# A Subjective Evaluation of Web-based Programming Grading Assistant: harnessing digital footprints from paper-based assessments

Yancy Vance Paredes, Po-Kai Huang, Hannah Murphy, I-Han Hsiao

Arizona State University, Tempe AZ 85281, USA
{yvmparedes,phuang24,hmurphy2,Sharon.Hsiao}@asu.edu

**Abstract.** In a blended learning environment, paper-based evaluation has still been the preferred assessment method of student's learning because of the convenience it offers. In programming learning, several automatic programming evaluation tools have been developed. However, they are less focused on grading paper-based programming problems. In this paper, we discuss a homegrown educational technology called Web-based Programming Grading Assistant (WPGA). It is a system that facilitates grading and feedback delivery of paper-based assessments. A classroom study was designed and feedback from the users were collected. This paper reports a subjective evaluation of the role of the system to their learning, the ease of using the system, and their feedback on some specific features. Results show that students generally gave positive feedback on the WPGA system. Furthermore, the inclusion of analytics and social/peer learning features were some of the recommendations solicited to improve the system.

**Keywords:** Feedback, Reflection, Programming Learning, Educational Technology.

## 1 Introduction

In a blended learning environment, paper-based evaluation has still been one of the most preferred assessment methods of student's learning because of its convenience in terms of preparation. Furthermore, it is easier to reduce the risk of academic dishonesty, which is prevalent in its online counterpart. However, it has its own drawbacks. For instance, the inconsistency among or even within graders [8]. This happens when graders evaluate large amount of test papers. Another drawback is the high turnaround time before students receive a feedback on their graded papers [1]. This traditional setup defeats the purpose of having an efficient feedback mechanism, which is essential to the learning process [7]. There have been quite several research projects in literature that underscore the importance of having timely feedback in student's learning. This raises several important questions: *Do students review their returned test papers? Do they even exert effort in understanding the mistakes they have committed? Are they only focused on the final score that they have received?* With the paper-based evaluation, it is not possible to determine whether students really do review their graded test papers.

To address the above-mentioned issues, specifically in programming learning domain, a web-based platform was developed called Web-based Programming Grading Assistant (WPGA). WPGA serves as a digital-and-physical conjunction platform, which supports traditional blended classroom conduct as well as harnesses modern digital footprints. This system is capable of capturing students' learning behaviors by tracing their learning activities. WPGA is designed such that it would stimulate the students to review and reflect on their graded assessments. Also, students will be able to access a digital copy of their graded physical test paper. It is hypothesized that student's learning will be positively impacted by their ability to review and reflect on their graded test papers online.

## 2 Feedback and Reflection on Learning

Positive feedback does not necessarily produce positive results on the growth of the students [7]. This positive feedback can be in a form of praise for accomplishing a given task. It is the "critical" and not the "confirmatory" feedback which is considered to be the most beneficial for learning [3]. Promptness in providing feedback is said to be very essential to the learning process [12]. Additionally, one of the studies also shows that slow feedback and no feedback at all had no significant difference [4]. Self-corrective feedback improves the performance in exams. The immediate feedback must be task-related. In some studies, exams where feedback is given individually (per question) are seen to be more effective than those with feedback given as a whole [12]. These results highlight the importance of having a timely feedback.

Aside from assessing what the students know, the assessments should also be able to allow the students to reflect and evaluate their performance. It enables them to understand their reasoning processes. There are two types of reflection: in action (during study and practice) and on action (after being assessed) [6]. Successful learners are believed to be aware of what they do not know. Therefore, the main focus of this work is to investigate how do students reflect on their understandings after they are assessed.

In the system developed, students will receive feedback in electronic form after their works are graded. This includes both formative and summative feedback. This will allow the students to react on the graded items and to take some notes for future reference. The objective of this work is to make it possible for the prompt delivery of feedback to students and to capture how students attend to it. Also, it seeks to understand students' reviewing and reflecting behaviors and how they impact on their learning.

## 3 Technology Support in Feedback Generation and Delivery

Several systems have already been developed to automate the generation of feedback in STEM courses. Examples of which include WEB-CAT [5] and ASSYST [11]. Pattern-matching techniques are applied to determine the correctness of the student's answers. The answers of the students are compared to the identified correct answers. Unfortunately, in programming learning, it is difficult to use this existing system because it does not consider the nature of the solution of the student. It does not capture the thought process or the ability of the student to understand the problem. Existing automated program evaluation is more focused on checking electronic source codes and not paper-based.

A few studies have attempted to address the problem. One proposed solution was to digitize the giving of feedback and provide a centralized grading interface for the graders. This is made possible by tablet grading system [2,10]. This helps in mass programming grading since default feedback can be stored and given to those solutions with similar mistakes. Furthermore, the student's identity can be anonymized which would eliminate the bias of the grader.

# 4 Design and Implementation of Web-based Programming Grading Assistant (WPGA)

A web-based system was developed to help in grading paper-based exams and in providing direct feedback online. The name of the system is Web-based Programming Grading Assistant (WPGA)[1]. It connects paper-based assessments to the cyber world and ensures instructors that they can still have paper-based exams without having to learn new tools. It has three key features: (1) digitalization of paper-based assessments, (2) augmented grading platform, and (3) reflective feedback delivery.

## 4.1 Digitalization of paper-based assessments

WPGA utilizes quick response (QR) codes to label and identify a hard copy of an exam of a student. Instructors upload their student rosters, which contain a list of student identification and names. An automatic document feeder is used to scan all the paper exams of the students. Afterwards, these images are uploaded to the online system. All the scanned exams will be stored as images and will be recognized and indexed using the QR code.

## 4.2 Augmented grading platform

After digitizing and labeling an exam, instructors can partition it into multiple sections. These sections can be assigned to different graders as WPGA allows multiple graders to grade them simultaneously. In effect, the system reduces the turnaround time of the distribution of grades. Also, graders' grading coherence will improve.

The grading interface is shown in Figure 1. The system allows multiple types of feedback to be given through its interactive buttons, which are located on the upper right corner. Every rubric default to a perfect score, which translates to a full understanding of the concept. For every click of the button, the color of the button changes: red (partial understanding) or grey (missing the concept). Also, the grade is decremented and the overall score is recalculated. Free form feedback can also be provided in the comment section. The grader can even write or mark on the question image (the

[1] http://cidsewpga.fulton.ad.asu.edu/gradingHelper/

paper exam image in the middle of interface). According to a previous study [8,9], graders prefer on typing their feedback rather than physically writing them on paper. They feel comfortable using digital grading system. Most importantly, it is possible for them to copy and paste comments given to similar mistakes since this can be a common scenario.

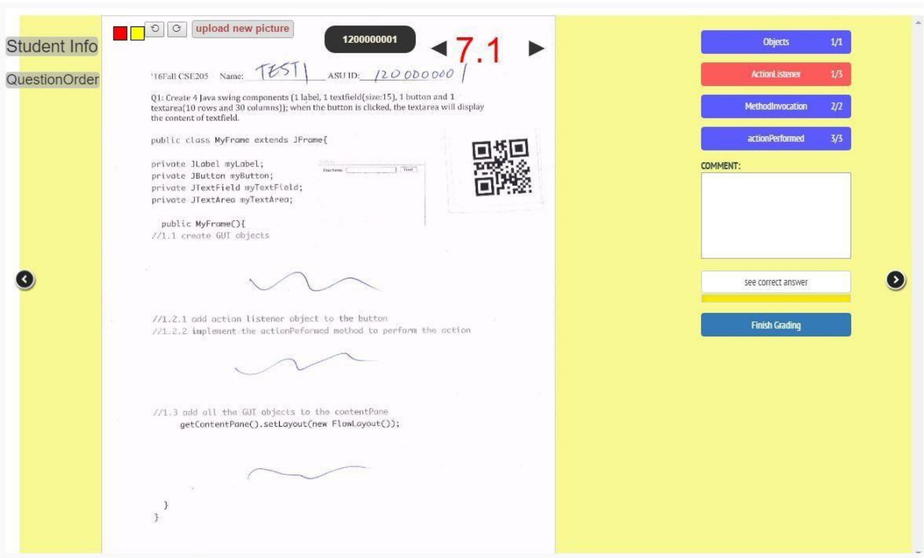


**Fig. 1.** Grading interface: grading scheme is tied to feedback buttons; free form comment text area; and correct solution details

### 4.3 Reflective feedback delivery

In the student's interface, there are two levels of view: exam level and question level. In the exam level (Figure 2), a general result of the exam is displayed. This includes the overall score and the points obtained by the student on each of the questions. In the question level, a detailed feedback is displayed. This feedback includes both summative and formative. In addition, students can take notes to reflect on the feedback given to them. There is also has a checkbox to signify whether they have already known how to solve the problem. They can also use a star bookmark to indicate the importance of the problem. These features allow them to filter questions for targeted review in the future.

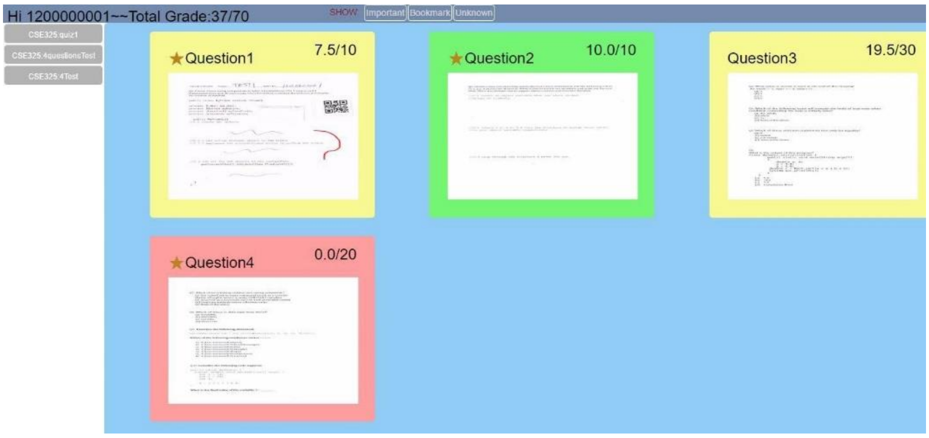


**Fig. 2.** Exam level student interface: the overall score and the points earned by the student on each of the questions

# 5 Findings

In the main conference paper [10], the reviewing and reflecting behaviors of the students using WPGA were tracked. This sequential data of activities was modeled using Hidden Markov Models (HMM). In this paper, we focused on the Human-Computer Interaction aspect of the system. To understand the general sentiment of the students about the system, they were asked to anonymously answer an online feedback form. The form is divided into five sections. The first section seeks to identify which CSE courses the student belongs to. The second section asks questions that determine the effect of using WPGA to the learning of the students. The third section contains questions that pertain to determining the usability of WPGA. The fourth section contains questions that seek to determine the awareness of the students to specific features of the system. Lastly, the fifth section solicits ideas from the students on new features to the system. A total of 11 questions were asked.

The collection of feedback was administered during the last week of the Fall semester in 2016. All the students who were using the system in any of their CSE courses were invited to fill the feedback form. They were informed that this survey will not affect their academic standing. In this manner, they can be honest about their answers. A total of 199 respondents answered the WPGA survey.

## 5.1 Role in learning

Figure 3 illustrates that 48.49% (aggregate of 36.87 and 11.62) of the students responded that WPGA was able to help them learn the class material better. On the other hand, 27.1% were undecided whether the system was able to help them or not. When asked whether they are going to use WPGA to help them in studying for an exam, 42.7% responded positively while 26.9% were uncertain or undecided. Since the system has just been recently introduced to the students, and was offered to support returned paper delivery, (we aimed to capture students' natural reviewing and reflecting behaviors, reported in the main conference paper), this might explain why less than half were using it to help them in their learning. Students were asked how they would normally review for a programming exam. The following were the popular choices: review the slideshows from lecture (78.2%), review assignments (68%), and create a study guide (45.7%). Based on this result, majority of the students identify their study preparation process involves with a range of review activities. It led us to believe that capturing and modeling the corresponding reviewing behaviors can predict learning outcome.

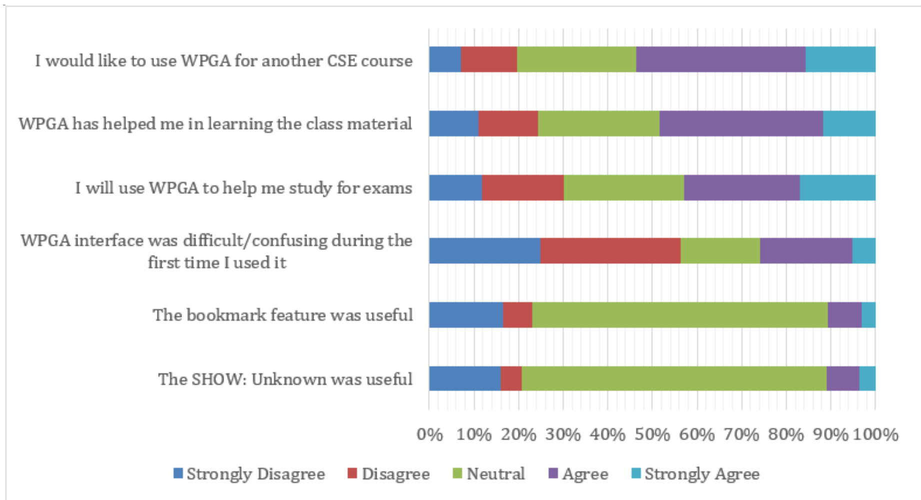


**Fig. 3.** Some of the survey questionnaire responses.

### 5.2 Ease of using the system

In terms of using the WPGA system, 56% of the students found it easy to use. Majority of the students felt comfortable using the system after taking 1-2 quizzes. A quick survey of the awareness of functionalities revealed that majority of the users already knew what the color-coding means for a given problem. Unfortunately, only few students know how to use the advanced features, which includes filtering important/bookmarked/unknown questions, bookmarking a question, or making a note. The possible explanations could be because they are using a new system and it was introduced only in the latter part of the semester. Also, during that time, there was no user manual that could have helped them familiarize the system. Another possible explanation could be students are just mainly focused on examining their exam scores and they do not bother exploring other features offered to them by the system. With WPGA, it is easier for them to access their exam scores virtually anywhere. However, this raises an issue on whether the other features of WPGA are indeed visible to the users.

### 5.3 WPGA specific interface questions

The students were asked to provide some feedback on some specific features of WPGA. When asked about the usefulness of the bookmark feature, majority (65.8%) had no opinion about it. This could be attributed to the fact that only few users know it or are aware on how to use it. The same goes for the “SHOW: Unknown” feature. Majority (68.2%) had no opinion about it because many them do not know that the feature exists.

The following were the features suggested to be included to improve the system: (1) make available to the students the analytics showing the overall performance of everyone in the class; and (2) include social and peer learning features which will allow them to communicate not only with the professor but with other students as well.

Students would like the ability to communicate with the professor or TA through WPGA because they wanted to ask for help for them to understand a specific question (63.4%). The next reason would be they wanted WPGA to be able to help suggest on what content to focus on (51.2%), which suggests more advanced predictive models (i.e. recommenders, intelligent tutors). Lastly, they wanted a facility on how to rebut about points assigned on a specific question on an exam (43%). Overall, based on the feedback provided by the students, 53.6% would like to use WPGA for another CSE course.

## 6 Conclusions and Future Work

We designed and deployed a system called Web-based Programming Grading Assignment (WPGA), which is a new educational technology that facilitates grading and feedback delivery of paper-based assessments. The system is designed to bridge the learning analytics from physical papers to digital electronic formats. From the classroom deployment and students’ survey results, we learned that most of students use WPGA only to view test results and questions but not use advanced functions like bookmark and note taking that could have helped them in reviewing. In addition, the system was just deployed during the latter part of the Fall semester in 2016 which limited its usage. Despite this, students generally gave positive feedback on the WPGA system. They even wanted to use it in their other CSE courses.

Using the feedback provided by the students, another version of WPGA will be developed to help improve the interaction of the users and the user interface. Through this, the students would become more aware of the existence of several important features. Furthermore, new functionalities, such as giving personalized feedback and recommendations to students based on their mistakes, will be explored to increase the impact of WPGA on learning.

## References


1. Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How Learning Works: Seven Research-Based Principles for Smart Teaching*. John Wiley & Sons.
2. Bloomfield, A., & Groves, J. F. (2008). A tablet-based paper exam grading system. *ACM SIGCSE Bulletin*, *40*(3), 83.
3. Cutumisu, M., & Schwartz, D. L. (2016). Choosing versus receiving feedback: The impact of feedback valence on learning in an assessment game. In *The 9th International Conference on Educational Data Mining*. Raleigh, NC.
4. Dihoff, R. E., Brosvic, G. M., Epstein, M. L., & Cook, M. J. (2004). Provision of feedback during preparation for academic testing: Learning is enhanced by immediate but not delayed feedback. *The Psychological Record*, *54*(2), 207.
5. Edwards, S. H., & Perez-Quinones, M. A. (2008). Web-CAT: automatically grading programming assignments. In *Proceedings of the 13th annual conference on Innovation and technology in computer science education* (Vol. 40, pp. 328–328). ACM.
6. Ertmer, P. A., & Newby, T. J. (1996). The expert learner: Strategic, self-regulated, and reflective. *Instructional Science*, *24*(1), 1–24.
7. Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112.
8. Hsiao, I.-H. (2016). Mobile Grading Paper-Based Programming Exams: Automatic Semantic Partial Credit Assignment Approach. In *Lecture Notes in Computer Science* (pp. 110–123).
9. Hsiao, I.-H., Govindarajan, S. K. P., & Lin, Y.-L. (2016). Semantic visual analytics for today's programming courses. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*. https://doi.org/10.1145/2883851.2883915
10. I.-Han Hsiao, Po-Kai Huang, and Hannah Murphy, "Uncovering reviewing and reflecting behaviors from paper-based formal assessment," in *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 2017, pp. 319-328.
11. Jackson, D., & Usher, M. (1997). Grading student programs using ASSYST. *ACM SIGCSE Bulletin*, *29*(1), 335–339.
12. Kulkarni, C. E., Bernstein, M. S., & Klemmer, S. R. (2015). PeerStudio. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale - L@S '15*. https://doi.org/10.1145/2724660.2724670