Grading and Feedback Tool for Online Document Submission
(Technical Project)

Analysis of Online Educational Software Tools used for Grading Student Work
(STS Project)

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Philip Hart

December 9, 2022

Technical Team Members:
Ketian Tu
Anubhav Acharya

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS

MC Forelle, Department of Engineering and Society

Panagiotis Apostolellis, Department of Computer Science

Introduction

A global paradigm shift occurred within education in the midst of the COVID-19 pandemic: instructors and students had to adapt to a technological infrastructure which supports remote learning in a vast variety of age groups. A striking example is a digital learning platform that was evaluated in primary and secondary mathematics classes for 12 weeks. This study found that while instructors use software tools primarily for lesson planning and grading, students use software tools primarily to complete their homework (Christopoulos & Springers 2021). In higher education, assignments grow more complex in an effort to model real-world problems and scenarios. Instructors, particularly with large classes, must have a system in place which tracks student learning while also allowing them to grade with consistency and transparency. A group of graders, such as teaching assistants, may also assist in the course requiring collaborative grading support for the chosen software to be exceptionally useful.

Grading software is dominating the evaluation systems for many higher education institutions, notably in engineering fields where assignments involve design ideation and implementation. A generalized term for these assignments is Project-Based Learning (PBL) (Kokotsaki, Menzies, Wiggins 2016). Throughout these assignments, students tackle problems and generate a deliverable that is then evaluated by the instructor. That deliverable may take many different forms depending on the teaching style, course content, and computer infrastructure available. An argument surrounding PBL is focused on how much support students need throughout learning based on problem solving (Webb & Moallem 2016), a crucial component of PBL. However, research shows that feedback containing useful resources are particularly useful for PBL because the student learning is dependent on the concepts engaged with throughout the development of a student's deliverable (Barron, Schwartz, Vye, Moore, Petrosino, Zech, & Bransford 1998). Thus, for instructors that integrate PBL assignments, grading must be executed to provide rich feedback in a timely manner. The software used by the course may be paramount to the success of these hands-on assignments. This is just one example of an assignment type that may be difficult to grade given current

software tools. The aim of this paper is to determine the benefits of grading software tools for evaluating non-traditional deliverables and to examine the tool's ability to provide quality, consistent feedback for student work. In other words, how much do instructors that assign open-ended design project really benefit from these tools while grading? Grading software is capable of transforming the educational experience for students and the effects of these systems are hypothesized to contribute to better learning and internal reflection.

Methodology

To solve the problem of how non-traditional formats of student work may best be graded, a literature review of non-traditional assignments and grading software used in higher education was conducted. An important feature of non-traditional assignments is that student solutions are not always arrived upon in the same way. Therefore, there must be a personalized aspect of grading in order to engage the student or group with the courses feedback process and use reflective practices to guide their chosen implementation (Almulla 2020). Thus, a review of feedback in higher education and its effectiveness in multiple contexts was examined. A software tool was developed that synthesizes the grading and feedback process and was analyzed for its usefulness for grading. This will provide some insight into the importance of this tool and may guide the direction of future grading tools that maintain a balance between grading and personalized feedback. Using the idea of 'configuring a user' (Woolgar 1990), the development of several tools was put into perspective. The intentions behind many of these tools are evaluated including the tool developed to solve the problem of grading non-traditional deliverables.

The State of Grading

Within higher education, assignments are intended to produce the following results: introduce and cement conceptual knowledge from a course, provide students with hands-on experience relating to conceptual knowledge gained in the course, or a combination of both. For certain fields, Project-Based Learning (PBL) is often used as a means of integrating theory and practice by means of problem solving

(Bergh, Mortelmans, Spooren, Petegem, Gijbels, & Vanthournout 2006). Information technology has been increasingly applied to PBL as a means to support different learning styles and enhance student learning outcomes (Guthrie 2010). Therefore, technology has completely transformed the types of assignments available for students. Instructors have thus began offering assignments that use the features of the internet for sharing and communicating what they learn through online deliverables (Lynch, Sage, Hitchcock, & Sage 2021). Depending on the course, these online deliverables may take many different forms. For example, engineering courses may use open-ended, client-driven, team-based problems commonly referred to as Model-Eliciting Activities (MEAs) (Yildirim, Shuman, & Besterfield-Sacre 2010). The problem for instructors is that the outcomes of MEAs can take various forms, and grading tools support only a limited number of submission types. Gradescope, for example only allows the submission of PDFs or images. In engineering courses that incorporate a long-term student design project, non-traditional online deliverables are a common outcome (Beneroso & Robinson 2022). Grading tools must therefore be flexible to be effective at evaluating these types of deliverables.

An integral component to evaluating PBL or MEAs is the feedback that students receive throughout project development. This feedback must be timely and consistent to assist students in their learning process (Hattie & Timperley 2007). Since these non-traditional assignments are worked on by students over a long period of time, splitting up the assignment into different deadlines ensures students are getting their work done and creates space for instructor feedback on student work. This feedback should aim to reduce the gap in understanding between the student's current understanding and the instructor's desired goal (Shute 2008). This is crucial for learning and grading transparency, particularly when conducted online where there are so many different concepts and processes to explore. The feedback for these assignments must therefore be personalized and easily accessible.

Various grading tools are currently fully integrated into many higher education universities and have been evaluated to determine the benefits of such tools for student learning. One of the most prevalent

software used in compulsory and tertiary STEM education is Gradescope (Yen, Karayev, & Wang 2020). Students upload a document to be graded and a specific rubric for that assignment can be used to grade the uploaded file. The benefits to instructors using Gradescope include speed, consistency, and flexibility while grading (Singh, Karayev, Gutowski, & Abbeel 2017). For large courses, a grading system with these benefits is valuable to save time, establish trust between student and instructor, and change the grading criteria as seen fit. On the other end of grading, students receive their grades in a timely manner enabling deeper reflection since the work and feedback is returned closer to the assignment's due date. The use of a rubric allows for the students to receive a more transparent view of the concepts being graded and gives them a better picture of what the full requirements of the assignment are. However, Gradescope was developed with a focus on grading paper-based work. Thus, the limitations of the tool are obvious when applied to long-term projects which involve multiple components. Further, Gradescope submissions can only be graded in 'passes', meaning that groups of graders in large courses cannot grade the same submission at the same time.

Another tool, iRubric, is a web-based rubric development, assessment, and sharing software, often integrated with a Learning Management System (LMS). An LMS, such as Blackboard, Moodle, or Canvas, provides students with access to all of a course's resources within one easily navigable page. iRubric can be incorporated to grade grade work that is uploaded to these sites and is useful generating customized rubrics for assignments. Similar to Gradescope, iRubric increases grading transparency and consistency in the assessment of student work. Further, iRubric streamlines the feedback process by allowing instructors to click on a rubric criterion while grading an assignment which prompts them to enter specific comments within a feedback box (Myers, Peterson, Mathews, Sanchez 2018). iRubric is excellent for developing rubrics for assignments, but is a generalized software, not meant for directly grading student work and providing feedback.

While many other grading tools exist, the aforementioned tools satisfy the effectiveness of modern grading software. Gradescope is excellent for grading certain forms of assignments and iRubric can be adapted to use for various assignment contexts. To continue the examination of software tools used for evaluating student work, I will now pivot toward software that lacks grading functionality but excels in providing feedback for student work.

Hypothes.is is an annotation tool for marking web pages and PDF files with highlights and comments. When used in a synchronous learning environment, Hypothes.is was found to promote communication and peer review (Grossu 2021), both of which are crucial to project-based learning. Digication, an online commenting tool, allows students to submit a screenshot of a web page, but commenting can only occur on the live website which lacks usefulness. Diigo, a social bookmarking tool, allows user to attach digital notes to web pages, and is useful for organizing content. Many of these tools have been studied in the context of student learning (Dennen, Cates, & Bagdy 2017, Lethinger & Haller 2007). However, an examination of annotation technology has shown to benefit grader's when generating feedback for students (Wolfe 2002).

The purpose of this review has been to introduce some non-traditional means of assessing student knowledge, as well as examine the grading tools available for instructors in higher education. Currently, grading software supports the evaluation of paper-based assignments best but lacks support for grading student work hosted somewhere on the web. Further, prominent grading software such as Gradescope does not support collaborative grading, a component sometimes necessary with courses that have multiple graders. Annotation technology is a strong alternative for evaluating non-traditional student work, but conversely lacks grading support. Additionally, a problem arises with how to organize a grading and feedback system for decentralized web-based deliverables, as a project could be organized into groups of multiple students or spread across multiple webpages.

Grading Outcomes

With the introduction of technology in education, the types of assignments offered to students has expanded to include non-traditional deliverables. Therefore, the tools available to graders should aim to support the grading of these deliverables. However, each tool is developed for a particular set of purposes guided by problems faced by their future users.

The review of technologies used in education has shown several paradigms present within the development such grading tools. Gradescope is well suited for grading an individual's problem-solving abilities and decreases grading time while ensuring equity and consistency for students throughout the grading process. iRubric's usefulness in grading is also due to its ability to ensure equity and consistency for students by standardizing the objectives of an assignment to a rubric. Throughout the development of these tools, a focus was placed on these requirements. A way of looking at this development process is through the lens of 'configuring the user' (Woolgar 1990). The creators of these tools were invested in how users in the future could use these tools and thus implemented requirements based on how they believed future users would use the tool. Based on the review, developers of Gradescope aimed to provide instructor with an efficient, consistent, and transparent grading process. These requirements were implemented, but with the nuance of limiting what may be submitted by a student so that the tool's features are guaranteed to work. Conversely, iRubric was developed for the purpose of generating rubrics. Since iRubric may be integrated with an LMS or used completely detached from where the assignment was submitted, it is more flexible in that it can be used for any type of assignment. If iRubric is used separately from an LMS, consistency and transparency are conserved while grading efficiency is lost. This reiterates the problem current faced in a new light: 'configuring the user' results in unimplemented requirements that may impede the grading process of non-traditional assignments.

Through findings in literature, annotation tools prove to be one step closer towards supporting the grading of non-traditional assignments hosted on the web. This is due to their ability to manage, mark, and comment on web pages. These tools were created with this purpose in mind. Using the idea of 'configuring

the user', annotation tools have not branched to supporting the evaluation of web pages and has instead focused on supporting self-learning and group-learning. A further iteration of an annotation tool could include features that support the grading process by associating the grade annotations created with a specific assignment. This grading process can be abstracted to the evaluation of deliverables in phases, for example, each phase (or 'assignment') focuses on specific learning goals for a semester-long project.

A first-of-its-kind grading tool, e^{2}logos, has been developed to integrate the necessary design requirements explored in this paper and evaluated through a usability study. The tool is based on Hypothes.is, a chrome extension that allows users to mark websites. Design requirements for the tool include within-context feedback and grading, personalized adjustment of score and feedback, collaborative grading, and general feedback and regrade requests. A rubric tab was added to the extension as well as points associated with annotations. The different user roles, 'Instructor', 'Grader', and 'Student' determines what annotations are shown at different points of the grading process. For example, the instructor can release the assignment from the e^{2}logos homepage, which allows students to see the grade and feedback for the assignment. Further features include the auto-selection of an assignment upon visiting a graded webpage, links on rubric deduction to navigate to the page that item was graded on, and the option to hide grader comments.

The tool was used in the 2022 Fall *Human-Computer Interactions in Software Development* course at the University of Virginia. Graders and students were invited to groups, a mechanic already implemented in Hypothes.is, in which they could interact with the assignments set up by the instructor. 506 rubric items were collected by the end of the semester. Then, a usability study was conducted to determine indicators of interest within the grading process. In comparison to Gradescope, findings found that while grading open-ended learning, the developed tool was perceived as more efficient and dependable. A shortcoming of this study is that data was only gathered from two sections in the same course and the functionality of the tool changed in small doses throughout the semester.

e^{2}logos ‘configures the user’ to interact with and grade on websites, a vastly different view compared to how other companies imagined their future users to use their grading tool. Within the actual functionality of the chrome extension, graders are configured to evaluate textual information, while using images, animations, or other deliverables as reference for student work progress. Students are thus able to creatively tack open-ended design problems and present their work in any viewable format they choose. The lack of restriction in submission type provides opportunity for the exploration of different tools available on the internet, assisting students in developing IT skills throughout the learning process. Such skills are valuable if not necessary in many fields.

Conclusion

The grading process of an instructor may guide students to higher learning objectives and is an important motivator for student success. Through a review of some unfamiliar types of assignment and different types of tools available for grading, paradigms in the development of these tools present challenges for grading such assignments. While these tools are excellent to use for their respective intended purposes, the use of the internet in a classroom presents a problem for how to efficiently grade deliverables hosted on the web. Each technology ‘configures a user’ a certain way, but a technological grading approach has not been applied to non-traditional assignments such as PBL or MEAs. An instructor that uses these tools for non-traditional assignments will not benefit greatly in terms of efficiency, while they will benefit only if using traditional assignments. Within higher education, students in design and STEM fields lack real experience and a way to provide them with a simulated experience is through non-traditional assignments. So far, annotation technology shows the most promise in supporting the evaluation of non-traditional online deliverables.

Through the development of an online annotation-based grading tool, there has been some data collected on how usable this type of tool is for the grading process. Findings concluded that it was more efficient and dependable than the leading software competitor. In the perspective of ‘configuring the user’,

students may develop their deliverable on various types of websites that are markable by the tool. This is indicative of a shift from an isolated submission site to numerous different types of markable websites. Each approach has their respective benefits for instructors. Design requirements extracted throughout the iterative design process of the tool may help inform the design of future grading tools and also provide solutions to some of the shortcomings of existing tools. Given the current benefits of tools for expediting grading, increasing consistency, providing transparency, and marking websites, a combination of these components may be a way to tackle the problem on evaluating online reports. Further, these components may most importantly save time for instructors that provide assignments emulating real life problems.

With the advances made in network technology and software, more instructors may be inclined to take advantage of online resources and challenge their students to learn through IT infrastructure. The benefits of current grading tools are vast, but not well fitted for long-term projects or IT deliverables that may take many different directions. Annotation technology shows a possible route for handling the evaluation of these projects and deliverables that incorporates the current paradigms of grading. While creating a new grading tool may lead to a better solution, teaching styles, experience, and IT skills contribute to what users may require out of the tool. Predicting such requirements and implementing them without sacrificing learnability and usability of the software is essential for such a tool to thrive within higher education.

References

Christopoulos, A., & Sprangers, P. (2021). Integration of educational technology during the Covid-19 pandemic: An analysis of teacher and student receptions. *Cogent Education*, 8(1), 1964690. https://doi.org/10.1080/2331186x.2021.1964690

Kokotsaki, D., Menzies, V., & Wiggins, A. (2016). Project-based learning: A review of the literature. *Improving Schools*, 19(3), 267–277. https://doi.org/10.1177/1365480216659733

Webb, A., & Moallem, M. (2016). Feedback and Feed-Forward for Promoting Problem-Based Learning in Online Learning Environments. *Malaysian Journal of Learning and Instruction*, Vol. 13, No. 2 Dec 2016, 1–41. https://doi.org/10.32890/mjli2016.13.2.1

Barron, B., Schwartz, D., Vye, N., Moore, A., Petrosino, A., Zech, L., & Bransford, J. (1998). Doing With Understanding: Lessons From Research on Problem and Project-Based Learning. *Journal of the Learning Sciences*, 7(3), 271–311. https://doi.org/10.1207/s15327809jls0703&4_2

Almulla, M. A. (2020). The Effectiveness of the Project-Based Learning (PBL) Approach as a Way to Engage Students in Learning. *SAGE Open*, 10(3), 215824402093870. https://doi.org/10.1177/2158244020938702

Beneroso, D., & Robinson, J. (2022). Online project-based learning in engineering design: Supporting the acquisition of design skills. *Education for Chemical Engineers*, 38, 38-47.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of educational research*, 77(1), 81-112.

Shute, V. J. (2008). Focus on Formative Feedback. *Review of educational research*, 78(1), 153-189.

Michael Yen, Sergey Karayev, and Eric Wang. (2020). Analysis of Grading Times of Short Answer Questions. *In Proceedings of the Seventh ACM Conference on Learning @ Scale (L@S '20)*. Association for Computing Machinery, New York, NY, USA, 365–368. https://doi.org/10.1145/3386527.3406748

Lynch, M., Sage, T., Hitchcock, L. I., & Sage, M. (2021). A heutagogical approach for the assessment of Internet Communication Technology (ICT) assignments in higher education. *International Journal of Educational Technology in Higher Education*, 18(1). https://doi.org/10.1186/s41239-021-00290-x

Van den Bergh, V., Mortelmans, D., Spooren, P., Van Petegem, P., Gijbels, D., & Vanthournout, G. (2006). NEW ASSESSMENT MODES WITHIN PROJECT-BASED EDUCATION - THE STAKEHOLDERS. *Studies in Educational Evaluation*, 32(4), 345–368. https://doi.org/10.1016/j.stueduc.2006.10.005

Guthrie, Cameron. (2010). Towards Greater Learner Control: Web Supported Project-Based Learning. *Journal of Information Systems Education*. 21. 121-130.

Yildirim, Tuba & Shuman, Larry & Besterfield-Sacre, M.. (2010). Model-Eliciting Activities: Assessing engineering student problem solving and skill integration processes. *International Journal of Engineering Education*. 26. 831-845.

Singh, Arjun & Karayev, Sergey & Gutowski, Kevin & Abbeel, Pieter. (2017). Gradescope: A Fast, Flexible, and Fair System for Scalable Assessment of Handwritten Work. 81-88. 10.1145/3051457.3051466.

Myers, D., Peterson, A., Matthews, A., & Sanchez, M. (2018). One team's journey with iRubrics. *Current Issues in Emerging eLearning*, 4(1), 248–261.

Grossu, D. (2021). Using the Hypothesis Tool in a Synchronous Learning Environment.

Dennen, V. P., Cates, M. L., & Bagdy, L. M. (2017). Using Diigo to engage learners in course readings: Learning design and formative evaluation. *International Journal for Educational Media and Technology*, 11(2).

Morris, M. R., Brush, A. B., & Meyers, B. R. (2007, October). Reading revisited: Evaluating the usability of digital display surfaces for active reading tasks. *In Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'07)* (pp. 79-86). IEEE.

Wolfe, J. (2002). Annotation technologies: A software and research review. *Computers and Composition*, 19(4), 471–497. https://doi.org/10.1016/s8755-4615(02)00144-5

Woolgar, S. (1990). Configuring the User: The Case of Usability Trials. *The Sociological Review*, 38(1_suppl), 58–99. https://doi.org/10.1111/j.1467-954x.1990.tb03349.x