

Rubrics on the Fly: Improving Efficiency and Consistency with a Rapid Grading and Feedback System

Adam C. Moyer¹, William A. Young II^{2,*}, Gary R. Weckman³, Red C. Martin², and Ken W. Cutright²

Abstract: Many learning management systems (LMS) used in higher education provide customizable rubrics that aid in the process of grading and providing feedback for many forms of assessments commonly used by educators today. Rapid Grade (RG) is a grading and feedback feature built into a non-commercialized LMS developed by a large, public, Midwestern university in the United States. In this research, RG was compared to a grading and feedback system found in one of the most utilized LMS found in higher education. It should be noted that the name of this particular LMS is not named. Using the Technology Acceptance Model (TAM) to validate that RG empirically improves upon existing methods, survey results indicate that RG is a significant improvement in terms of ease of use and usefulness when grading and providing feedback for a given assessment. The RG framework, as well as the specific results of the TAM, are presented.

Keywords: teaching, learning management systems, rubrics, system development.

Introduction

One of the most time-consuming tasks from a teaching perspective is the process of grading assessments that have been submitted by students. These duties would be much less exhausting if the act providing useful feedback were omitted; however, constructive criticism is a major part of the educational equation (Hall, et al., 2001; He, Hui, & Quan, 2009). In this paper, a methodology related to the student-teacher feedback loop will be presented. Before outlining the scope of the paper, elements of the student-teacher feedback loop will be defined. The feedback loop is straightforward. Students are given assessments that have been prepared by a subject matter expert. They are then asked to complete and submit their assessments often through a learning management system (LMS). Based on the type of assessment given, instructors evaluate the students' performance by providing some sort of numeric score. In addition to this score, instructors provide written feedback so that the student can learn from their mistakes or successes (Butler, 2011).

From an instructor's perspective, the type of assignments chosen is often based on a few considerations. For example, instructors might choose one type of assessment over another simply based upon the size of the class. If the class size is large, instructors may choose scalable methods like multiple-choice questions. These types of assessments are not always easy to create, but can be graded easily by a LMS. However, some instructors could choose other types of assessments, such as reports or projects, if the size of the class is not large. In many cases, these types of assessments usually take much longer to evaluate, due to the variety of answers that are submitted.

¹Department of Management Information Systems, College of Business, Ohio University, Athens, Ohio 45701

²Department of Management, College of Business, Ohio University, Athens, Ohio 45701, youngw1@ohio.edu

³Department of Industrial and Systems Engineering, Russ College of Engineering and Technology, Ohio University, Athens, Ohio 45701

With that said, the most recent trend in higher education is that the student-to-faculty ratios are increasing (Spinellis, Zaharias, & Vrechopoulos, 2007). Thus, providing consistent, useful feedback in a timely manner will only become more challenging for instructors in the future.

Of course, there are many other things that instructors take into account when making decisions on what type of assessment to use. The decision could be based on the type of material being covered in the class. If, for example, an instructor is teaching a course in computer programming, some assessment types, like ones built with multiple-choice questions, may not be the most accurate choice of assessment in order to evaluate a student's knowledge of learning outcomes (Cheang, Kurnia, Lim, & Oon, 2003). Certainly, other factors come to mind when deciding what type of assessments should be given to students. Examples include whether or not grading assistance is available to instructors; the time that the instructor has to dedicate to grading; and whether technology can be used and managed effectively. Though only a few reasons why instructors choose certain types of assessments were provided, there is usually one common thread among all assessment types. This thread is the act of providing constructive criticism and/or assigning points to the particular assessment in question. In order to simplify the remaining portions of this manuscript, this act will simply be referred to from this point forward as "grading."

Providing consistent grading and feedback on a particular assessment is often a difficult challenge for instructors. While grading assessments, there are many things that might contribute to inconsistencies. Providing consistent feedback is perhaps the biggest challenge instructors face because inconsistencies can potentially mean that those students do not share a common experience within the classroom. For example, depending on the complexity and the number of students submitting assignments, providing feedback can be very time-consuming. Too often, grading for extended periods can result in inconsistent assessment and feedback (Klein & El, 2010). This fatigue occurs when instructors are too tired to continue grading and must often take breaks from the work. This period of discontinuous work can result in inconsistent feedback, since some of the comments and/or points that were given to previously completed work are not as fresh in the instructor's mind as they once were. Using predetermined evaluation methods, such as rubrics, is one way to reduce the inconsistencies of grader fatigue. However, this is only part of the solution (Ramey, VandeVusse, & Gosline, 2007). Having a good rubric can go a long way towards making grading less time consuming for instructors, as well as providing feedback that is more consistent to students; however, there are certainly limitations (Cross, 1990).

The scope of this work includes the presentation of a framework that is both easy to use, easy to understand, and can be built in an LMS, which allows instructors to provide faster, more consistent grading and feedback to students. This framework, which will be referred to as Rapid Grade (RG) from this point forward, has been implemented into an LMS that has been used within a college of a large, public, non-profit Midwestern university in the United States. In order to validate the proposed methodology, expert users of the LMS which RG has been built into is benchmarked against a well-known and highly utilized LMS platform within higher education. It should be noted that the specific name of the vendor of this popular LMS is intentionally not mentioned within this manuscript. Thus, this particular LMS will simply be referred to as XYZ henceforth. Therefore, RG will be benchmarked against XYZ based on survey results, which will be more formally described in a later section, in order to determine the perceived usefulness and ease of use of the RG technology.

Presenting a new framework that could be developed within any LMS environment is significant in many ways. For example, instructors can leave feedback that is more consistent at a more rapid pace than what is currently offered in popular LMS used in higher education. This is

especially true when class sizes are large, or when an instructor is teaching multiple class sections at a time. The implemented framework also allows instructors to start the grading process without a pre-existing rubric. In other words, instructors can develop “rubrics on the fly” with the grading process ongoing. Simply put, the framework is robust and flexible enough to offer students with more consistent feedback, which from an instructor perspective is faster and easier to use than the benchmark LMS, or XYZ.

Literature Review

A review of literature is provided to obtain a broad view of what others have studied as it relates to rubrics as a grading and feedback system. This review provides an overview of these systems and goes more in-depth, with an overview of current state-of-the-art technologies and a brief description of the general deficiencies of these systems. Finally, since the framework of RG is presented as a LMS methodology, a brief overview of the software development strategies used to develop the framework is presented.

Feedback and Rubric Systems

There are many benefits to utilizing rubrics. For example, rubrics can be used by teachers to communicate their expectations to students; students to self-assess their work; by teachers who want to communicate learning objectives and outcomes with other teachers; and, of course, by teachers to provide grading and feedback, which is the focus of this article.

There are many ways to provide feedback to students. In terms of using rubrics as a feedback mechanism, there is a wide variety of strategies found in research. For example, the use of rubrics within quantitative courses is rather low (Riddle & Smith, 2008). However, some practitioners have used rubrics within these types of courses successfully (Riddle & Smith, 2008). In fact, they not only list a classification of rubrics into various categories such as checklists, analytical, and hybrid-based rubrics, they also provide insight as to which rubric type works best for the assignments given.

As previously noted, rubrics can be used as a feedback method and have been shown to improve teaching (Cooper & Gargan, 2009; Wolf & Stevens, 2007). However, in some cases there is a lack of quantitative research on the ability of the rubric value to enhance academic performances. In one such study, rubrics were found to be more significant to academic performance than college year, major, pre-test score, and gender (Howell, 2011).

A rubric is simply a set of criteria that is used to help articulate the gradations of the quality of work the student has demonstrated in an assessment (Goodrich, 1997). Rubrics as feedback mechanisms are also useful to students, as they clearly state what an instructor is looking for in order to demonstrate a mastery level of course content. Providing students access to the rubric before an assignment is submitted can make the assessment process to a student very clear (Andrade, 2000). Therefore, since the baseline of expectations has been established, providing consistent feedback is often easier (Andrade, 2000).

Current State of Rubrics as Feedback Systems

There have been astonishing findings when it comes to using customized feedback systems. For example, one study showed that using computer-assisted grading rubrics were almost 200% faster than traditional hand grading without rubrics, more than 300% faster than hand grading with

Journal of Teaching and Learning with Technology, Vol. 4, No. 2, December 2015.
jotlt.indiana.edu

rubrics, and nearly 350% faster than typing the feedback into an LMS (Anglin, Anglin, Schumann, & Kaliski, 2008). Not only did this particular system reduce the time required from the teacher perspective, their results indicated that utilizing an online system did not negatively affect students' attitudes or their overall satisfaction of receiving feedback. Though these types of systems are not widely used, there are popular systems that are used within GK-12 and higher education. For example, according to market share, Blackboard®, Moodle®, and Desire2Learn® (CampusComputing, 2013) are the top three LMSs used today, with a market share of 41%, 23%, and 11%, respectively. Based on these commercial options, some researching practitioners have studied the effectiveness from a student and teacher perspective with success. For example, one study showed that from a teacher's perspective there is a 40% reduction of marking time and improved student satisfaction with feedback (Atkinson & Lim, 2013).

Teachers do not always have the luxury of having access to custom-built or commercial LMS. Even if they do, commercial systems are limited in terms of their capabilities especially when it comes to more subjective assignments (Thompson & Ahn, 2012). Therefore, another option for teachers is to utilize rubric systems that are available online. These systems allow teachers to build, share, and utilize pre-developed rubrics for the classroom. For example, these sites include RubiStar (4Teachers.org, 2012), Rubrics for Teachers (TeacherPlanet.com, 2012), and TeAch-nology (2012).

Deficiencies with Current Rubrics as Feedback Methods

Some researchers state that one limitation of using rubrics as a feedback system is that rubrics are not always self-explanatory (Andrade, 2005). Not only is this a limitation to students, but it is also a limitation for instructors. Rubrics that are unclear often are not reliable and could result in grading inconsistencies even for the same body of work (Andrade, 2005). Therefore, though rubrics can improve consistency with feedback and make the process faster for instructors, they do have their limitations. However, some studies show that these limitations are improved upon when rubrics are co-developed, or when they go through a process of peer-review before they are used as a grading tool (Andrade, 2005).

Computers can assist with the process of grading and providing feedback for student assessments (Chen, 2004). However, there is certainly room for improvement, as it is often very cumbersome to work with a modern LMS in order to provide individualized feedback. For example, if an instructor is using XYZ as their LMS for a particular class, and the instructor is trying to view all assignments that were submitted, several steps are needed in order to provide a grade and feedback for an individual student. This process of completing several steps is then repeated for every student, which can be a tedious, draining task.

Another limitation of using rubrics is that it is very difficult to accommodate every grading scenario into the pre-defined rubric categories. The limitation here is that if pre-defined categories are used within an LMS environment, instructors often do not have a quick solution to provide students with individualized comments. Even if customized comments are left, they are often not available to reuse when a similar grading scenario occurs. For example, instructors are at the mercy of having a complete rubric before the grading process occurs. If there are issues that were not discovered when the rubric was created, then an instructor is left with a decision to make. Does the instructor correct the rubric for the current student and utilize it to grade all the remaining assessments, or does the instructor take the time to go back and redesign the rubric to make all of the necessary changes and then grade all of the assessments that were once complete? In addition,

if an instructor is using a rubric within an LMS across multiple class sections, then it is not easy to apply the changes made in one class section to another. LMSs also do not allow users to modify previously graded assignments easily if changes are made to the rubrics (Dornisch & McLoughlin, 2006).

Software Development Overview

The development of the proposed LMS feedback framework called RG is based on two existing software development approaches called Systems Development Life Cycle (SDLC) and Rapid Application Development (RAD). Therefore, a brief overview is provided.

Systems Development Life Cycle (SDLC)

SDLC is an approach that can be described in general terms as a process for planning, creating, testing, and deploying information systems (Centers for Medicare & Medicaid Services, 2008). The processes within SDLC can be broken down into phases, activities, and tasks in order to achieve the overall goal of the project (US Department of Justice, 2012). The motivation behind SDLC is that projects can be managed in order to ensure that users get what they want in a timely and cost-effective manner (Necco, Gordon, & Tsai, 1987). The primary phases of the SDLC include initiation, system concept development, planning, requirements analysis, design, development, integration and test, implementation, operations and maintenance, and disposition.

Rapid Application Development (RAD)

RAD describes a methodology for building information systems. In comparison to SDLC, it is more of a user-centered approach (Mackay, Carne, Beynon-Davies, & Tudhope, 2000). For example, in RAD, users are more closely associated with the development process. In this process, users are considerably more integrated into the process than just supplying information and signing off on systems that have been developed. In RAD, users actively participate in the development process, with the motivation being that the applications that are built are completed in a short amount of time (Abdullah, Mateen, Sattar, & Mustafa, 2010). Development comes in the form of iterative prototyping (Beynon-Davies, 2000). In general, RAD is a form of incremental and iterative development, where previous versions of a system or application are tested by users and feedback is gathered to make incremental improvements upon previous versions of the system (Avison & Fitzgerald, 2003).

Methodology

The emphasis for the remainder of this manuscript is on the systems development framework of a feedback and grading methodology called RG. In part of the presentation for this methodology, RG will be explained through technology examples as they relate to the LMS features available to instructors. However, before this framework is presented, a brief introduction of RG is given, followed by the system development methodology, which will be referred to as **PBR²** (or **Plan, Build, Run, and Repeat**). Finally, survey results will be investigated with the Technology Acceptance Model (TAM) (Davis, 1989) used to compare the perceived usefulness and ease of use between RG and XYZ, which is one of the most widely-used LMS in higher education (CampusComputing, 2013).

Rapid Grade Overview

RG is a part of a larger LMS system called **ISMS**, or **I**ntegrated **S**ite **M**anagement **S**ystem, which has been developed and is currently being utilized by a college in a large Midwestern university. More specifically, ISMS, and therefore RG, is built using Microsoft C# programming language, and has been deployed on a Microsoft ASP.Net web server environment.

There are many functions of the LMS component of ISMS, which could be discussed. However, the scope of this section is strictly limited to the functionality of RG. *Figure 1* shows a screen capture of RG when it was utilized to grade a particular assignment in a management information systems course. The screen capture shows a partial roster of the class, which has been blurred in order to protect the identity of the students. The screen capture also shows basic functionality such as controlling which students are selected, which in turn controls other functions such as adding points, assigning points, or assigning comments to the selected students. Also shown in the figure is a button for RG, which is located below the numeric scores for each student. Once activated, a secondary window is opened, which is located in the bottom right portion of the figure. The secondary window shows the total points available for an assignment as well as the pre-defined rubric dimensions, which are relevant to the particular assignment. For example, under the total possible points available, the first dimension that can be assigned by a simple click of the mouse is “All Requirements Met,” which, if clicked, would mean that zero points would be taken off the student’s assessment. However, please note that this is not the case for the particular example shown in the figure. What is selected is the first dimension that is applied to the student’s grade, “Code: Inadequate Comments”, followed by the rubric comments that students would see once they selected the particular assignment on the LMS. While grading an assignment, instructors can simply select any available rubric dimension that applies to the student’s assessment. If a given rubric dimension is not available, or has not previously been defined, instructors can simply add rubric dimensions by hitting the “Add a New Dimension” button. Even after the creation of a new rubric dimension, instructors can simply hit the “Edit” button to make changes that will be retroactively applied to students who were assigned a given dimension on their assessment. In fact, if personalized feedback is needed, rather than the default feedback that was created when the rubric was later added in, then an instructor can simply hit the “Edit” button and has the option to apply customized feedback for a given instance, rather than changing the default rubric dimension for all the students. The point deductions can be modified in a similar fashion.

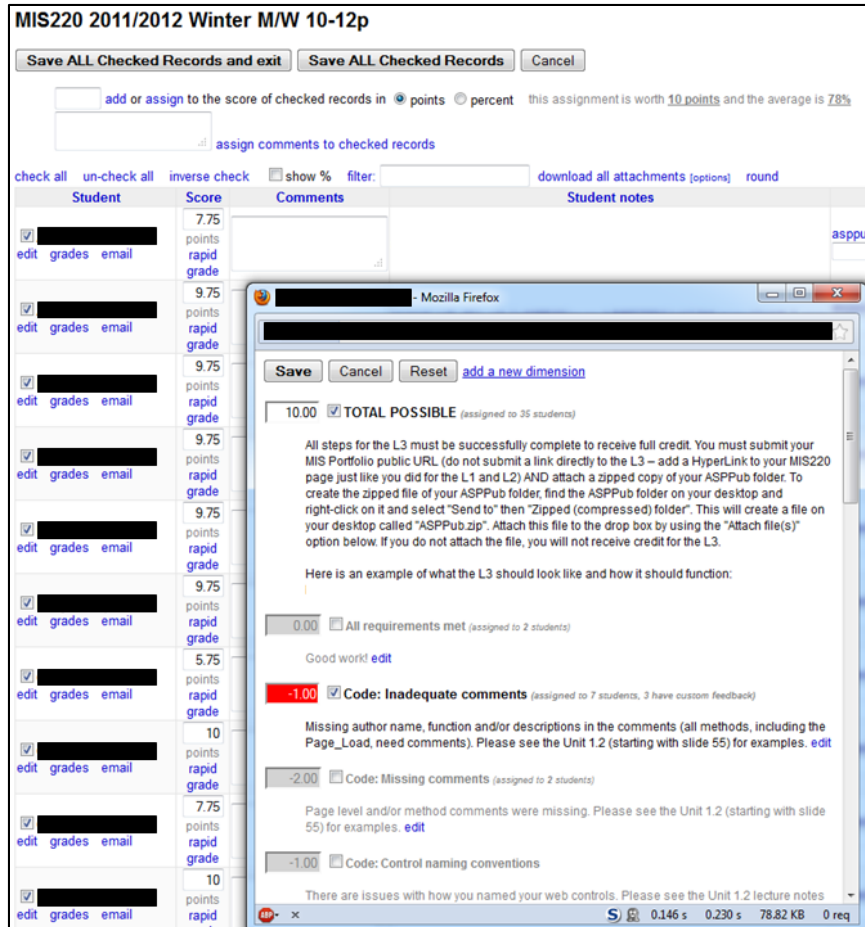


Figure 1. Example Rapid Grade

PBR²

RG is a novel feedback and grading system in comparison to the current state-of-the-art technologies offered on websites and LMS environments. Ultimately, the goal of RG is to create an easy-to-use grading system that provides feedback that is more consistent for instructors and students. In order to develop this system, a hybrid systems development technique was created which merges parts of SDLC and RAD. This hybrid approach will be called **PBR²**, which stands for **Plan, Build, Run, and Repeat**.

The remaining sections provide a more thorough explanation of **PBR²** as it relates to the development of RG. Thus, examples of RG will be given in order to provide an illustration of how **PBR²** was utilized in the development process. However, before these sections are presented, a brief overview of **PBR²** is presented. In brief, **Planning** involves understanding the players or those whom the solution will help, as well as the non-technical problems and user requirements. Requirements are prioritized into “need” and “want” categories, and then a strategy is developed to produce a potential solution. **Building** involves creating a functioning prototype through a constant feedback loop involving the players. While building, a prototype is made available for testing, comments, and constant feedback. **Running** is described as the point where the system is made available to users. However, it is important to note that anything not considered a “need” is

moved into the Repeat phase. **Repeating** allows the process to start over and build upon what has already been implemented. For example, the “wants” from a prior iteration are simply re-evaluated and are potentially classified as “needs” in a later iteration. If a prior “want” is re-classified as a “need,” the feature requirement is then built into the Build and Run phases. Thereafter, repeating provides a constant feedback loop to improve the system continuously.

Plan

In terms of PBR² as it relates to the development of RG, a few considerations of the Plan phase are shown below. In this case, instructors are considered the users of RG. For example, instructors were asked for input in order to develop a feedback system. The process of categorizing their “needs” and “wants” were assessed. For example, the users stated that they “wanted” or “needed” a feedback system that:

- improved the performance of the user’s ability to provide feedback through enhanced consistency and speed of the grading process - a **“need”**
- is easy to understand and use - a **“need”**
- allows for easy and immediate historic feedback comparisons to permit continuous improvement - a **“want”**
- addresses situations where the student has no computer access - a **“want”**

In terms of the brief list above, two states were classified as a “need” and two were categorized as a “want.” If a particular feature is determined to be a “need,” the aspects related to the request are addressed in the planning stage. Features determined as a “want” are simply not addressed until the next iteration, or until the Repeat phase of PBR² is invoked.

For the first iteration of the development of RG, a prototype was developed. This plan is shown in *Figure 2*. To provide this explanation, the assumption will be made that an instructor does not have a pre-defined rubric to utilize for an assessment that needs grading. For example, when an instructor starts the grading process, they must make a decision about whether or not to create a new rubric dimension. If this is desired, the instructor simply creates the rubric dimension, where the description that will be available to the student is defined with a corresponding point deduction. The instructor will continue to grade additional questions, with new dimensions added as they are needed. Again, the assumption is that there is not a pre-defined rubric available to the instructor. However, if one was available, all of the rubric dimensions could be added before the grading process started. Assuming that the instructor has added all of the dimensions relevant to the first student’s assessment, the instructor has decisions to make while grading the next student’s work. For example, if the rubric dimension is available and applies to the next student’s work, then the instructor can simply click on the dimension that needs to be assigned to the student’s assessment. The instructor would simply create additional dimensions as they are needed.

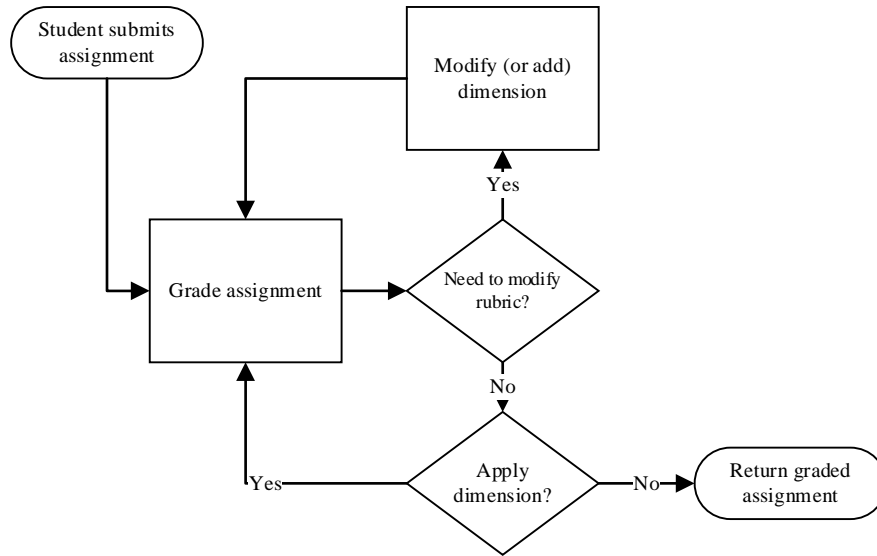


Figure 2. Example of Adding New Dimensions

During the decision making process shown in *Figure 2*, the instructor must decide whether or not a new dimension is to be created, or if a previous one needs to be modified. In terms of adding a new dimension, this has already been discussed. However, what was not discussed was the act of modifying an existing dimension. Before presenting the slightly more complex decision process shown in *Figure 3*, where the instructor can decide if customized feedback is used, a simple example of modifying an existing rubric dimension is given. Given that rubric dimensions have been created, instructors have the ability to edit the dimensions at any time. These edits might be related to a more descriptive comment associated with a given rubric or a change in point values. The more complicated scenario of modification is presented in *Figure 3*. The added complexity is associated with the decision of how edits should be managed and applied across the course roster. For example, assume for a given assessment that about half of the student assignments have been graded and the instructor realizes that a change is necessary. Perhaps the instructor wants to take off more or fewer points on a given dimension, or perhaps the comments are not as detailed as desired. The instructor can choose either to apply the changes to a dimension across all previous assignments that have been graded, or to make a decision to customize feedback or point values to the specific student assignment, which is currently under review. This customized feedback is useful for many situations, including adding a personalized response to a student, which could ultimately promote an instructor's social presence in the classroom. In any case, giving instructors the power of retroactively applying changes to previously graded assignments saves time and improves the accuracy of grading course assignments.

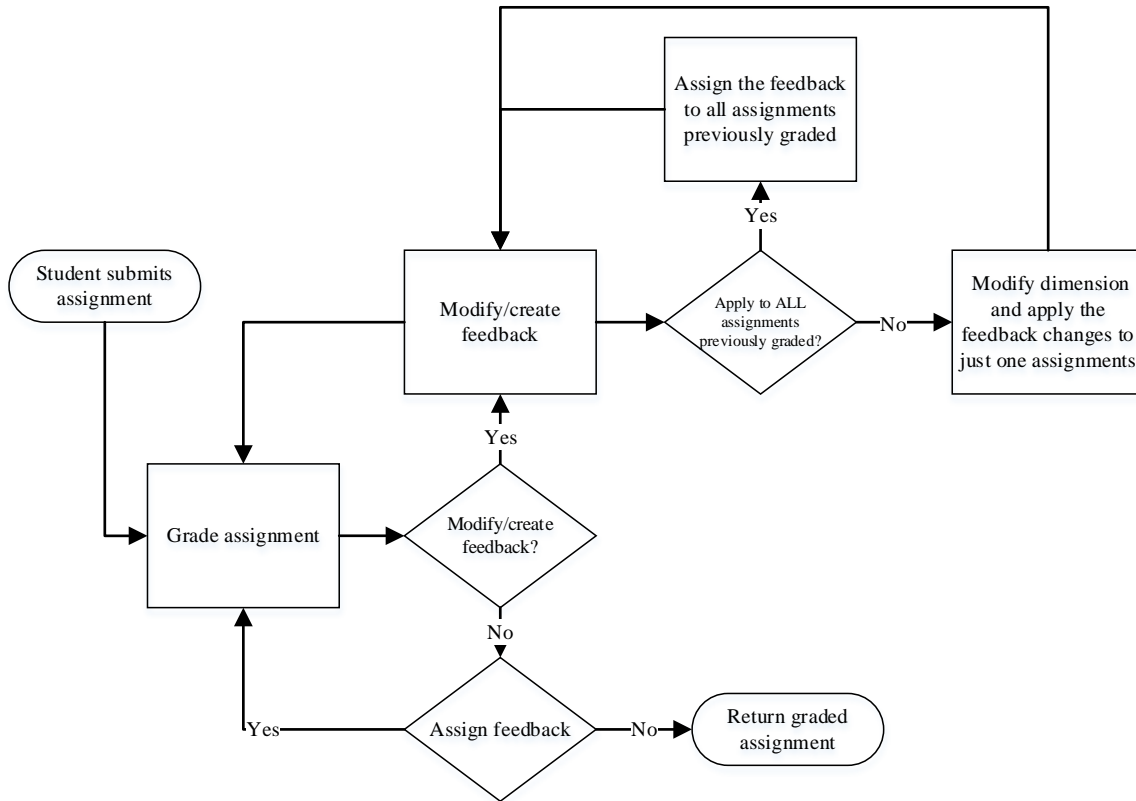


Figure 3. Example of Rubric Dimension Modifications

Build

Based upon the input from instructors and teaching assistants, a prototype of RG was developed to address the “needs” listed in the Plan phase. During the Build phase, a simple yet effective feedback loop was followed, which is shown in *Figure 4*. Simply put, as prototypes were built and reached a stable state, feedback was requested from a small team of users. If the users were not satisfied with the solution presented, then more input was collected and the application was revised. This process continued until the user group was satisfied. Once this occurred, the component of the prototype was released to a larger group of users, which occurred in the Run phase.

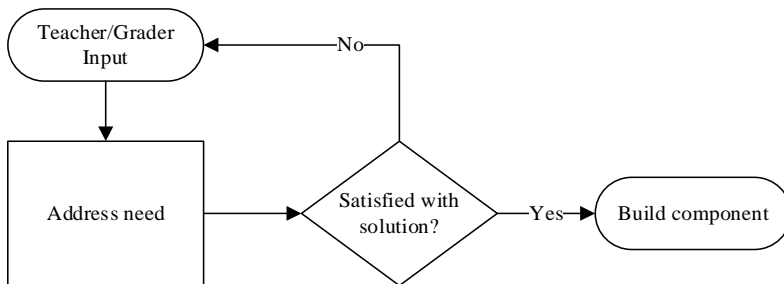


Figure 4. Example of Build Feedback Loop

Run

From the development side of the process, the Run phase is nothing more than making the software application available to a larger base of users. During this phase, users work with the technology so that ideas can be generated for future enhancements and feature additions. With that said, a pool of “needs” and “wants” are generated by the users and are later assessed when the process is repeated.

For the remainder of this particular section, examples of features built into RG that were deemed acceptable by the small group of users in the Build phase of the project are presented. For example, the screenshot shown in *Figure 5* displays various controls that can be set by the instructor that relate to the visibility of the rubric to students. In this particular screenshot, a question is proposed to the instructor. The question is asking the instructor whether the rubric developed for the assessment should be hidden or made available to students prior to the student submitting their work.

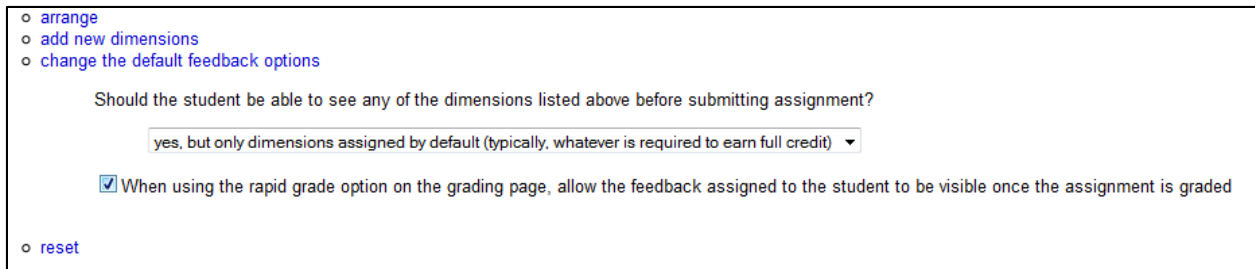


Figure 5. Feedback Visibility Controls

When users were asked whether they built rubrics before an assessment was to be graded or during the grading process, the feedback was mixed. Since the process of creating rubrics on the fly was previously discussed, the following figure shows an example of when rubric dimensions were built before the grading process. As one can see, the window for creating new rubric dimensions is straightforward. Users define the point values, which can be positive or negative. Positive point values might mean a bonus, whereas negative values would deduct points from a given assessment. The user then fills out a name of the new rubric dimension and gives a description. As noted, these dimensions can be edited later.

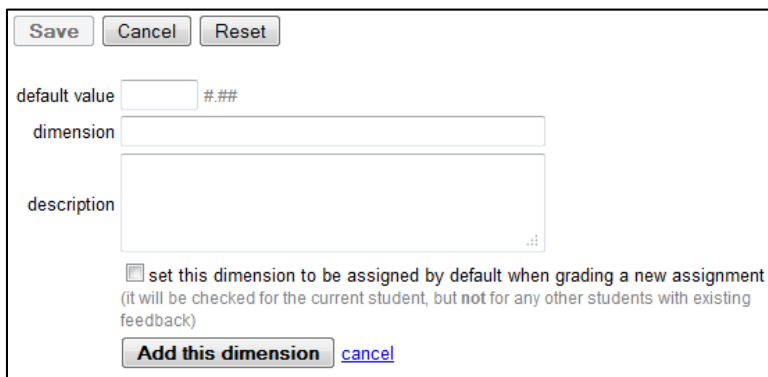


Figure 6. Creating a New Rubric Dimension

Figure 7 shows one way in which rubrics can be managed. In this figure, users can modify the visibility setting as well as a number of other features. For example, users can add or delete dimensions from the “Manage” page. They can also select whether or not the dimension is selected by default when RG is used to grade a student’s assignment. In addition, point values, as well as textual information, can be edited. One important feature included in this screen capture is the “Arrange” button. This simply allows instructors to change the order of the rubrics that have been developed. This is useful in making the grading process more user-friendly. In other words, dimensions can be aligned in such a manner that is directly related to the questions being assessed.

Should the student be able to see rubric before submitting assignment? | yes, but only dimensions assigned by default (typically, the requirements)

default value	dimension	description	add	reset	arrange	
	<input checked="" type="checkbox"/>	30.00	Requirements	1.) Show the code and description of products whose warehouse quantity is smaller than the average warehouse quantity		
	<input type="checkbox"/>	-3.00	Query 6	This query required pulling fields from ViewPurchaseOrder and filtering by using a where clause on "supplierNum in (00004,00006,00009)".		
	<input type="checkbox"/>	-3.00	Query 7	By creating a new query which joined ViewPart.SubCode and Product.prodcutCode (just like we did in class - see Lecture 14 slides		
	<input type="checkbox"/>	-3.00	Query 8	By creating a new query which joined ViewPart.SubCode and Product.prodcutCode (just like we did in class - see Lecture 14 slides		
	<input type="checkbox"/>	-15.00	Missing Queries	You were missing responses for 1-3, 7 & 8. Did you forget to do these or did you not understand how to do them? Let me know if you		

Figure 7. Editing Existing Dimensions

A final example of a feature of Rapid Grade is shown in Figure 8. This screen capture demonstrates the process an instructor would use when grading an assignment. Though the screen capture shows a limited number of dimensions, the instructor would simply need to select the appropriate dimensions that would apply to the assignment being graded.

-3.00 Query 8
By creating a new query which joined View

-15.00 Missing Queries
You were missing responses for 1-3, 7 & 8

-30.00 Requirements not met
Other than creating ViewWarehouseInvent

20.00 Partial Credit
Partial credit was offered for re-submitting

24 / 30 (total points)

[if you need to add a new dimension, click here](#)

Figure 8. Assigning existing feedback

Repeat

During the final phase of PBR², feedback is solicited from the pool of users. Ideally, the size of the pool would be based upon the scope of the feature requests made, as well as the size of the development team. Once sufficient feedback is obtained, the iterative process of PBR² is repeated.

Results

In order to develop an effective solution that meets the pedagogical needs of the faculty and students, instructional technologies should be validated (Machado & Tao, 2007). Since there is often confusion between verification and validation, a definition is provided. Verification is the act of ensuring that the system is built correctly. In other words, does the system function as it was intended to function from a technical perspective? Validation, on the other hand, is the act of ensuring that the right system has been built. In other words, does the technology accomplish the goals of the users (Miser & Quade, 1988)? In order to validate that RG functions more efficiently than feedback and grading mechanisms within XYZ, a survey based on the TAM (Davis, 1989) was sent to higher education instructors who have had experience using both RG and XYZ. Thus, the results that will be presented in this manuscript are based upon a convenient sample. Before presenting the remaining portions of the results, a brief overview of TAM is provided.

Technology Acceptance Model

It should be noted that acceptance models are in short supply for researchers developing technology (Davis, 1989). Researchers claim that most models to validate systems are based upon subjective measures and suggest that they are often invalid. Thus, the TAM was chosen in order to test the hypothesis that RG improves upon the functionality of XYZ as a grading and feedback-based methodology.

The TAM survey consists of fifty-six factors, twenty-eight of which are dedicated to the evaluation of RG, and the other twenty-eight to XYZ. The model consists of a set of validated questions that provides quantitative insight as to the perceived level of usefulness of technology, as well as the perceived ease of use (Davis, 1989). Perceived usefulness is the extent to which a user of the system expects the system to improve personal efficiency. Perceived ease of use describes the extent to which a system is easy to use.

This survey was developed in and distributed from ISMS. This decision was simply made out of convenience to collect the survey data electronically. Each question of the TAM is based upon a linear, five-point Likert scale (Likert, 1932).

Data Collection

In total, twenty-three survey participants responded to the survey who had experience in both online systems. A one-tailed paired t-test was used to determine if there was a statistical difference between RG and XYZ for any categories and questions found in the TAM. For hypothesis testing, it was suggested that t-tests provide a good indication of the Likert response location (Meyers, Gamst, & Guarino, 2005). As noted, TAM consists of two evaluation categories, with the first being the perceived usefulness and second being the perceived ease of use. Each category consists

of fourteen factors. It should be noted that when the factors were presented to the participants, they were not randomized. The factors, which will later be presented in squared brackets, were presented in numeric order to the survey participants.

The survey utilized the standard convention of a 5-point Likert scale, where each response category was assigned a linear value. For example, “Strongly Disagree,” “Disagree,” “Undecided,” “Agree,” and “Strongly Agree” were assigned integer values of 1, 2, 3, 4, and 5, respectively. Before running this analysis, the aggregated correspondent data was preprocessed in a manner such that a positive difference between RG and XYZ would indicate that RG was a potential improvement over XYZ. Likewise, if the difference were negative, this would indicate that XYZ was a potential improvement over RG.

Data Preprocessing

The data was preprocessed to order to make interpreting the results easier to understand. For example, if the individual paired responses for XYZ were subtracted from RG (i.e. $RG_i - XYZ_i$), it would be logical that a positive difference would indicate that RG was an improvement over XYZ. For example, the first factor that appears in the survey is “My job would be difficult to perform without the technology.” Assuming one survey participant replied with a value of 5 (i.e. “Strongly Agree”) for RG and a 3 (i.e. “Undecided”) for XYZ, the difference would be 2, which would indicate that the survey participant more strongly agreed with the notion that RG made his or her work easier to perform. On the other hand, if a survey participant replied with a value of 3 for RG and a 5 for XYZ, the difference would be -2, which would indicate that the survey participant more strongly agreed with the factor that XYZ made his or her work easier to perform.

Based upon the wording of some factors that appeared in the survey, an adjustment in terms of the point values assigned to the Likert scales needed to be made in order to keep the results consistent for the purpose of interpretation. For example, the fifteenth factor is “I often become confused when I use the technology.” Thus, if a survey respondent indicated an answer of 5 (i.e. “Strongly Agree”) for RG and a 3 (i.e. “Neutral”) for XYZ the difference would be 2, but the interpretation of +2 has a different meaning. In this particular example, a +2 would **not** indicate that RG was an improvement over XYZ; rather, it would actually indicate that XYZ was an improvement over RG, meaning that the system was less confusing to work with for this particular example. Thus, to overcome these issues the values for the 5-point Likert were adjusted so that “Strongly Disagree,” “Disagree,” “Undecided,” “Agree,” and “Strongly Agree” were assigned values of 5, 4, 3, 2, and 1, respectively. Thus, after this adjustment was made, the interpretation of positive and negative differences would be consistent based upon any factor in the survey. Finally, Table 1 and Table 2 show the numbering convention used after the data preprocessing was complete. Factors that the standard Likert score changed (i.e. flipped), are indicated by an asterisk (i.e. *).

Table 1. Perceived Usefulness Factors (RG-XYZ)

#	Factor
1.	My job would be difficult to perform without the technology.
2.	Using the technology gives me greater control over my work.
3.	Using the technology improves my job performance.
4.	The technology addresses my job-related needs.
5.	Using the technology saves me time.

-
6. The technology enables me to accomplish tasks more quickly.
 7. The technology supports critical aspects of my job.
 8. Using the technology allows me to accomplish more work than would otherwise be possible.
 9. Using the technology reduces the time I spend on unproductive activities.
 10. Using the technology enhances my effectiveness on the job.
 11. Using the technology improves the quality of the work I do.
 12. Using the technology increases my productivity.
 13. Using the technology makes it easier to do my job.
 14. Overall, I find the technology useful in my job.
-

Table 2. Perceived Ease of Use Factors (RG-XYZ)

#	Factor
15.	I often become confused when I use the technology.*
16.	I make errors frequently when using the technology.*
17.	Interacting with the technology is often frustrating.*
18.	I need to consult the user manual often when using the technology.*
19.	Interacting with the technology requires a lot of mental effort.*
20.	I find it easy to recover from errors encountered while using the technology.
21.	The technology is rigid and inflexible to interact with.*
22.	I find it easy to get the technology to do what I want it to do.
23.	The technology often behaves in unexpected ways.*
24.	I find it cumbersome to use the technology.*
25.	My interaction with the technology is easy for me to understand.
26.	It is easy for me to remember how to perform tasks using the technology.
27.	The technology provides helpful guidance in performing tasks.
28.	Overall, I find the technology easy to use.

Data Analysis

Ultimately, to determine if RG was an improvement over XYZ on average (i.e. μ), a paired, two-sample, one-tailed t-test was needed to make a conclusion about the statistical hypothesis. In particular, the null hypothesis (or the status quo) is that XYZ outperforms or is at least equal to the mean performance of RG. Thus, the alternative factor is that the mean performance of RG outperforms, or is an improvement to, the XYZ system. Therefore, the two-sample, one-tailed hypothesis statement is shown below, where RG and XYZ are abbreviated as RG and XYZ, respectively.

- $H_0: \mu_{RG} \leq \mu_{XYZ}$
- $H_1: \mu_{RG} > \mu_{XYZ}$

After stating the hypothesis test, the significance of the test was selected. For this particular study, an alpha value of 0.05, or 5%, was selected. As described in the previous section, the results from the survey respondents were preprocessed and the difference between RG and XYZ

responses were determined. Thus, the two-sample hypothesis test shown above can be rewritten in an equivalent form to:

- $H_0: \mu_{RG} - \mu_{XYZ} \leq 0$
- $H_1: \mu_{RG} - \mu_{XYZ} > 0$

A statistical hypothesis test derived from the TAM data collected can be performed in many ways. The strategy of presenting the results in the following section will start at a high aggregated level, followed by additional analyses that are more detailed. Thus, the structure of the results is as follows. First, an overall test of twenty-eight paired factors of RG – XYZ will be conducted and presented. In this case, the “overall case,” both the perceived usefulness and ease of use are grouped together. Second, factors associated with the perceived usefulness are aggregated, analyzed, and presented. Third, factors associated with the ease of use are aggregated and the testing results and discussions are presented. Fourth, results of the individual factors are presented. The presentation of the general descriptive statistics will be shown along with the paired, two-sample, one-tailed t-test results.

Overall

Table 3 shows the descriptive statistics for the overall test conducted from the survey results. The sample size for this test is 644, which represents the twenty-eight factors on perceived usefulness and the ease of use from the twenty-three survey participants. From this table, the mean of the difference between RG and XYZ is positive, a value of 0.9829, which suggests that there is evidence to support the t-test analysis. In an overall sense, RG performs better than XYZ as a grading and feedback tool.

Table 3. Overall Descriptive Statistics

N	Mean	SE	Mean	StDev	Coef Var	Sum of Sqrs.	Minimum	Q1	Median	Q3	Maximum	Range	Skewness
644	0.9829	0.0591	1.5	152.61	633	2069	-4	0	1	2	4	8	-0.36

To determine if the alternative hypothesis that the performance of RG is greater than XYZ is correct, a paired, two-sample, one-tailed t-test was performed, and the results are shown in Table 4. Based upon this aggregated study, the p-value is less than α of 0.05, which indicates that the null hypothesis can be rejected. Thus, it appears that RG is an overall improvement to the grading and feedback system of XYZ.

Table 4. Overall Test for $H_0: \mu_{RG} \leq \mu_{XYZ}$, $H_1: \mu_{RG} > \mu_{XYZ}$

N	Mean	StDev	SE Mean	95% Lower Bound	T	P
644	0.9829	1.5000	0.0591	0.8856	16.63	0.000

Perceived Usefulness

From the previous section, the testing suggests that RG is an improvement over XYZ; however, in order to provide more detailed results, this section presents the results of just the questions related to the perceived usefulness of the online system. Based upon the descriptive statistics found in Table 5, the sample size is 322, which is based upon the fourteen factors associated with the perceived usefulness and the twenty-three survey participant. The mean of this sample is 0.7267, with a standard deviation of 1.3603.

Table 5. Perceived Usefulness Descriptive Statistics

N	Mean	SE Mean	StDev	Coef Var	Sum	Sum of Sqrs.	Minimum	Q1	Median	Q3	Maximum	Range	Skewness
322	0.7267	0.0758	1.3603	187.18	234	764	-3	0	1	2	4	7	-0.64

Figure 7 shows the results of the t-test where, like the overall test, the p-value is less than alpha. This indicates that from a perceived usefulness standpoint based on the TAM, RG is more useful than XYZ on an aggregated level.

Table 6. Perceived Usefulness Test for H0: $\mu_{RG} \leq \mu_{XYZ}$, H1: $\mu_{RG} > \mu_{XYZ}$

N	Mean	StDev	SE Mean	95% Lower Bound	T	P
322	0.7267	1.3603	0.0758	0.6017	9.59	0.000

Ease of Use

Thus far, statistical testing indicates that on a high, aggregated level, RG is an improvement over XYZ in terms of its overall performance and from a perceived usefulness standpoint. In addition to these tests, factors related to the ease of use were investigated, and the descriptive statistics for this sample are shown in Table 7. Like the perceived usefulness sample, the sample size was 322, which is based upon the fourteen factors associated with the ease of use and the twenty-three survey participants. The mean of this sample is 1.2391, with a standard deviation of 1.5891.

Table 7. Ease of Use Descriptive Statistics

N	Mean	SE Mean	StDev	Coef Var	Sum	Sum of Sqrs.	Minimum	Q1	Median	Q3	Maximum	Range	Skewness
322	1.2391	0.0886	1.5891	128.24	399	1305	-4	0	1	2	4	8	-0.33

Table 8 shows the results of the t-test. Similar to the overall and perceived usefulness, the p-value is less than alpha, indicating that from an ease-of-use standpoint, RG is perceived to be better than XYZ on an aggregated level.

Table 8. Ease of Use Test for H0: μ RG \leq μ XYZ, H1: μ RG $>$ μ XYZ

N	Mean	StDev	SE Mean	95% Lower Bound	T	P
322	1.2391	1.5891	0.0886	1.093	13.99	0.000

Individual Factors

Prior testing was based upon the aggregated results or the overall results, as well as being based upon the category (i.e. perceived usefulness or ease of use) and factors associated with those categories from the TAM. This section investigates the individual factors related to the survey. Therefore, to be consistent with the presentation of the prior tests, Table 9 shows the descriptive statistics for all twenty-eight factors that appeared on the survey.

Table 9. Individual Factors Descriptive Statistics

#	N	Mean	SE Mean	StDev	Coef Var	Sum	Sum of Sqrs.	Minimum	Q1	Median	Q3	Maximum	Range	Skewness
1.	23	0.174	0.342	1.642	944.06	4	60	-3	0	0	1	4	7	0.44
2.	23	0.870	0.262	1.254	144.24	20	52	-3	0	1	2	3	6	-1.25
3.	23	0.957	0.277	1.331	139.12	22	60	-3	1	1	2	3	6	-1.18
4.	23	0.217	0.259	1.242	571.12	5	35	-3	0	0	1	3	6	-0.30
5.	23	1.261	0.290	1.389	110.15	29	79	-3	1	1	2	4	7	-1.07
6.	23	1.348	0.285	1.369	101.55	31	83	-3	1	2	2	4	7	-1.40
7.	23	0.087	0.288	1.379	1585.60	2	42	-3	-1	0	1	3	6	-0.17
8.	23	0.435	0.265	1.273	292.79	10	40	-3	0	1	1	3	6	-0.64
9.	23	0.870	0.290	1.392	160.04	20	60	-3	0	1	2	3	6	-1.08
10.	23	0.739	0.261	1.251	169.26	17	47	-3	0	1	1	3	6	-1.13
11.	23	0.652	0.248	1.191	182.65	15	41	-3	0	1	1	2	5	-1.37
12.	23	1.130	0.297	1.424	125.97	26	74	-3	1	1	2	4	7	-0.77
13.	23	0.826	0.241	1.154	139.71	19	45	-3	1	1	1	3	6	-1.57
14.	23	0.609	0.272	1.305	214.43	14	46	-3	0	1	1	4	7	-0.26
15.	23	1.609	0.360	1.725	107.24	37	125	-3	1	2	3	4	7	-0.96
16.	23	1.478	0.301	1.442	97.54	34	96	-1	0	2	2	4	5	0.15
17.	23	2.043	0.324	1.551	75.92	47	149	-1	1	2	3	4	5	-0.40
18.	23	1.130	0.423	2.029	179.53	26	120	-4	0	1	3	4	8	-0.77
19.	23	1.261	0.303	1.453	115.22	29	83	-2	1	1	2	3	5	-0.69
20.	23	0.609	0.249	1.196	196.51	14	40	-2	0	1	1	3	5	-0.19
21.	23	1.478	0.273	1.310	88.60	34	88	-1	0	1	3	4	5	0.05
22.	23	1.217	0.332	1.594	130.95	28	90	-1	0	1	3	4	5	0.20

23.	23	1.174	0.249	1.193	101.61	27	63	-1	0	1	2	3	4	-0.01
24.	23	1.478	0.349	1.675	113.32	34	112	-2	0	1	3	4	6	-0.09
25.	23	0.957	0.390	1.870	195.53	22	98	-3	0	1	2	4	7	-0.39
26.	23	0.870	0.357	1.714	197.08	20	82	-3	0	1	2	4	7	-0.37
27.	23	0.565	0.280	1.343	237.52	13	47	-2	0	0	2	3	5	0.16
28.	23	1.478	0.349	1.675	113.32	34	112	-3	0	2	3	4	7	-0.98

Table 10 shows the t-test results for each factor on the TAM used in this investigation, which is summarized more clearly in Table 11.

Table 10. Individual Questions Test for H0: $\mu_{RG} \leq \mu_{XYZ}$, H1: $\mu_{RG} > \mu_{XYZ}$

#	N	Mean	StDev	SE Mean	Bound	T	p-value
1.	23	0.174	1.642	0.342	-0.414	0.51	0.308
2.	23	0.870	1.254	0.262	0.420	3.32	0.002
3.	23	0.957	1.331	0.277	0.480	3.45	0.001
4.	23	0.217	1.242	0.259	-0.227	0.84	0.205
5.	23	1.261	1.389	0.290	0.764	4.35	0.000
6.	23	1.348	1.369	0.285	0.858	4.72	0.000
7.	23	0.087	1.379	0.288	-0.407	0.30	0.383
8.	23	0.435	1.273	0.265	-0.021	1.64	0.058
9.	23	0.870	1.392	0.290	0.371	3.00	0.003
10.	23	0.739	1.251	0.261	0.291	2.83	0.005
11.	23	0.652	1.191	0.248	0.226	2.63	0.008
12.	23	1.130	1.424	0.297	0.621	3.81	0.000
13.	23	0.826	1.154	0.241	0.413	3.43	0.001
14.	23	0.609	1.305	0.272	0.141	2.24	0.018
15.	23	1.609	1.725	0.360	0.991	4.47	0.000
16.	23	1.478	1.442	0.301	0.962	4.92	0.000
17.	23	2.043	1.551	0.324	1.488	6.32	0.000
18.	23	1.130	2.029	0.423	0.404	2.67	0.007
19.	23	1.261	1.453	0.303	0.741	4.16	0.000
20.	23	0.609	1.196	0.249	0.180	2.44	0.012
21.	23	1.478	1.310	0.273	1.009	5.41	0.000
22.	23	1.217	1.594	0.332	0.647	3.66	0.001
23.	23	1.174	1.193	0.249	0.747	4.72	0.000
24.	23	1.478	1.675	0.349	0.878	4.23	0.000
25.	23	0.957	1.870	0.390	0.287	2.45	0.011
26.	23	0.870	1.714	0.357	0.256	2.43	0.012
27.	23	0.565	1.343	0.280	0.085	2.02	0.028
28.	23	1.478	1.675	0.349	0.878	4.23	0.000

Based on the results shown in Table 11, in only four of the twenty-eight total factors is there not enough evidence to reject the null hypothesis that RG was an improvement on XYZ as a

grading and feedback system. These questions are 1, 4, 7, and 8, which correspond to “My job would be difficult to perform without the technology,” “The technology addresses my job-related needs,” “The technology supports critical aspects of my job,” and “Using the technology allows me to accomplish more work than would otherwise be possible”. It should be noted that the p-value associated with Question 8 was very close to alpha; however, as it currently stands, there is not enough evidence to conclude that RG allows users to accomplish more work than would otherwise be possible. The remaining questions, however, are all statistically significant wherein the null hypothesis can be rejected.

Table 11. Individual Factors Hypothesis Summary

#	Question	Reject Null	p-value
1.	My job would be difficult to perform without the technology.	No**	0.308
2.	Using the technology gives me greater control over my work.	Yes	0.002
3.	Using the technology improves my job performance.	Yes	0.001
4.	The technology addresses my job-related needs.	No**	0.205
5.	Using the technology saves me time.	Yes	0.000
6.	The technology enables me to accomplish tasks more quickly.	Yes	0.000
7.	The technology supports critical aspects of my job.	No**	0.383
8.	Using the technology allows me to accomplish more work than would otherwise be possible	No*	0.058
9.	Using the technology reduces the time I spend on unproductive activities.	Yes	0.003
10.	Using the technology enhances my effectiveness on the job.	Yes	0.005
11.	Using the technology improves the quality of the work I do.	Yes	0.008
12.	Using the technology increases my productivity.	Yes	0.000
13.	Using the technology makes it easier to do my job.	Yes	0.001
14.	Overall I find the technology useful in my job.	Yes	0.018
15.	I often become confused when I use the technology.*	Yes	0.000
16.	I make errors frequently when using the technology.*	Yes	0.000
17.	Interacting with the technology is often frustrating.*	Yes	0.000
18.	I need to consult the user manual often when using the technology.*	Yes	0.007
19.	Interacting with the technology requires a lot of mental effort.*	Yes	0.000
20.	I find it easy to recover from errors encountered while using the technology.	Yes	0.012
21.	The technology is rigid and inflexible to interact with.*	Yes	0.000
22.	I find it easy to get the technology to do what I want it to do.	Yes	0.001
23.	The technology often behaves in unexpected ways.*	Yes	0.000
24.	I find it cumbersome to use the technology.*	Yes	0.000
25.	My interaction with the technology is easy for me to understand.	Yes	0.011
26.	It is easy for me to remember how to perform tasks using the technology.	Yes	0.012
27.	The technology provides helpful guidance in performing tasks.	Yes	0.028
28.	Overall I find the technology easy to use.	Yes	0.000

To summarize the information found in Table 11, the list below shows the most significant factors based upon the highest sum in relation to perceived usefulness.

- The technology enables me to accomplish tasks more quickly [6]
- The technology saves me time [5]
- The technology increases my productivity [12]
- The technology improves my job performance [3]
- The technology reduces the time I spend on unproductive activities [9]

Finally, in addition to the list above, the list below shows the most significant factors based upon the highest sum related to the perceived ease of use.

- Interacting with the technology is *NOT* frustrating [17]
- I *NEVER* become confused when using the technology [15]
- Overall, I find the technology easy to use [28]
- I do *NOT* find it cumbersome to use the technology [24]
- The system is *NOT* rigid and inflexible to interact with the technology [21]

Conclusions

From a low-level standpoint, RG outperformed XYZ in nearly every category that was evaluated in the TAM. In fact, the null hypothesis was rejected in twenty-four out of the twenty-eight individual tests that were performed. With this evidence, it is not surprising that from a slightly higher-level standpoint, the null hypotheses stating that the perceived usefulness and ease of use of RG were better than or equal to XYZ were also rejected. Finally, when all of the respondent data was grouped together, the tests indicated that the RG technology was accepted from an overall standpoint.

From a high-level standpoint, the results of this investigation suggest that students and teachers would benefit if commercialized LMS environments would adopt a similar flexible rubric system. The primary features of RG that make the system unique include:

- The ability for teachers to start the grading process with or without a pre-defined rubric
- The ability for teachers to add or modify rubric dimensions during the grading process
- The ability to control whether modifications are enforced across all graded and non-graded assignments or if one-time-only modifications are only applied to special cases
- The ability to manage rubric dimensions so that one can easily delete, modify, or copy from one assignment or course section

From a development standpoint, the RG component was straightforward to implement. However, RG was developed by a seasoned software engineer who was also an instructor and had a clear vision of what was either needed or wanted. The development continued as other instructors used the LMS within their classrooms. Therefore, RG is considered to be highly customized, which can be considered an advantage or a disadvantage. The advantage is that the features that could be developed into RG are endless. However, very few practitioners have the knowledge or resources to develop or modify existing code within non-commercialized LMS environments. In general, most teachers are using commercialized software that cannot be changed. Thus, one disadvantage is simply that this type of flexible feedback system does not exist in the most popular LMS environments that are used in GK-12 and higher education.

Currently, many of the LMS environments used within education have a plethora of features. From a flexibility standpoint, this is good. However, these features are not usually developed from a usability standpoint. If features such as RG are not developed with a usability perspective, then the practitioners do not benefit because the features are often considered too cumbersome to use and, therefore, they are limited in terms of their benefit. Thus, one conclusion that can be drawn from this study is that commercial suppliers of LMS environments should consider a framework like TAM when developing features for their customers. To say that a feature

exists is not sufficient anymore. Features must be developed with the customer in mind and they must be developed in a manner that is easy for the customers to use. If these conditions are satisfied, then features like RG can be used in a variety of classes that vary regarding the forms of assessments that are used.

Overall, a vast amount of evidence supports the claim that RG is an improvement over the XYZ system in terms of its **usefulness** and **ease of use** as an online grading and feedback system. As this investigation shows, RG enhances the way in which users perform their work. From this study, users are more efficient in grading assignments while also being able to provide high-quality feedback, which is more than what XYZ currently offers at no extra cost. In fact, in terms of the ease of use, there were no individual categories in which the null hypothesis failed to be rejected. Based on these results, an empirically-validated conclusion can be made that RG is **less confusing** to use, **less frustrating** for users to interact with, **less cumbersome** for users, and is an **easy** system to use. The simplicity of RG allows you to accomplish more with less effort, primarily by staying out of your way when grading. Spending less time worrying about making simple mistakes and being able to easily correct them when they inevitably occur is invaluable. Having the right tool for the job is sometimes all that you need.

References

- 4Teachers.org. (2012, 3 7). *RubiStar Home*. Retrieved 3 7, 2012, from RubiStar: <http://rubistar.4teachers.org/>
- Abdullah, T., Mateen, A., Sattar, A. R., & Mustafa, T. (2010). Risk Analysis of Various Phases of Software. *European Journal of Scientific Research*, 40(3), 369-373.
- Andrade, H. G. (2000). What Do We Mean by Results? Using Rubrics to Promote Thinking and Learning. *Educational Leadership*, 57(5), 13-18.
- Andrade, H. G. (2005, Winter). Teaching With Rubrics: The Good, the Bad, and the Ugly. *College Teaching*, 53(1), 27-31.
- Anglin, L., Anglin, K., Schumann, P., & Kaliski, A. (2008). Improving the Efficiency and Effectiveness of Grading Through the Use of Computer-Assisted Grading Rubrics. *Decision Sciences Journal of Innovative Education*, 6(1), 51-73.
- Atkinson, D., & Lim, S. (2013). Improving assessment processes in Higher Education: Student and teacher perceptions of the effectiveness of a rubric embedded in a LMS. *Australasian Journal of Educational Technology*, 29(5), 651-666.
- Avison, D. E., & Fitzgerald, G. (2003). Where now for development methodologies? *Communications of the ACM*, 46(1), 78-82.
- Beynon-Davies, P. M. (2000, July). 'It's lots of bits of paper and ticks and post-it notes and things . . .': a case study of a rapid application development project. *Information Systems Journal*, 10(3), 195-216.

Butler, D. A. (2011). Closing the loop 21st century style: providing feedback on written assessment via MP3 recordings. *Journal of Australasian Law Teachers Association*, 4 (1&2), 99-107.

CampusComputing. (2013). A Profile of the LMS Market. *The national Survey of Computing and Information Technology*, 1-35.

Centers for Medicare & Medicaid Services. (2008). *Selecting Development Approach*. Retrieved from <http://www.cms.hhs.gov/SystemLifecycleFramework/Downloads/SelectingDevelopmentApproach.pdf>

Cheang, B., Kurnia, A., Lim, A., & Oon, W.-C. (2003). On automated grading of programming assignments in an academic institution. *Computers & Education* 41, 121-131.

Chen, P. M. (2004, May). An Automated Feedback System for Computer Organization Projects. *IEEE Transactions on Education*, 47(2), 232-240.

Cooper, B. S., & Gargan, A. (2009). Rubrics in Education Old Term, New Meanings. *Phi Delta Kappan*, 91(1), 54-55.

Cross, P. K. (1990). Teaching to Improve Learning. *Journal on Excellence in College Teaching*, 9-22.

Davis, F. D. (1989, September). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319-340.

Dornisch, M. M., & McLoughlin, A. S. (2006). Limitations of web-based rubric resources: Addressing the challenges. *Practical Assessment, Research & Evaluation*, 11, 1-8.

Goodrich, H. (1997, January). Teaching for Student Performance. *Educational Leadership*, 54(4), 14-17.

Hall, R., Butler, L., McGuire, S., McGlynn, S., Lyon, G., Reese, R., & Limbach, P. (2001). Automated, Web-Based, Second-Chance Homework. *Journal of Chemical Education*, 78(12), 1704.

He, Y., Hui, S. C., & Quan, T. T. (2009). Automatic summary assessment for intelligent tutoring systems. *Computers & Education*, 53, 890-899.

Howell, R. (2011). Exploring the Impact of Grading Rubrics on Academic Performance: Findings From a Quasi-Experimental, Pre-Post Evaluation. *Journal on Excellence in College Teaching*, 22(2), 31-49.

Klein, J., & El, L. P. (2010). Impairment of teacher efficiency during extended sessions of test correction. *European Journal of Teacher Education*, 26(3), 379-392.

Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*.

- Machado, M., & Tao, E. (2007). Blackboard vs. moodle: Comparing user experience of learning management systems. *Frontiers In Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports*, (pp. 7-12). doi:10.1109/FIE.2007.4417910
- Mackay, H., Carne, C., Beynon-Davies, P., & Tudhope, D. (2000, October). Reconfiguring the User: Using Rapid Application Development. *Social Studies of Science*, 30(5), 737-757.
- Meyers, L. S., Gamst, G. C., & Guarino, J. A. (2005). *Applied Multivariate Research: Design*. Sage Publications, Inc.
- Miser, H., & Quade, E. (1988). *Handbook of Systems Analysis: Craft Issues and Procedural Choices*. USA: Wiley.
- Necco, C. R., Gordon, C. L., & Tsai, N. W. (1987, December). Systems Analysis and Design: Current Practices. *MIS Quarterly*, 11(4), 461-476.
- Ramey, S., VandeVusse, L., & Gosline, M. (2007). Using a Written Communication Rubric to Improve Students' Writing. *International Journal of Learning*, 13(10), 67-74.
- Riddle, E., & Smith, M. (2008). Developing and Using Rubrics in Quantitative Business Courses. *The Coastal Business Journal*, 7(1), 82-95.
- Spinellis, D., Zaharias, P., & Vrechopoulos, A. (2007). Coping With Plagiarism and Grading Load: Randomized Programming Assignments and Reflective Grading. *Computer Applications in Engineering Education*, 15(2), 113-123.
- TeacherPlanet.com. (2012, 3 7). *Rubrics for Teachers*. Retrieved 3 7, 2012, from Teacher Planet: <http://www.rubrics4teachers.com/>
- TeAch-nology. (2012, 3 7). *Rubrics and Rubric Makers*. Retrieved 3 7, 2012, from teAchnology: http://www.teach-nology.com/web_tools/rubrics/
- Thompson, M., & Ahn, B. (2012). The Development of an Online Grading System for Distributed Grading in a Large First Year Project-Based Design Course. *American Society for Engineering Education*, 2012-3467.
- US Department of Justice. (2012, 3 7). *DOJ Systems Development Life Cycle Guidance Chapter 1*. Retrieved 3 7, 2012, from US Department of Justice: <http://www.justice.gov/jmd/irm/lifecycle/ch1.htm>
- Wolf, K., & Stevens, E. (2007). The Role of Rubrics in Advancing and Assessing Student Learning. *The Journal of Effective Teaching*, 7(1), 3-14.