Ampel-Grader: Traffic-Light Feedback for Self-Regulated Learning in Digital Exercises

Author(s): Laurenz Neumann*, Maximilian Kissgen, Michał Słupczyński, Stefan Decker
RWTH Aachen University
*laurenz.neumann@dbis.rwth-aachen.de

ABSTRACT: Rising student numbers and limited teaching resources present challenges for grading and timely feedback in courses. As a solution, we present the Ampel-Grader, an automated grading tool to offer students traffic-light-based feedback on their exercises for self-regulated learning. Further, a Learning Analytics Dashboard was developed, providing instructors with insights into student performance. We employ a comparative analysis of student performance in a database course before and after the introduction of the tool, examining the impact on learning outcomes. User surveys indicate that the tool is received as a low-cost, real-time feedback system that enhances students' performance. The results suggest that the Ampel-Grader promotes self-regulated learning while reducing the workload of instructors.

Keywords: Automatic Grading, Self-regulated Learning, Digital Exercises, Learning Analytics, Jupyter Notebooks

1 INTRODUCTION

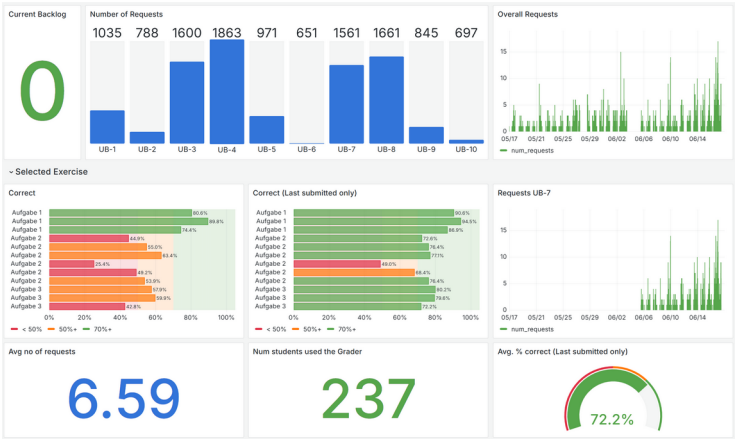
In learning environments, feedback is crucial for the success and effectiveness of tasks (Sadler, 1989). Continuous, formative feedback supports Self-Regulated Learning (SRL) by motivating regular self-assessment (Clark, 2012), but excessive feedback can hinder genuine understanding. Jupyter Notebooks (Project Jupyter et al., 2019) are widely used for interactive digital exercises, and nbgrader[1] extends them with automated grading. While many tools provide basic feedback on task completion and error logs (Keuning et al., 2019), some offer detailed feedback but are limited to specific tasks or platforms. LOOM (Duss, 2020) provides percentage-based feedback with visualizations but is restricted to certain question types and requires tasks to be written specifically for the tool. Similarly, the model by Steen et al. (Steen et al., 2017) offers hints and explanations but is dependent on a Learning Management System (LMS). Other approaches also rely on platforms like Moodle (Corino et al., 2022; Eitemüller et al., 2023) or require manual grading. To address these limitations, we propose the *Ampel-Grader* (AG), named after the German word “Ampel” for “traffic light”, a low-cost automated checking service for Jupyter-based exercises using nbgrader. Our solution supports various task types and is LMS -independent. Students receive feedback via traffic lights without detailed scores. Additionally, we provide a Learning Analytics Dashboard (LAD) for educators to monitor student progress and inform instructional decisions. Unlike existing tools, AG combines formative exercise assessment with LAD functionality without relying on an LMS or specific teaching environment (Nyland, 2018; Schwendimann et al., 2017).

[1] https://github.com/jupyter/nbgrader

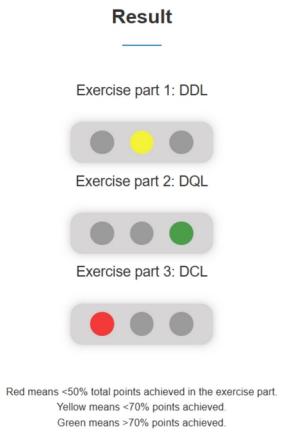
Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

2 IMPLEMENTATION

The AG application comprises three components: a backend, a frontend, and a LAD. The Django-based[2] backend handles exercise corrections with dedicated Grading-Workers, enabling concurrent processing and enhanced security. Feedback can be configured on sub-exercise level, enabling granular feedback with individual traffic lights.



(1) Learning Analytics Dashboard



(2) Student Results Page

Figure 1: Screenshots of the AG LAD and result interface. The dashboard (Fig. 1a) shows general and assignment-specific data for "UB-7," including request and correctness statistics. The result UI (Fig. 1b) gives partial feedback: red for <50%, yellow for <70%, and green for bonus-passing scores.

The frontend provides a form for students to upload completed Jupyter notebooks for grading. Grading results are displayed as traffic lights: one per sub-exercise or one for the entire exercise if sub-exercises are not configured (see Figure 1b). Django's built-in OpenID authentication system prevents unauthorized access. Teaching staff can upload exercise notebooks and set grading availability periods, e.g. to encourage early hand-ins, as well as rate limiting. To provide insights based on submission to the AG, the software includes a LAD for teaching personnel (see Fig. 1a). Built with Grafana[3], the dashboard visualizes key metrics on AG usage. It displays average points achieved per exercise and tracks submission metrics. Instructors can assess exercise difficulty and student performance using the average score for all submissions and the average score of each student's most recent submission. It also tracks average submissions per student for feedback needs and performance. In the test phase of AG we made adjustments: First, we added rate limits as the LAD suggested some students tried to brute-force multiple choice solutions. Second, issues with the exercises came up on the LAD before the final submissions, so we allowed to change exercise parameters to fix them.

3 EVALUATION SETUP

AG was deployed in a university database course for Computer Science bachelor students across two semesters as an optional tool along with exercises. We analyzed results from weekly homework assignments and final exams with and without the AG. Students required 50% of total points for exam eligibility and earned a grading bonus at 70%, represented by a yellow light for ≥50% and green for ≥70%. AG data includes submission times, anonymized student emails, and points per exercise. Points

[2] https://www.djangoproject.com

[3] https://grafana.com/grafana

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

were collected from Moodle, the course LMS, and matched with exam results via student identifiers or emails, manually linked when necessary. Unmatched entries were excluded, and students without AG database identifiers were treated as non-users. We compared average student performance between the current and preceding semesters, the latter without AG. In the semester without AG, students received a grading of the exercise after the submission deadline, lacking any preliminary feedback. In the lecture evaluations students were asked to give their opinions about AG. Analysis data is available on GitHub[4]. To assess expert perspectives, ten educators completed a questionnaire after a demo at the JTEL Summer School 2025[5], containing ten open-ended questions on perceived effectiveness, benefits, and limitations for SRL.

4 RESULTS AND DISCUSSION

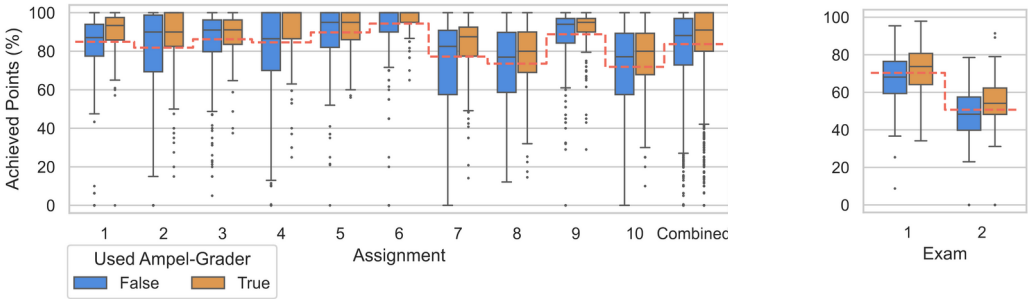


Figure 2: Student performance in exercises (left) and exams (right) for the first term, separated by prior AG use. Boxes are orange for tool use and blue for non-use. An additional boxplot shows all exercises combined. Average performance per exercise/exam is marked by the red dashed line. Students who used the tool generally performed better.

To evaluate the AG's impact, we analyzed the performance of students in two terms of a course in which the AG was used and compared the performance of students who used the tool with those who didn't in exercises and exams. The data set for the exercises in the first term contained around 526 – 596 students per exercise, apart from the tenth one, in which only 392 participated. 336 of those students could be matched in the AG. In the second term where the AG was used, the exercises were done by 575 – 636 students. 372 Students could successfully be matched. In the first term we evaluated, the first exam had 399 participants, while the second one had 128. In the second term, the first exam had 411 participants, while the second one had 157. As seen in Figure 2, Students generally scored higher when they had used the AG. In summary, the score increased on average by 4.36/20 points or 21% based per exercise in the first term and by 4.3% in the second term. To examine the AG's influence on student performance compared to a whole course in which the tool was not available, we used the 2023 term, where exercises were completed without the tool, and compared it to the 2024 term, where the AG was deployed. The comparison showed a 9.17% improvement in average grades following its introduction. Finally, students highlighted in the lecture evaluations that they enjoyed using AG and that it gave them more confidence in their submissions. Expert evaluations identified the system as promising for supporting SRL and self-assessment, emphasizing its ability to deliver immediate, on-demand feedback and stimulate reflective thinking. However, experts noted that the feedback may not be fit for a students' knowledge level and lacked actionable guidance. They recommended integrating more personalized feedback, such as contextual explanations to strengthen its pedagogical impact.

[4] https://anonymous.4open.science/r/LAK26-ampel-eval-6253

[5] https://ea-tel.eu/jtelss25

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

Comparative analysis of exercises and exams indicates that the AG can improve student performance, as evidenced by higher overall outcomes compared to the semester without it. However, the findings are context-bound to a computer science database lecture; generalization to other disciplines/courses remains uncertain. The tool's dependence on Jupyter notebooks may limit its use in specialized exercise setups, and its intentionally simplified feedback system—designed to reduce configuration complexity—can restrict deeper knowledge acquisition compared to more detailed SRL-oriented feedback. Nonetheless, expert evaluations recognized AG as a valuable aid for student learning. Overall, AG increases student performance and some trust more in the correction process, but a smaller number of students still make mistakes that the grader would help to correct or don't use it at all. In our opinion, AG reduced the grading effort as it catches issues in students' submissions before the deadline. The LAD allowed preliminary insights into whether the exercise had a fitting degree of difficulty or if there were some problems with some exercises specifically.

5 OUTLOOK

Even though AG was received well in evaluation, educators asked for personalized feedback and improvement hints. The tool could be expanded for this purpose in accordance with self-regulated-learning principles. This additional data could then be incorporated into the dashboard for more comprehensive feedback. Studies in further courses could then provide a clearer picture of the AGs' efficacy.

REFERENCES

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, *24*(2), 205–249.

Corino, E., Ghirotto, S., Gilli, L., & Bovo, D. (2022). Adaptive exercises and formative assessment for English remedial action. In D. Ifenthaler, P. Isaías, & D. G. Sampson (Eds.), *Orchestration of learning environments in the digital world* (pp. 3–19). Springer International Publishing.

Eitemüller, C., Nienaber, C., Fend, M., Dünnebier, U., & Eickelmann, B. (2023). Digitalization of multistep chemistry exercises with automated formative feedback. *Journal of Science Education and Technology*, 453–467.

Project Jupyter, Ragan, E., Barmak, H., & Box, M. (2019). nbgrader: A tool for creating and grading assignments in the Jupyter Notebook. *Journal of Open Source Education*, *2*(16), 32.

Keuning, H., Jeuring, J., & Spronck, P. (2019). A systematic literature review of automated feedback generation for programming exercises. *ACM Transactions on Computing Education*, *19*(1), Article 3.

Duss, K. (2020). *Formative assessment and feedback tool: Design and evaluation of a web-based application to foster student performance*. Springer Gabler Wiesbaden.

Nyland, R. (2018). A review of tools and techniques for data-enabled formative assessment. *Journal of Educational Technology Systems*, *46*(4), 505–526.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144.

Schwendimann, B. A., & Dillenbourg, P. (2017). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, *10*(1), 30–41.

Steen, A., Haugen, K. S., & Olsen, D. H. (2017). Atomized feedback, an approach to automatic formative assessment. IATED (International Academy of Technology, Education and Development).

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)