increase the flexibility for the system, a web service should also be investigated. This web service would allow an instructor to bypass the interactive process for grading. This would only be recommended when an adequate supply of correct and incorrect responses for each gradable item that has been identified. Online courses could benefit from the availability of this feature. An example where this web service may be useful is when an instructor uses the AGLS to allow students to upload a small practice exercise. These smaller problems can be immediately graded and feedback sent to the student and instructor for review. In order to create an effective web service portion of the AGLS, an extensively refined library including the majority of expected answers would need to be created. As the AGLS is utilized, its internal library will evolve and the web service enhancement can be addressed.

A few grading restrictions surfaced during the development of the AGLS. Table 9 lists the limitations and their associated Grading Module. During continual development and additions made to the AGLS, these problems will be addressed and incorporated into the system.

Grading Module	Limitation
Access	"ID" cannot be the Primary Key of a Table being tested in the AGLS.
Access	Cannot test Field Type for a Query.
Access	Forms cannot be accessed and graded.
Access	Reports cannot be accessed and graded.
Access	Formatting cannot be graded.
Excel	Cell values can only range from A1 to Z99.

Table 9: AGLS grading limitations and their associated Grading Modules

5.6 Future Research - Impact on student learning

The impact on overall student learning indicated by increased grades was one area the authors desired to study. For this research it was a challenge to develop a clean 'control group'. Once the AGLS was implemented, no instructors of the introductory course chose not to use it; therefore there was no available control group with the same assignments, homework, or instructors in the same academic year. The grades from previous semesters when the AGLS was not available were also compared; however, since instructors or assignments had changed during the two year period there was a possibility of instructor bias.

One impact on grades which is planned for a future study is the use of pre-grading. As noted earlier, over 33% of instructors offered pre-grading on assignments. This permitted students to submit their projects early. The projects are then re-graded and comments posted to assist the students. The students could resubmit by the due date for an updated grade. Future plans include measuring the impact of students who take advantage of pre-grading versus those who don't on future projects and tests in the class.

Another interesting extension of this research will be to determine if the same findings and benefits accrue from other automatic grading systems such as procedure systems and case-based grading systems. This research would involve finding other colleagues that have implemented other automated systems and are running similar experiments.

Finally, future research needs to encompass the means to increase the 'quality' of the feedback comments. Prior research indicates that enhanced feedback does increase learning, thus a better means to provide students higher quality feedback needs to be investigated and incorporated into the system.

6. CONCLUSIONS

From a literature review on student learning, feedback has been identified as a key component of student success. To promote effective learning, students need to be given multiple real-world exercises and need to receive rapid and meaningful feedback. With an increase in class sizes and the time involved in manually grading assignments, instructors have adopted methods that do not promote student learning.

To understand the impact of the AGLS on feedback, three hypotheses were developed. Two null hypotheses were rejected and one was accepted. Specifically, the data analysis from the three tests performed yielded the following results:

- The use of the AGLS ***will increase*** the quantity of feedback provided
- The use of the AGLS ***will not*** affect the quality of feedback provided
- The use of the AGLS ***will decrease*** the amount of time for a project to be graded.

A significant finding is that in the pre-implementation control set, over 30% of the students received no feedback on their projects, while in the post-implementation set, less than 1% received no feedback comments. In addition, the average grading lag was reduced by 20 days.

The implementation of the AGLS provides an innovative approach for automated grading – especially on assignments with more complexity. The system was built to support the learning concepts found in the literature of increased receptiveness and retention of knowledge by students. Prior research indicates that improved quantity of feedback, adaptive learning nature, and timely responses do facilitate student success over time. Although the authors were disappointed that the quality of feedback did not increase, the increased quantity and more timely feedback are significant benefits of the system.

REFERENCES

Ahonjemi, T. and Karavirta, V. (2009). Analyzing the use of a rubric-based grading tool. Proceedings of the fourteenth annual ACM SIGCSE Conference on Innovation and technology in Computer Science Education, 333–337.

Anglin, L., Anglin K., Schumman, P., and Kalinski, J., (2008). Improving the efficiency and effectivement of grading through the use of computer assisted grading rubrics. Decision Sciences Journal of Innovative Education, 6 (1), 51-73 .

Conole, G. and Warburton, B. (2005). A review of computer-assisted assessment. Research in Learning Technology, 13(1), 17–31.

Crews, T., and Murphy, C. (2008). CASEGRADER: Microsoft Office Excel Casebook with Autograding Technology. Boston, MA: Thomson Course Technology.

Debuse, J., Lawley, M., and Shibl, R. (2007). The implementation of an automated assessment feedback and quality assurance system for ICT courses. Journal of Information Systems Education, 18 (4), 491-502.

Dewald, N. (1999). Web-based Library Instruction: What Is Good Pedagogy? Information Technology and Libraries, 18 (1), 26-31.

Dodrill, W., Lidtke, D. K., Brown, C., Shamos, M., and Fosberg, M., (1981). "Plagiarism in computer sciences courses (Panel Discussion)," Proceedings of the twelfth SIGCSE technical symposium on computer science education, 26-27.

Draper, S., Brown, M., Henderson, F. and McAtee, E. (1996). Integrative evaluation and emerging role for classroom studies of CAL. Computers Education, 26 (1), 13-20.

EMC/ Paradigm Publishing (2007). College Catalog: Computer Technology: Snap Web-based Training and Assessment: Snap 2007. (Online). Retrieved: 7/1/2010: http://www.emcp.com/product_catalog/index.php?GroupID=1926 [October 13, 2007].

Gagne, R. (1965). The Conditions of Learning (1st Ed.). New York: Holt, Rinehart and Winston.

Gagne, R. (1985). The Conditions of Learning (4th Ed.). New York: Holt, Rinehart and Winston.

Gagne, R. (1988). Mastery Learning and Instructional Design. Performance Improvement Quarterly, 1 (1), 7-18.

Gagne, R., Briggs, L. and Wager, W. (1992). Principles of Instructional Design (4th Ed.). Fort Worth, TX: HBJ College Publishers.

Heinrich, E., Milne, J., Ramsay, A., and Morrison, D. (2009). Recommendations for the Use of E-tools for Improvements around Assignment Marking Quality. Assessment and Evaluation in Higher Education, 34 (4), 469 - 479.

Janicki, T., and Steinberg, J. (2003) Evaluation of a Computer – Supported Learning System. Decision Sciences Journal of Innovative Education. 2 (1), 203-223.

Kay, D. (1998). "Large Introductory Computer Science Classes: Strategies for Effective Course Management," ACM SIGCSE Bulletin, 30, 131-134.

Kulhavy, R. (1977). Feedback in written instruction. Review of Educational Research, 47 (1), 211-232.

Martin, F., Klein, J. and Sullivan, H. (2007). The Impact of Instructional Elements in Computer-Based Instruction. British Journal of Educational Technology, 38 (4), 623 - 636.

Murray, T., (1998). Authoring Knowledge Based Tutors: Tools for Content, Instructional Strategy, Student Model and Interface Design. Journal of the Learning Sciences, 7 (1), 5-64.

National Research Council (1999). BeFIT: Being Fluent in Information Technology. Washington, DC: National Academy Press.

Phillips, T., Hannafin, M., and Tripp, S. (1988). The effects of practice and orienting activities on learning from interactive video. Educational Communication and Technology, 36, 93-102.

Rakes, G. (1996). "Using the Internet as a tool in resource based learning environment". Educational Technology, 6 (2), 52-29.

Reiser, R. and Dick, W. (1996). Instructional Planning: A guide for teachers (2nd Ed.). Allyn and Bacon Publication, Columbus OH.

SAMS2007 (2007). Course Technology - SAM 2007 Training V1.0. Retrieved: 3/17/2010: http://www.course.com/catalog/product.cfm?isbn=978-1-4239-1305-4

Schiller, S. (2009), Practicing Learner-Centered Teaching: Pedagogical Design and Assessment of a Second Life Project. Journal of Information Systems Education, 20 (3), 369 – 382.

Stattrek (2010), Hypothesis Test of the Mean, (Online), Retrieved: 7/1/2010: http://stattrek.com/Lesson5/Mean.aspx

Tan, C. (2009), Assessment via WebCT Quizzes: Offline Grading Process with Customized Feedback, Decision Sciences Journal of Innovative Education. (1), 321-326.

Yadin, A., and Or-Bach, R. (2010). The Importance of Emphasizing Individual Learning in the "Collaborative Learning Era", Journal of Information Systems Education, 21 (2), 185-194.

AUTHOR BIOGRAPHIES

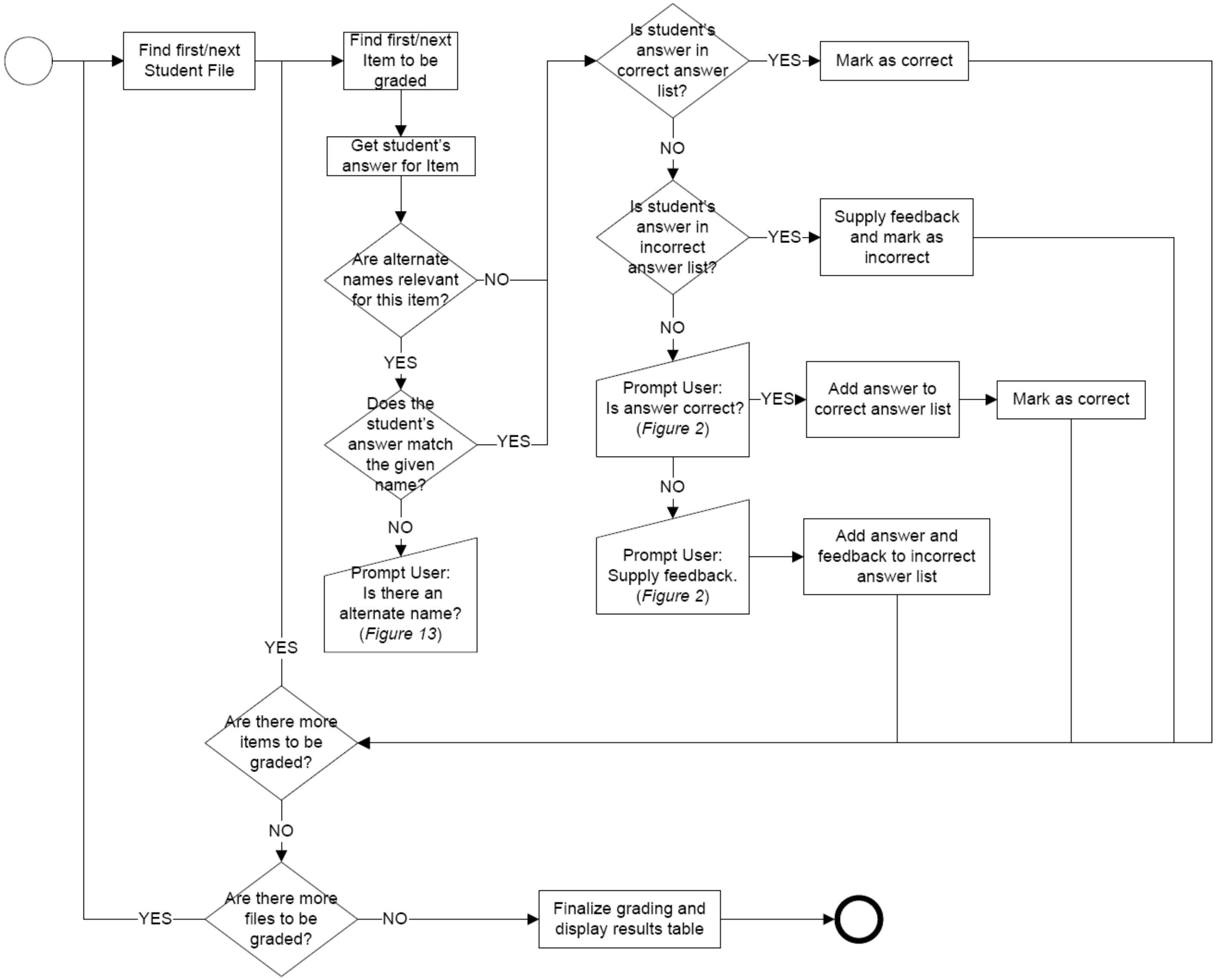
Kevin Matthews is an instructor of Information Systems at the University of North Carolina Wilmington. His teaching interests include web development and web technologies, programming concepts and logic, and business application development. He has interests for research in learning theory, technology-aided teaching and pedagogy, learning systems, and agile software development.

Thomas N Janicki is the Progress Energy/Gordon Hurbert Professor of Information Systems at the University of North Carolina. His primary research area involves utilizing web-based technologies to enhance the traditional classroom-learning environment. He is the corporate advisory board coordinator which brings in over twenty five business professionals from the community into the department to assist in program development and increased opportunities for students. He has published over 25 articles in journals such as Decision Science Institute Innovative Education, International Journal of Innovation and Learning and Information Systems Education Journal.

Ling He is currently an Associate Professor of Accounting in the Department of Accounting, College of Business and Management at Saginaw Valley State University. Her research interests include accounting information systems, managerial accounting, database management systems, information security, electronic commerce, and PAC learning theory. She has published papers in Decision Support Systems, Journal on Computing and Journal of Database Management.

Laurie J. Patterson is an associate professor of Computer Science at the University of North Carolina Wilmington as well as the coordinator of the Information Technology minor. Her primary research areas involve how technology is used in academic settings and differences in technology use between genders. She has published in many conference proceedings and in IEEE's Computer.

APPENDIX A: Process Flow of Adaptive Grading in the AGLS



APPENDIX B: Gradable Tasks

MS Access Gradable Items

Table names
Field attributes (names, types, default values)
Field Sizes (where appropriate)
Primary Keys
Data entered into tables and the result of queries
Query names
Fields shown in queries
Criteria (=, < >, <. >, < =, > =, LIKE, *, %, BETWEEN, AND, OR)
Parameter Queries (using [ ])
Calculated columns/aggregate functions (SUM, COUNT, AVERAGE, GROUP BY, etc.)
Number of rows of data (min, max, must match)
Sorting for reports
Properties for fields of a report

MS Excel Gradable Items

Static Cell contents (strings, numbers, etc.)
Cell text formatting (alignment, bold, italics, underline, font face, type, etc.)
Cell type (percent, currency, text, decimal places, etc.)
Sheet formatting (page orientation, etc.)
Formulas (basic arithmetic, functions (SUM, MIN, MAX, etc), IF statements)
Cell references (absolute, relative, combination/partially absolute)
Charts (title, chart type, etc.)
Sheet names (Sheet1, Sheet2, Answer Report, etc.)
Scenario Manager (scenario summary, changing cell, scenario names, scenario values, result cells, scenario result values)
Solver (default values, constraints, target cell, adjustable cells, answer report, extension cases, answer report answers, final value answers)

APPENDIX C: Technologies Employed

MS SQL 2005/2008 Database

Language: ASP.Net (VB)

Keys to grading:

Excel:

1) Student files are renamed with .zip extension
2) Compressed files are unzipped which exposes XML files (format, sheets, formulas, styles, etc)
3) Particular XML files are parsed to determine:
    a. Styles (formatting)
    b. Sheets (what sheets are part of worksheet)
    c. Values (specific values in cells)
    d. Formulas (relationships)
    e. Absolute or Relative Addressing
    f. Graph existence and graph type
    g. Graphing ranges
    h. Scenario properties
    i. Solver properties

Access:

Student files are opened in one of three way to expose properties needed for grading

- ADOX – exposes catalog to permit investigation of table names, query names, field names field types, sizes, formatting
- ADO – for properties and records in queries without parameters
- OBDC – exposes those queries that have input parameters (permits entry of parameters for grading

Handling of misspellings or various field / table names

A key hurdle in grading Access projects was the coding of ‘alternative names’ for table names, field names. Students liked to enter the same field with various naming (i.e. instructorID, instructor_ID, instructor ID). The system needed to flag an error if the naming was different, but it then needed to know and keep track of the incorrect name in order to check for field size, type, values etc. If the incorrect name was not considered the system would mark many items in error since it could not find the correct name in the catalog. The system now permit instructors to indicate if alternative names may be accepted (and only penalized once) and continue drilling down for additional properties.

STATEMENT OF PEER REVIEW INTEGRITY

All papers published in the Journal of Information Systems Education have undergone rigorous peer review. This includes an initial editor screening and double-blind refereeing by three or more expert referees.

ISSN 1055-3096