Work-in-Progress: An Investigation into the Benefits of Tagged Assessment Scores

Prof. Robert M Nickel, Bucknell University

Robert M. Nickel received a Dipl.-Ing. degree in electrical engineering from the RWTH Aachen, Germany, in 1994, and a Ph.D. in electrical engineering from the University of Michigan in 2001. During the 2001/2002 academic year he was an adjunct instructor at the Department of Electrical Engineering and Computer Science at the University of Michigan. From 2002 until 2007 he was a faculty member at Penn State University. Since the fall of 2007 he is a faculty member at the Electrical and Computer Engineering Department of Bucknell University, Lewisburg, Pennsylvania. While being a faculty at Bucknell he spent the 2010/2011 academic year as a Marie Curie Incoming International Fellow at the Institute of Communication Acoustics, Ruhr-Universität Bochum, Germany. During the 2022/2023 academic year he was on sabbatical as a Guest Researcher at the Electronic Systems of Medical Engineering Department, Technische Universität Berlin, Germany. Prof. Nickel's research interests include signal processing, artificial intelligence, machine learning, and engineering education.

Dr. Alan Cheville, Bucknell University

Alan Cheville studied optoelectronics and ultrafast optics at Rice University, followed by 14 years as a faculty member at Oklahoma State University working on terahertz frequencies and engineering education. While at Oklahoma State, he developed courses

Dr. Stu Thompson, Bucknell University

Stu is an associate professor and chair of the department of Electrical and Computer Engineering at Bucknell University, in Lewisburg, PA. While his teaching responsibilities typically include digital design, computer-related electives, and senior design, his focus

Dr. Rebecca Thomas, Bucknell University

Rebecca Thomas is the inaugural director for the Pathways Program at Bucknell University, where she oversees the rollout of Bucknell's ePortfolio initiative. She is also an Associate Teaching Professor in the Department of Electrical and Computer Engineering. She holds a B.S. and M.Eng. in Electrical Engineering from the University of Louisville and a Ph.D. in Electrical Engineering from North Carolina State University.

Dr. Stewart Thomas, Bucknell University

Stewart J. Thomas received the B.S. and M.Eng. in Electrical Engineering from the University of Louisville in Louisville, Kentucky in 2006 and 2008, respectively, and the Ph.D. in Electrical and Computer Engineering from Duke University in Durham, North Carolina in 2013. He has served on the organizing committee for the IEEE International Conference on RFID series since 2014, serving as the Executive Chair in 2022, with research interests in areas of low-power backscatter communications systems and IoT devices. He is also interested in capabilities-based frameworks for supporting engineering education. He is currently an Assistant Professor at Bucknell University in the Electrical and Computer Engineering Department, Lewisburg, PA USA.

Jenny Tilsen, Bucknell University

Dr. Jenny Tilsen has a background in Science and Technology Studies and STEM Education. She is also the creator or STEMtelling, a pedagogical storytelling tool.

Sarah Appelhans, Lafayette College

Sarah Appelhans is an Assistant Professor of Engineering Studies at Lafayette College. She earned her PhD in Cultural Anthropology at the University at Albany (SUNY) where she conducted research on

the cultural factors that contribute to inequalities in engineering. As a postdoc at Bucknell University, she was the resident ethnographer in the Electrical and Computer Engineering Department, exploring applications of Amartya Sen's capabilities approach in engineering education. Her current book project, On the Bleeding Edge: Gender, Immigration and Precarity in Semiconductor Engineering, investigates the intersections of gender, race/ethnicity, and immigration status among semiconductor engineers.

Philip Asare, University of Toronto

WIP: An Investigation into the Benefits of Tagged Assessment Scores

Introduction and motivation

This work-in-progress paper discusses the development of a tagged assessment scoring system and its potential benefits. Our motivation stems from the fact that course grades, while widely used, have several limitations when it comes to measuring student learning, progress, and potential [1, 10]. Grades fail to recognize the nuanced and individualized strength and weaknesses of a student [2], they lack precision [3], are prone to exhibit subjectivity and inconsistencies [4], overemphasize performance vs. learning [5], have cultural and contextual biases [6], and provide very limited feedback to the learner, the teacher, as well as any potential future employer. In addition, they frequently do not adequately assess skills such as creativity, collaboration, emotional intelligence, or resilience for example.

The reasons why grades are still a predominant feature of higher education are many-fold. For example, they offer a means for standardization and comparison [7] and are associated with an air of "objectivity and fairness," despite the fact that they are frequently not [8]. One of the barriers in moving away from a course grade as a single number that "measures" the learning/performance of a student is that the software and data systems that colleges and universities use are deeply rooted in the course grade paradigm. The affordances that these systems offer, in turn, dictate a great deal of how student assessment is structured, evaluated, and reported. It stands to reason that, if one broadens the affordances that these systems provide, one may be able to better recognize the nuanced and individualized areas of improvement for students and thereby provide better feedback to the learner as well as the teacher.

To that end, we have developed software that provides features that are not readily available with most learning management systems. We refer to this software as the *RED System* since it arose in conjunction with an NSF RED grant awarded to our department. The key aspect of the tool is that every assessment score can be associated with a number of category labels called *tags* [9]. In addition, the software provides various tools to easily represent configurable subsets of data in graphical form. The combination of tagged data with graphical representations is intended to allow faculty and students to review performance data in significantly more nuanced form than with traditional LMS systems. By offering these features we want to assess how they influence the grade sharing and communication practices for both students and faculty. To accomplish this, we surveyed faculty and student responses to the testing of the new software in their courses. The results of this preliminary study will be used to further develop the new software and to create new research instruments and a research design for a more robust qualitative study.

Software description

As mentioned before, the RED System allows faculty to share course scores, assignment feedback, and grades with students and advisors. In that sense the software is similar to a Learning Management System (LMS), but with five important differences:

1. The software allows for, both, sharing of numeric scores as well as textual feedback.
2. All score and/or feedback items can be labelled to belong to one or multiple *tags*.
3. Numeric scores can not only be reviewed in table form but also in a variety of graphical representations.

4. The data chosen for a tabular or graphical representation is selectable via a *tag-logic*, which allows for a fine grained review of virtually arbitrary slices of the data.
5. The internal data structure of the software supports a longitudinal analysis of the data across multiple courses, semesters, and ultimately the entire curriculum.

The software is still revised, extended, and updated in an ongoing, iterative fashion based on discussions with the entire research team. A few key aspects of the platform are discussed in detail in the following subsections. A comprehensive description of all aspects of the RED System, however, has to be omitted due to space limitations.

	Lab					
Name	**LB#01Rep**	**LB#01Out**	**LB#01Tech**	**LB#01Back**	**#TAGS**	**#TAGS**
Jon Hamm	19.5	40.0	100.0	35.0	**EE**	**Team 1**
Elisabeth Moss	18.0	39.0	97.5	36.0	**CpE**	**Team 1**
Vincent Kartheiser	18.0	40.0	100.0	32.0	**CpE**	**Team 2**
January Jones	19.0	40.0	100.0	40.0	**EE**	**Team 2**
Christina Hendricks	0.0	37.0	92.5	0.0	**EE**	**Team 1**
Bryan Batt	19.5	40.0	100.0	36.0	**CpE**	**Team 1**
Mason Vale Cotton	19.0	40.0	100.0	37.0	**EE**	**Team 2**
Matthew Weiner	19.5	40.0	100.0	37.0	**CpE**	**Team 1**
Scott Hornbacher	20.0	40.0	100.0	37.0	**EE**	**Team 2**
Andre Jacquemetton	18.0	37.0	92.5	37.0	**EE**	**Team 2**
Maria Jacquemetton	20.0	34.0	85.0	39.5	**CpE**	**Team 2**
Max Score	20.0	40.0	100.0	40.0		
#TAGS	**Lab**	**Lab**	**Lab**	**Lab**		
#TAGS	**Raw**	**Raw**	**Normalized**	**Raw**		
#TAGS	**Writing**	**Writing**	**Experiment**	**Info-Lit**		
#TAGS	**Teamwork**	**Ethics**	**Circuits**	**Writing**		
#DATES	10/6/25	10/6/25	10/6/25	10/6/25		

Figure 1: An example for the structure of a spreadsheet containing numerical assessment scores, tags, and assignment dates. Tags can be given to assignments as well as individual students. The listed names are *not* the names of actual students.

Score/feedback and tag management

Student scores, feedback text, and tags etc. are provided to the system via a spreadsheet. An example with an excerpt from such a spreadsheet is shown in Figure 1. The system permits the upload of an Excel file *or* it can read the information from a Google Sheet. If one uses a Google Sheet as a master spreadsheet then the upload can be accomplished with a single click. Any type of score calculation that is supported with spreadsheets is implicitly supported by the system as well since the data that is stored with the software is just the result of a calculation and not the corresponding formula. This feature makes the software a lot more flexible than the limited calculation modes supported by traditional LMS systems.

In addition to scores and feedback text, every assignment and every student can be assigned a *tag* as shown in Figure 1. These tags, such as “Lab” or “Homework”, “Teamwork”, “Information-Literacy”, “Writing”, “Team 1” or “Team 2”, etc. can be used to assign one or multiple categories to each respective score. This way it becomes possible to assign a different score to each component of an assignment and, later, link data of similar components across different assignments, course sections, semesters, and even across an entire curriculum. Furthermore, the

various tags can be used via a *tag-logic* to create almost arbitrary slices across the data for visualisation. The tag-logic and the visualizations are discussed in the next section.

Automated visualizations and the tag-logic
In addition to viewing scores and feedback text in tabular form it is also possible to view score data in the form of histograms and timelines. Which data items are selected to be included in the display can be chosen with a tag-logic. The tag-logic allows the specification of a simple binary yes/no result for each data item based on a logic expression that involves the operations *and*, *or*, or *not* in conjunction with tags. As such it is possible to create nearly arbitrary slices of the data. One may, for example, only look at lab assignments that involve a writing component, or consider only all of those assignments from student team 1 that have an information-literacy tag. The tag-logic thereby provides an easy way to peruse, browse, and view the data from a multitude of perspectives and angles. Faculty have the ability to look at data from individual students, as well as the ability to look at an aggregation of data from all students or a group of students.

Research methodology
In this preliminary study, we utilized survey research methods [11] to capture initial perceptions, attitudes, and experiences of faculty and students about grades, existing grading methods and their perceptions of the new software after using it throughout a semester-long course. To gather this information, faculty and students were surveyed in a pre survey and post survey to describe their thoughts on existing grading methods and the new software. Faculty and students were surveyed separately as they maintain different roles that influence their perspectives in understanding the impact of both grades and the new software on teaching and learning. Due to the exploratory nature of this work-in-progress study, the study is not representative. Instead, the survey is purposeful by including only a small selection of faculty and students.

Data collection and analysis
This study was conducted in an undergraduate electrical and computer engineering department at a small liberal arts university in the east coast region of the United States. All of the faculty in the department (which is the same department that the corresponding authors on this paper are affiliated with at the time of this writing), were invited to participate. Of the 12 faculty in the department, three consented to participate by completing the faculty pre and post survey, and 30 students participated in one class by completing the pre and post survey. Most of the survey questions were multiple choice or 5-point Likert scale questions. Other questions were designed to be more open-ended due to the small sample size and to gain as much detail as possible from the participant's backgrounds and experiences. The pre surveys were taken at the beginning of the Fall semester, prior to using the software in the courses. Post surveys were taken about half way through the Spring semester. At the time of the post survey, both students and faculty had been using the RED System for roughly 1.5 semesters of continual use.

Due to the page limitation imposed on WIP papers we are, unfortunately, unable to include a complete list of the questions used in our pre and post surveys. The faculty and student pre surveys had 6 questions each. Faculty and students were asked about existing grading practices and their challenges outside of the RED software. The faculty post survey had 9 questions, which asked primarily about the impact of the RED software on the faculty’s workflow, including

observed benefits and challenges, and the perceived impact on learning and grading. The student post survey contained six questions. Five of these asked students to rate easiness of access, faculty/student communications, rapidity of feedback, detail of feedback (w.r.t. tags), and clarity of feedback on a Likert scale. The last question was an open ended question on how to potentially improve the system.

To analyze our data, we developed themes from both student and faculty survey responses that were generated primarily from the open ended responses using content analysis [12]. We conducted line by line coding to identify these themes to support us in answering our research questions and to provide a direction for our further studies. Due to the brevity of this WIP, we report on only two themes from each of the faculty and student open ended responses. In addition to these themes we report on a few key aspects of student responses to the Likert scale questions to explain perceptions and experiences.

Results, discussion, and limitations

Results from the faculty and student pre survey provided us with an understanding of what current grading technologies are used by the faculty in our sample, as well as why faculty might be interested in trying a new software. All three of the faculty indicated that they are using the grading platforms Moodle, Perusall, and Gradescope, as well as spreadsheets and paper based grade trackers (e.g., binders, notebooks, etc.). One faculty stated that they use a custom linux program for grading exams. All of the faculty indicated some level of frustration with the current grading software. One respondent described, “always fighting something” in regards to either a current technology system or time. Other responses were more indicative of current grading ideologies, expectations and policies, as one faculty wrote “There is a tension I run into in keeping students as up-to-the-date informed of their grade as possible, but also making sure the current grade is correct. The more up-to-date, the more the class seems to center around "the grade" and always questioning points. The less up-to-date it is, the more the class gets to focus on the work itself.” These responses suggest that faculty willing to test out a new grading software, might be frustrated with current options or have concerns about how grading functions in broader conversations. Results from the student pre-survey provided us with an understanding of what students value and care in relation to grading and communicating about grades. Especially encouraging was that 48% of the students stated that a detailed understanding of how they were performing in a course, not just overall, but also *with respect to individual learning goals and tasks*, was *extremely* important to them. Less than 5% stated that it was only slightly important or not important to them at all. Grading platforms other than the RED System were judged by nearly 50% of students as only moderately useful, with 32% of the students rating them as less than moderately useful. Our pre surveys, thus, provided ample support for the motivation for our software development.

Faculty Theme 1: Time Management

In the post surveys faculty generally reported that the system made grade sharing faster and with more ease, though the degree varied. One respondent noted it "dramatically changed my workflow" and made it "a lot easier to share grades with students," while another who had previously built their own Google Sheets system said simply that it "made it easier to get grades to students." However, the time savings were not uniform — one respondent estimated it "cut it down somewhat, maybe an hour or two a week, probably closer to an hour," suggesting modest

gains. However, a different faculty member found that increased grade sharing actually created more work: "because I used the (RED System) messaging system, some students would question many of their grades and I would need to spend some time looking back and sending them a response." The benefit for some may reflect how cumbersome prior systems were rather than a fundamental improvement in workflow. Responses suggested including videos and templates to help enhance the user experience, which may influence time devoted to using the software.

Faculty Theme 2: Shifts in perception of grading and teaching
For some faculty the system prompted a reconsideration of how they approach grading and student feedback more broadly. One respondent described moving away from thinking within "the limits of what Moodle automated mainly because it was easier," feeling instead "more creative about grades, and a bit more free to experiment." Another described using graphical feedback to "adjust my course plan twice during the semester," suggesting the system supported more active, data-informed teaching decisions. However, this was not universal. One respondent admitted they were "still getting my head around" the system's possibilities, and another was candid that they were "not using the tagging feature nearly as much as I should." A more cautionary note came from one respondent who questioned whether the system genuinely expands faculty and student agency, asking "does this system take away some agency they have over their grade?" This speaks to a larger consideration about grade-sharing technology and the student-faculty relationship.

Across the five Likert scale questions on the student post survey over 51% of respondents rated the RED System positively, about 24% rated it neutrally, and 25% rated it negatively. The majority of students agreed that the software made it both easier and faster to access grade information. Particularly encouraging was that 70% of the students reported that the software made it easy for them to access information on how they were performing on *specific metrics* in a course, which was one of the prime motivators for the software. However, they still recommended themes for improvements in the software for interaction and usability. We generated two themes on how students thought the RED System can be improved for student experience.

Student Theme 1: Grade Visibility
Students made recommendations to improve on how they could see their grade overall for a class, as well as individual assignments. One student noted that "when I am looking at my grades I should be able to easily tell my standing in the class without having to enable and disable settings," while another described feeling like they "had no idea what my grade in the course was until after it was complete," contrasting this unfavorably with Moodle where they had "an understanding of how I was doing in the middle of the semester." The interface itself was a source of confusion. One student described it as difficult "to distinguish which grades corresponded to which assignments," and another found the flexibility of the system overwhelming, noting "there is so much flexibility for me when observing results I'm a bit overwhelmed." Suggestions included a running overall grade estimate, mentioned by multiple respondents in various forms: "a final grade estimation system," "a general grade at the end," and "I would like to see my final grade at will." This could reflect how individual professors share grades and at what times during the semester.

Student Theme 2: User Experience
Beyond navigation, students described the visual presentation of grades as a barrier to engagement. Several called for a cleaner UI, with one noting it was "very difficult to distinguish which grades corresponded to which assignments." Another felt that "the having of normalized and raw scores in one area makes the screen cluttered and too much going on at once." Students provided practical suggestions: color coding by grade range ("green 90-100%, yellow 70-90%"), automatic percentage calculations per assignment, tabbed organization by assignment type such as homework, exams, and labs, and notifications when grades are posted. Notably, in the open ended question, a few students explicitly expressed satisfaction with using the new software. Again, this could suggest the experience was uneven and may partly reflect how individual professors configured their gradebooks.

The faculty data suggests that while the software reduced some time devoted to grade sharing and communicating, its greater value could be on shifting perceptions to teaching practices. This is still largely dependent on individual faculty motivation and technical comfort. Students experienced the system differently, with many reporting confusion around basic grade visibility and interface navigation, raising a tension between the flexibility faculty valued and the clarity students needed. Taken together, these findings suggest that the system's potential may be constrained less by its technical capabilities than by the gap between how faculty configure and understand it and how students are able to access and interpret the information it presents.

This study contained two limitations related to the population samples studied. The first limitation is that the faculty who participated in the study are involved in the design and the development of the LMS software that they are studying. Their input as a participant-reseracher is valuable at this stage of the development, as it speaks to the dual experience of user and designer. However, it is important to note that the faculty experiences need to be interpreted with this positionality in mind. The second limitation is that the students' samples were only from one of the participating faculty's classes. The two other participating faculty did not include student perspectives from the surveys in their sample. This can give a partial and incomplete understanding of the entire student sample populations. Future research of this software will incorporate a broader faculty and student population to avoid incomplete data collection.

Conclusion
In this WIP, we invited faculty and students to test out a new software system that we developed to provide faculty and students with a diversified view of scores with *tags*. We sought to generate some initial themes from survey data of how students and faculty interact with the software, specifically about how it influenced grade sharing and communication practices. These themes and initial results will inform the development of the software, future research instruments, and the research design of a larger qualitative study on how the new software system can influence teaching and learning.

Acknowledgment – This material is based upon work supported by the National Science Foundation under Award Number 2022271 (IUSE/PFE: RED Innovation: Enabling Convergence in Undergraduate Engineering through Structural Change).

References

[1] S. R. Hamilton, C. M. Saviz, D. A. Saftner, T. Kunberger, "What's in a Grade? Current Practices and Strategies to Evaluate Learning in Engineering Courses," ASEE Annual Conference and Exhibition, Portland, OR, June 23-26, 2024.

[2] F. Musekamp, J. Pearce, "Assessing engineering competencies: the conditions for educational improvement," Studies in Higher Education, vol. 40, no. 3, pp. 505-524, 2015.

[3] J. D. Allen, "Grades as Valid Measures of Academic Achievement of Classroom Learning," The Clearing House: A Journal of Educational Strategies, Issues and Ideas, vol. 78, no. 5, pp. 218-223, 2005.

[4] W. B. Walstad, L. A. Miller, "What's in a grade? Grading policies and practices in principles of economics," The Journal of Economic Education, vol. 47, no. 4, pp. 338-350, 2016.

[5] M. Gettinger, J. K. Seibert, "Contributions of study skills to academic competence," School psychology review, vol. 31, no. 3, pp. 350–365, 2002.

[6] L. Borghans, B. H. Golsteyn, J. J. Heckman, J. E. Humphries, "What grades and achievement tests measure," Proceedings of the National Academy of Sciences, vol. 113, no. 47, pp. 13354–13359, 2016.

[7] B. Roth, N. Becker, S. Romeyke, S. Schäfer, F. Domnick, F. M. Spinath, "Intelligence and school grades: A meta-analysis," Intelligence, vol. 53, pp. 118–137, 2015.

[8] D. Clark, R. Talbert, Grading for Growth. New York, NY, Taylor and Francis, 2023.

[9] G. R. Burks, J. R. Amos, C. W. Castleberry, "Embedded Tagging and Radar Map Shape Analysis for Assessing Student Outcomes," 126th ASEE Annual Conference and Exhibition, Paper ID #27587, 2019.

[10] A. A. Lipnevich, T. R. Guskey, D. M. Murano, J. K. Smith, "What do grades mean? Variation in grading criteria in American college and university courses," Assessment in Education: Principles, Policy & Practice, vol. 27, no. 5, pp. 480-500, 2020.

[11] F. J. Fowler Jr, Survey research methods. Sage Publications, September 2013.

[12] J. Saldana, Fundamentals of Qualitative Research (Understanding Qualitative Research). 1st Edition, P. Leavy (Editor), Oxford University Press, 2011.