Designing a Learning Analytics Dashboard to Provide Students with Actionable Feedback and Evaluating Its Impacts

Xiaojing Duan[1], Chaoli Wang[1] and Guieswende Rouamba[2]
[1] *University of Notre Dame, Notre Dame, IN, USA*
[2] *Purdue University Fort Wayne, Fort Wayne, IN, USA*
xduan@nd.edu, chaoli.wang@nd.edu, grouamba@pfw.edu

Keywords: Learning Analytics Dashboard, Machine Learning, Self-regulated Learning, Actionable Feedback

Abstract: Various educational settings have begun to increasingly leverage the power of data analytics to optimize the learning environment and enhance the learning experience for students. However, despite this effort, significant research gaps still exist around utilizing educational data mining to provide students with actionable feedback and assess the comprehensive impact of data-informed feedback on students. In this study, a learning analytics dashboard was designed to provide students with actionable feedback to advance their self-regulated learning skills and improve their course performance. A rigorous inquiry using mixed methods was also conducted to study the dashboard's impacts on students. It found that students' use of the dashboard was positively correlated with their course performance, and those who viewed the dashboard had higher course ranks. In addition, it showed that students' use of the dashboard was positively correlated with their homework submission time, and those who viewed the dashboard submitted homework earlier as the course progressed. The inquiry also revealed that students had mixed feelings about the dashboard, including motivation and anxiety.

1 INTRODUCTION

With the emergence of computer-supported learning environments, much data about the learners and their context became available. As a result, a large body of research in the learning analytics and educational data mining community has begun to leverage the power of data to provide learners a better learning experience optimized to their individual preferences and needs (Clow, 2013; Siemens and Baker, 2012). The initial focus of that effort was to identify the at-risk students and provide them with just-in-time and personalized support to prevent potential drop-outs (Wong and Li, 2020; Syed et al., 2019; Choi et al., 2018). Although these studies reported success with their interventions targeted at-risk students, it is unclear what effects these systems have on a broader range of students or how to scale them to support all students.

Other studies have explored using learning analytics dashboard (LAD) as a form of personalized feedback for all students, such as Course Signals (Arnold and Pistilli, 2012) and StepUp! (Santos et al., 2013). However, the information presented in those dashboards is not actionable: students cannot use it to change their learning behavior and achieve better course performance. Additionally, few of the existing LADs were built upon established learning theories (Sawyer, 2014). The lack of theoretical background threatens the effectiveness of LADs. Furthermore, what impact LADs have on students is an important topic that remains underexplored. As suggested by Wong and Li (2020), there is a great need for more rigorous inquiries on the impact of LADs on students.

To bridge the gaps mentioned above, we designed a LAD that targets all students, not only those at-risk. The primary goals of the dashboard include: (1) raising students' awareness of the correlation between their learning progress and learning behavior; (2) motivating students to adopt more effective learning behaviors and improve their self-regulated learning (SRL) skills. The dashboard design is grounded in the widely adopted SRL model (Winne and Hadwin, 1998). The rationales are as follows: (1) LADs can assist students in evaluating their current state of SRL and progression towards their learning goals (Kim et al., 2016); (2) LADs can motivate students and assist them in reflecting on their SRL process (Muldner et al., 2015).

Our dashboard provides students with actionable feedback on their weekly learning progress and their patterns on the key learning activities that were proved to influence course performance. Those key activities were identified through machine learning (ML) models and confirmed by the subject matter experts. This study helps educators and researchers better understand the dashboard's impacts on students. It examined the correlation between students' use of the dashboard and their course performance. It also assessed the association between students' use of the dashboard and their learning behavior change. In addition, it investigated what emotions were triggered by viewing the dashboard and why. This study addresses the following research questions:

- RQ1. What learning activity features strongly influence students' course performance?
- RQ2. How do we design a LAD that provides students with actionable feedback on the key learning activities?
- RQ3. What impacts does this dashboard have on students' course performance, learning behavior, and emotions?

In the remainder of the paper, we first review related work in Section 2. Next, we describe the methodology for the present study in Section 3, followed by the results in Section 4. Then we discuss the findings in Section 5, followed by the limitations and future opportunities in section 6. We conclude the study in section 7.

2 RELATED WORK

Identification of Influencing Learning Features. Determining what learning features to present is one of the most important steps in LAD design. Matcha et al. (2019) suggested using the user-centered design approach to discover new features. Ott et al. (2015) used predictors previously identified to predict student success and visualized them. Feild (2015) conducted an exploratory data analysis to determine the indicators that were accustomed to their study subjects. They analyzed data points, including days of the semester, days of the week, hours of the day, and start and submit times of students' assignments. To determine the learning activity features for our dashboard, we followed the process suggested by Bodily and Verbert (2017). It includes reviewing the literature, conducting exploratory data analysis, and using a theoretical framework. We also utilized ML techniques and discussed the identified features with course instructors and other domain experts.

ML in Educational Data Mining. A variety of ML techniques, such as classification, regression, deep neural networks, and reinforcement learning, have been utilized to predict students' course performance and model their learning behavior (Baradwaj and Pal, 2011; Okubo et al., 2017; Zhou et al., 2018; Moreno-Marcos et al., 2018; Wan et al., 2019; Deng et al., 2019). Additionally, ML techniques have been adopted to model learning sequences and detect learning strategies. For example, Akpinar et al. (2020) used pattern mining and Natural Language Processing (NLP) models to extract learning strategies from students' clickstream data and found those strategies are correlated with students' course homework grades. Jovanovic et al. (2019) clustered students' pre-class activities and identified learning strategies from those activities that are correlated with students' course grades. In our study, we followed the human-centered artificial intelligence principle. We discussed the results of ML models with the course instructors, so the features we selected were better aligned with the course design.

LAD. Most LADs provide at-a-glance views of various information collected from the learning environment, such as the frequency of logins, click sequences, and time spent on a task by utilizing information visualization techniques. Examples include PerformanceVis (Deng et al., 2019), Moodle (Podgorelec and Kuhar, 2011), LOCO-Analyst (Ali et al., 2012), SNAPP (Bakharia and Dawson, 2011), and Students Success System (Essa and Ayad, 2012). Those examples demonstrate that the initial effort on utilizing LADs to support students focused on highlighting students at risk of academic failure and proposing interventions for instructors. Our dashboard aimed to provide all students with actionable feedback on their learning activities, motivate them to develop more effective learning activities, and help them perform better.

SRL and Feedback. The SRL model defined by Winne and Hadwin (1998) is widely used in research related to computer-supported learning. It includes five recursive components: conditions, operations, products, evaluation, and standards (COPES). In this COPES model, feedback occurs internally when students evaluate their learning against the goals they set for themselves. However, as shown by Bjork et al. (2013), students may not be able to assess themselves accurately. More specifically, the underachieving students tend to overestimate their learning while the overachieving students tend to underestimate themselves. Winne and Jamieson-Noel (2003) also found that students' self-reports were not completely aligned with their own actions. Additionally, Malmberg et al. (2014) showed that students' misperception between their learning progress and learning outcome could lead to the choice of less effective learning practices. Therefore, our dashboard is designed to provide students with external feedback enhanced through intuitive and interactive visualizations, help them develop a more accurate estimation of their learning progress, and thus better self-regulate their learning. We followed the conceptual model proposed by Sedrakyan et al. (2020) that incorporates SRL theory, feedback theory, and LAD design. According to that model, LADs can provide four different types of feedback: cognitive, behavioral, outcome, and process-oriented feedback. Our dashboard focused primarily on behavioral, outcome, and process-oriented feedback.

3 METHODOLOGY

3.1 Study Context and Data Collection

The study was conducted on an introductory coding course required for all business major students at the University of Notre Dame. The primary goal of this course is to help students learn the fundamentals of coding and develop fluency with the Python programming language. The course consisted of 13 modules, and each module introduced a new programming topic. The course was delivered using the flipped classroom model in which students were expected to complete several learning tasks before the in-person class. During the in-person session, instructors guided students to do more practice on the topics they learned in the preparation materials. Except for the in-class participation, all the other learning activities were conducted in Canvas (a learning management system), Vocareum (a cloud programming system), and Panopto (a video streaming and management system). When students interacted with those systems, their learning activities were captured in the system logs. We collected data from the course when it was offered in Spring 2021 and Fall 2021. There were 45 and 69 students enrolled in the two semesters, respectively. After consolidating the data, we captured the activity data that reflect students learning behaviors and habits, as described in Table 1.

The course performance data we captured included students' final letter grades (on a range from A, A-, B+, B, B-, C+, C-, D, to F) and calculated grades (on a scale from 100 to 0).

3.2 Feature Extraction

One of our goals with the collected data was to identify the learning activity features that greatly influenced students' course performance. As we learned from the literature review, the significance of the features in predicting student success depends on the setting and structure of the course. Therefore, we consulted with the course instructors and extracted a set of features that are better aligned with the course design, as described in Table 2. The features representing students' click count on the course materials are commonly used in student success prediction. The new features proposed in this study include how many days before the deadline a student opened a homework (tutorial/lab), submitted it for the first/last time, and how many days passed between the opening of the homework and the first/last submission. Those features are proxies for how early students started homework and how much time they spent on it.

3.3 Model Building and Evaluation

We ran several ML models on the extracted features to determine their predictive significance for the course grade. The models we used include logistic regression, k-nearest neighbors, naïve Bayes, decision tree, and random forest regression. We evaluated those models' performances using accuracy, precision, recall, F-measure, and area under the PRC. The evaluation showed the random forest regression model had the best performance. Another reason we chose this model was that its result was easy to interpret, so we could discuss it with the instructors.

3.4 Dashboard Design

After identifying the learning activity features that strongly influenced students' course performance, our next step is to design a dashboard that provides students with actionable feedback on those important features. We followed a user-centered iterative process to create the dashboard. The design process is illustrated in Figure 1.

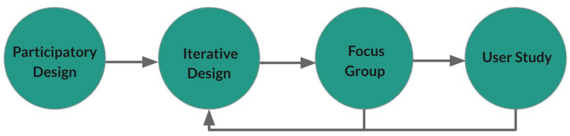

Figure 1: User-centered iterative design process.

Participatory Design. We started with a participatory design session with the course instructors. In this session, we discussed the important learning activity features determined by the random forest regression model. They confirmed that the features with higher predictive significance represented the critical learning activities for the course from the pedagogical perspective. They also commented that watching lecture videos was a key learning activity, although the model did not rank its significance as high as the lab survey responding activity. They wanted to cultivate video-watching behavior. Therefore, we decided to show students' tutorials submission and video watching activities in the dashboard. After each module in the dashboard, we also included their accumulated grade rank to raise students' awareness of the correlation between their learning activities and performance. To motivate students to regulate their learning

Table 1: Collected data field names and descriptions.

Name	Description	System
overview_click	Date and time when a student clicked on the module overview pages	Canvas
quiz_review_click	Date and time when a student clicked on a quiz to review	Canvas
discussion_view	Date and time when a student viewed an online discussion thread	Canvas
discussion_post	Date and time when a student posted an online discussion thread	Canvas
survey_respond	Date and time when a student responded a homework survey	Canvas
recording_click	Date and time when a student clicked on a class recording	Canvas
hw_start_time	Date and time when a student opened a tutorial/lab	Vocareum
hw_submit_time	Date and time when a student submitted a tutorial/lab	Vocareum
submit_count	Submit count on the tutorials/labs by a student	Vocareum
video_watched_second	Watch length in seconds of the lecture videos by a student	Panopto
video_watched_perc	Watch percentage of the lecture videos duration by a student	Panopto
video_watched_time	Date and time when videos were watched by a student	Panopto

Table 2: Extracted feature names and descriptions.

Name	Description
overview_count	Click count on the module overview pages by a student
assgn_count	Click count on the tutorials/labs by a student
quiz_count	Review count of the quizzes after submission by a student
discussion_view_count	View count of the online discussions by a student
discussion_post_count	Post count on the online discussions by a student
survey_count	Respond count to the lab surveys by a student
recording_count	Click count on the class recording pages by a student
due_start_days	Days before the deadline a student opened a tutorial/lab
due_first_days	Days before the deadline a student submitted a tutorial/lab for the first time
first_start_days	Days between a student opened a tutorial/lab and submitted it for the first time
final_start_days	Days between a student opened a tutorial/lab and submitted it for the last time
submit_count	Submit count on the tutorials/labs
video_sum_mins	Watch length in minutes of the lecture videos by a student
video_avg_diff	Average watch percentage of the lecture videos by a student

behaviors, we decided to show the tutorial submission and video watching patterns of students ranked in the top 25% of the class. We chose the top 25% as the benchmark because the course had a competitive grading policy, which requires the average GPA to fall into the range of 3.2 to 3.5. We sketched a dashboard prototype using paper and pencil in the participatory design session.

Iterative Design. After the participatory design session, we converted the paper sketch into a prototype using Tableau (an interactive data visualization software focused on business intelligence). There were a couple of reasons we chose Tableau to create the dashboard. First, the university where the study was conducted has deployed a Tableau server with the single-sign-on feature, which means students can access the Tableau dashboard with their university IDs. Second, Tableau supports user-filtering, which allows us to map a student with his/her data so users can only see their own data in the dashboard. Finally, we chose a line graph for the accumulated grade rank because it is easy to interpret. We decided to use a Gantt bar graph to compare an individual's tutorial submission and video-watching behaviors with those of the top 25% performers. The Gantt bar graph shows the contrast clearly and is easy to understand.

Focus Group. With the dashboard prototype, we had a focus group with two of the course instructors and a learning research director from the same university. During the focus group, we had an unmoderated discussion while the participants interacted with the prototype. The reason for the unmoderated discussion was to observe how the participants interacted with the dashboard and record their feedback. For example, one instructor interacted with the line chart showing the students' accumulated grade rank trend. She was concerned that showing the exact grade rank would mislead/discourage students because a low rank does not necessarily mean a bad grade. The other instructor suggested showing only when the rank falls into or outside the top 25%. They also changed their minds to include the lab submission activity, which is another homework component. Additionally, they suggested changing the dashboard layout so the explanatory text is closer to the associated graphs and users can easily reference them when viewing the graphs. We redesigned the dashboard based on their feedback and presented the new design for more feedback. We repeated this process three times until they were satisfied with the design.

User Study. We conducted a user study with four students ($n_{female} = 2$, $n_{male} = 2$) who took this course in Spring 2021 to evaluate the dashboard design. They were in the same year and similar majors as the targeted audience of the dashboard. We used the think-aloud protocol during the study and asked them several pre-designed questions. The participants joined the study via Zoom. During the study, the interviewer shared with the participants a link to a dashboard that was built on some hypothetical data and told them to assume those data were theirs. Then the interviewer asked the participants to share their screens in turn when they interacted with the dashboard and encouraged them to talk aloud about their thoughts. From this study, We learned that all participants could interpret all the charts in the dashboard as they were designed. They also cross-referenced them and drew some high-level insights such as "*homework activity and video activity are kind of like variables going into the grades percentile range, so you can look at those to explain*" and "*if you start earlier, it'll work out better*". Their overall impression of the dashboard usability was "*the color is good*", the message "*is very clear*", the text " *makes a lot of sense and it helps explain*", and the dashboard contains "*the right amount/mix of information*". They all viewed the dashboard as a useful tool and made comments like "*yeah I definitely do think it's useful*" and "*I feel like the two graphs cover different aspects of the course pretty well, so I think it's good*". Additionally, they offered helpful suggestions on the improvement of the dashboard design. For the chart displaying a student's grade rank, they suggested emphasizing that the rank was based on the accumulated grades. So, we repeated that information in the tooltip of the chart. For the chart showing a student's video watching pattern, one participant suggested adding labels showing "*what module the videos attached*" so students can easily cross-reference their grade rankings with video watching patterns. We followed his suggestion and added the module information to the video labels.

3.5 Empirical Study

Study Protocol. To study the dashboard's impacts, we conducted an empirical study with students who took the course in Fall 2021. Fifty-five out of the enrolled 69 students agreed to participate voluntarily. We created a dashboard for every participant using his/her homework (tutorial&lab) submission, video watching, and grade data. We also integrated the dashboard into the course site in Canvas so participants can easily access it. The dashboard was first released after the class completed Module 7 and started Module 8. It contained students' accumulated grade ranks after Modules 3, 4, 5, 6, and 7. Modules 1 and 2 did not have any graded assessments. It also included students' homework submissions and video watching patterns in the first seven modules. Since then, the dashboard has been updated weekly. A reminder email was also sent to the participants, informing them that the dashboard was updated with the latest module they had completed.

Survey. Throughout the empirical study, participants were encouraged to submit their feedback through a Google Form survey embedded in the dashboard. The survey started with one multiple-choice question asking how the participants felt after viewing the dashboard. The question provides a list of emotions that address the activity emotions (enjoyment, confusion, boredom), prospective outcome emotions (hope, motivated-to-improve, anxiety, and hopelessness), and retrospective outcome emotions (pride, relief). This instrument also measures both positive and negative emotions as well as the activating and deactivating emotions. Combining both the valence and activation dimensions, as suggested by Pekrun et al. (2011), the emotions we tried to measure fall into these four categories: positive activating (enjoyment, hope, motivated-to-improve, pride), positive deactivating (relief), negative activating (confusion, anxiety), and negative deactivating (hopelessness, boredom). Participants can choose multiple emotions in their answers. On the second update of the dashboard, we added two more questions. One asked how easy or hard it was to understand the dashboard; the other asked what actions the participants would likely take after viewing the dashboard.

Statistical Analysis. We captured participants' interaction with the dashboard, such as when they accessed it and how many times they viewed it. We used this data in conjunction with the participants' course performance and learning activity data to investigate if the participants who viewed the dashboard performed better or demonstrated any behavior change. We used Pearson's correlation analysis to evaluate any correlation between viewing the dashboard and grade rank. We also compared the grade ranks of participants who viewed the dashboard and those who did not. Because the grade ranks of both groups were not in a normal distribution, we used

Table 3: Predictive significance of the extracted features.

Features	Predictive Significance
due_first_days	0.4527
survey_count	0.1552
video_avg_diff	0.1003
overview_count	0.1002
discussion_view_count	0.0661
assgn_count	0.0660
quiz_count	0.0425
recording_count	0.0170

the one-tailed Mann–Whitney U test to determine if there was a significant difference in the grade rank between those two groups. In addition, we compared the tutorial/lab submission habit change between participants who viewed the dashboard and those who did not. The change measured in hours was not in normal distribution for either group, so we used the one-tailed Mann–Whitney U test again to test if those two groups demonstrated a significant difference.

Interview. We conducted a semi-structured interview with five of the empirical study participants ($n_{female} = 2$, $n_{male} = 3$). The goal was to understand their reactions to the dashboard and its impacts on them. The interview followed the think-aloud protocol and was complemented by some pre-designed questions. The questions include how participants used the dashboard, what is their perceived usefulness of the dashboard, what actions they took after viewing the dashboard and why, and how they felt after using the dashboard and why. The participants joined the interview via Zoom. Before the interview started, the interviewer explained the goal to the participants and received their oral consent for audio and video recordings of the interview. Next, the interviewer screen-shared a dashboard with hypothetical data to guide the conversation. The interview lasted for around 30 minutes. After the interview, we performed thematic analysis on the transcript.

4 RESULTS

4.1 RQ1: Features Identification

Predictive Features. The result of the random forest regression model is presented in Table 3. As it indicates, the top three features that are most significant in predicting the course grade are due_first_days ($p = 0.4527$), survey_count ($p = 0.1552$), and video_avg_diff ($p = 0.1003$). The model's mean absolute error (MAE) is 2.089, and the root-mean-square error (RMSE) is 2.272. Given the predicting label (course calculated grade) is on a scale from 0 to 100, this error rate is acceptable.

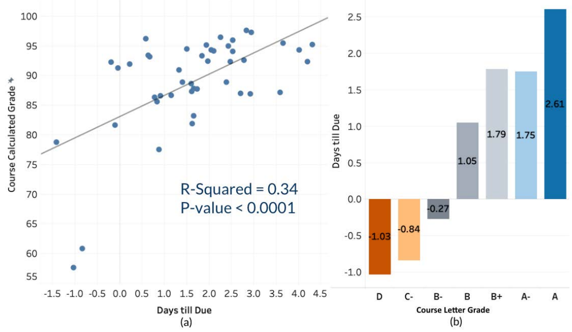

Figure 2: (a) Linear correlation between due_first_day feature and calculated grade (b)Patterns on due_first_day feature by different grades group.

4.2 RQ2: Dashboard Design

Predictive Features Evaluation. To further evaluate the significance of the identified predictive features, we visualized their correlations with the course grade and the feature patterns demonstrated by students in different grade groups. For example, as shown in Figure 2(a), the due_first_days feature has a positive linear correlation with the course calculated grade. Figure 2(b) shows the first submission of students in the A group (2.61 days before the deadline) was earlier than the students in any other grades group. On the other hand, the first submission of students in the D (1.03 days after the deadline) and C- (0.84 days after the deadline) groups were later than the students in any other grades group. When we showed those visualizations during the focus group, the instructors agreed that the top features reflect the important learning activities from the pedagogical perspective. They were not surprised that how early students submitted the homework had the most significant influence on their course performance. It is due to the flipped classroom nature of this course. Students were expected to learn the course content in the tutorials and practice them before the in-person class. The earlier they started that task and the more time they had to study the material, the better their performance was. If the students put off that task, they might not have sufficient time to learn the content, which leads to an undesired outcome.

Dashboard. Figure 3 shows the version of the dashboard that the instructors and learning research director were all satisfied with. It also incorporates the suggestions we received. The dashboard contains an embedded feedback form and three charts:

- The YOUR FEEDBACK button directs viewers to the Google Form.
- The grades percentile rank chart shows if a student's accumulated grade rank after a module is ranked in the class top 25% or not. The green cir-

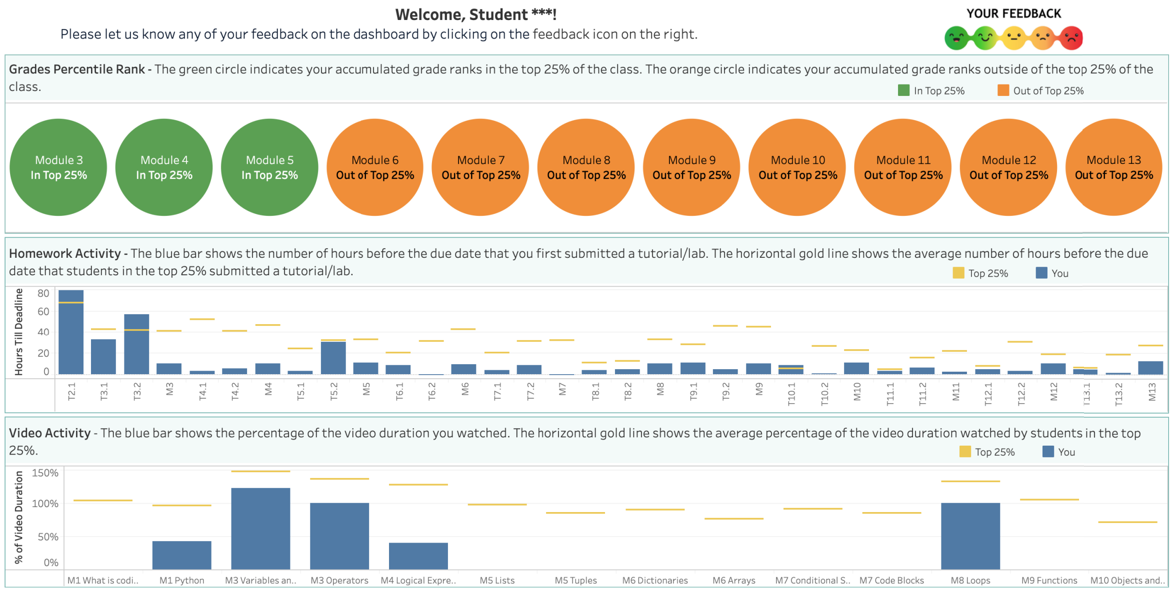

Figure 3: Our final dashboard design. Top to bottom: grades percentile rank, homework activity, and video activity charts.

cle indicates the grade rank falls into the top 25% and the orange circle indicates the grade rank falls outside of the top 25%.

- The homework activity chart shows how early a student submitted a tutorial or lab for the first time compared to the top 25% performer's average submission status. The blue bar indicates the number of hours before the deadline that a student submitted a tutorial/lab for the first time. The horizontal gold line shows the average number of hours before the deadline that the top 25% performers had their first submission on the same tutorial/lab.
- The video activity chart shows how much of the lecture video a student watched compared to the top 25% performers' average watched ratio. The blue bar shows the percentage of the video duration a student watched. The horizontal gold line shows the average percentage of the video duration watched by students in the top 25%.

Dashboard Evaluation. The dashboard was evaluated through the embedded survey and the interview. The insights we gained from the survey and the interview are as follows:

- Usability of the dashboard: among the 23 responses to the survey question on how easy or difficult it was to understand the dashboard, 9 (37.5%) were very easy, 11 (45.8%) were easy, 1 (4.2%) was neutral, and 3 (12.5%) were hard. Among the five interview participants, 3 (60%) thought the dashboard was very easy to understand, and 2 (40%) thought it was easy to understand. All five interviewees agreed that students could use the dashboard without any training.
- Perceived usefulness of the dashboard: among the five interview participants, 3 (60%) commented that the dashboard raised their awareness of the correlation between their grade and learning activity, and 2 (40%) held a neutral opinion. When asked if they like to have a similar dashboard for other courses, 3 (60%) interviewees answered with strongly yes, 1 (20%) was yes, and 1 (20%) was neutral.
- Follow-up actions after viewing the dashboard: among the 23 responses to the survey question asking the participants what actions they will likely take after viewing the dashboard: 12 (52.3%) were starting the tutorials & labs early, 5 (21.7%) were watching more lecture videos, 3 (13%) were nothing because they were ranked in top 25%, and 3 (13%) were nothing because they feel confident with their study strategies. Among the five interviewees, one shared he "*reviewed the tutorial more*" after viewing the dashboard. Another interviewee added she spent "*more time to complete the assignments*".
- Impact of the dashboard: when asked about the dashboard impact, one interviewee shared, "*I do think that it has been motivating and driving for me. It's just kind of a reminder every week to keep working hard*". Another participant added, "*I would say it's definitely motivating for me because like I said I've been outside the top 25%*

Table 4: Pearson correlation between students' viewing count of the dashboard and their grade rank.

	Pearson Coefficient	P-Value
Grade_Rank_M9	0.26	0.033
Grade_Rank_M10	0.30	0.011
Grade_Rank_M11	0.35	0.003
Grade_Rank_M12	0.35	0.003
Grade_Rank_M13	0.35	0.003

Table 5: Grade rank comparison between students who viewed the dashboard and those who did not.

	Viewed	Not Viewed
Student_Count	49	20
Mean_Rank (Std)	0.55 (0.27)	0.34 (0.29)
Median_Rank	0.57	0.24

for a bit. So every week, I want to get the email, I hope that I've hopped back in the top 25%". The third participant shared that the dashboard raised his awareness of the correlation between his grades and learning activity. He commented, "*I was in the top 25% for the start and then there were a couple homework assignments that I started real late or videos I didn't watch and that's what I dropped out of top 25%*".

4.3 RQ3: Dashboard Impacts

Impact on Students' Course Performance. The Pearson correlation analysis showed a positive correlation between the number of times a student viewed the dashboard and their grade rank. As shown in Table 4, the number of times students viewed the dashboard is positively correlated to their accumulated grade rank after Module 9, $r(67) = 0.26$, $p = 0.033$. More importantly, the coefficient increases to 0.3 after Module 10 and 0.35 after Modules 11, 12, and 13. In the meantime, the p-value decreases to 0.011 after Module 10 and 0.003 after Modules 11, 12, and 13. It indicates that the positive correlation becomes more significant as the course progresses. The analysis also showed that the number of times students viewed the dashboard was positively correlated to their accumulated grade rank change from Module 8 to Module 13, $r(67) = 0.31$, $p = 0.01$. We also compared the final accumulated grade rank (after Module 13) of students who viewed the dashboard and those who did not. As the results in Table 5 show, both the average and median grade ranks of students who viewed the dashboard are higher than those who did not. We performed a one-tailed Mann-Whitney U test to test if that difference is statistically significant. The result is $U = 699$, $n_{Viewed} = 49$, $n_{Not\ Viewed} = 20$, $p = 0.003$, showing that the group who viewed the dashboard has a higher rank than the group who did not.

Table 6: Submission behavior change comparison between students who viewed the dashboard and those who did not.

	Viewed	Not Viewed
Student_Count	49	20
Mean_Change (Std)	8.62 (20.43)	-1.29 (9.63)
Median_Change	5.11	0.94

Impact on Students' Learning Behaviors. Regarding the dashboard's impact on students' learning behaviors, we focused on their homework submission behavior because the instructors recommended they watch the lecture videos to prepare for the final exam. That recommendation might influence students' video watching behavior change. We defined the submission behavior change as the difference of students' average submission statuses between Modules 9-13 (after the dashboard was released) and Modules 3-8 (before and during the dashboard release). The Pearson correlation analysis showed a positive correlation between the number of times students viewed the dashboard and their homework submission behavior change, $r(67) = 0.33$, $p = 0.006$. We also compared the submission behavior change of students who viewed the dashboard and those who did not. As shown in Table 6, on average, students who viewed the dashboard submitted homework 8.62 hours earlier than what they did before the dashboard was released. On the contrary, students who did not view the dashboard submitted homework 1.29 hours later than what they did before the dashboard was released on average. We performed a one-tailed Mann-Whitney U test to test if that difference is statistically significant. The result is $U = 655$, $n_{Viewed} = 49$, $n_{Not\ Viewed} = 20$, $p = 0.015$, showing that the group who viewed the dashboard demonstrates a more positive submission behavior change than the group who did not.

Impact on Students' Emotions. We received 38 responses to the survey question asking how the participants felt after viewing the dashboard. As Figure 4(a) shows, 35 of the reported emotions were negative, including anxiety, hopelessness, confusion, and boredom. Nineteen of the reported emotions were positive, including pride, relief, motivated to improve, hope, and enjoyment. The top negative emotion was anxiety, and the top positive emotion was pride. Since the participants can answer this question by choosing multiple emotions, we broke down the responses. As Figure 4(b) shows, some of the responses include both positive and negative emotions at the same time. For example, 5.4% of the responses were motivated to improve and anxiety, 2.7% were pride and anxiety, and another 2.7% were pride, relief, and anxiety. We included a similar question during the interview and asked the participants to elaborate on why they had

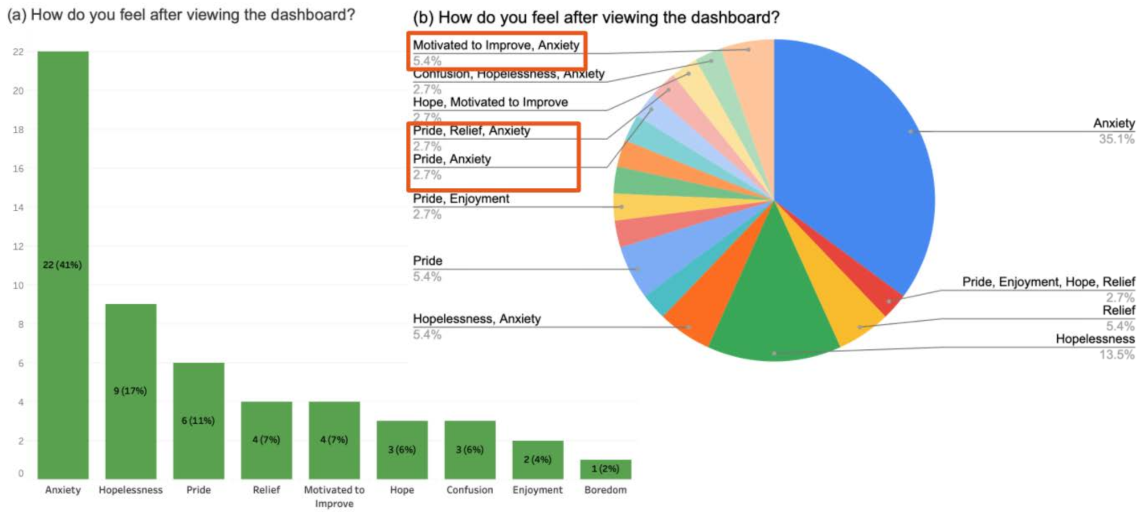

Figure 4: Reported emotions after viewing the dashboard.

certain emotions. Four of the 5 participants answered that question. All of the four responses were motivated to improve. One response goes like this, "*The dashboard motivates me to do better when I am behind the top 25% of the class in a category, and it makes me feel satisfied when I am in the top 25% of a category*". Another one was "*while it could be disappointing to see a[n] orange circle, it does positively motivate me to work towards becoming the top 25%*". The third one was "*rather than inciting negative emotions, it encourages me to improve my studying habits for the next module. If I see that I have performed in the top 25%, that is a pretty positive note so it incites positive emotions for me*". The fourth one was "*it generally prompts positive emotions, especially when I see that I fall within the top 25%. I enjoy seeing how my learning activities impact my grade and what I can do to improve my grade when I fall out of the top 25%*". These findings are consistent with the literature suggesting that social comparison generally appears to be motivating; however, some students do not like it (Bennett and Folley, 2021).

5 DISCUSSION

RQ1 focused on identifying the learning activity features that strongly influence course performance. The workflow we used to answer this question can be summarized into three steps: feature engineering, model building & evaluation, and verifying the identified features through visualizations. Other educational researchers or practitioners can adopt this workflow to identify the influencing features from their own data. Our result showed one of the new features we proposed in this study had the highest significance in predicting students' course performance. That feature measures how early before the deadline a student submitted homework for the first time. It is a proxy for measuring students' level of procrastination and time-on-task. Since accurately measuring time-on-task has been a challenge, our study sheds light on how to overcome that challenge. Our result also underscores the importance of the ML model's explainability. With an easy-to-interpret model, we can engage the instructors and other domain experts to evaluate the result effectively.

The answer to RQ2 describes the user-centered iterative design process for our dashboard. It demonstrates the benefits of involving the instructors, learning scientists, and students in the design process. By observing how they interact with the dashboard prototypes and listening to their feedback, we developed a deeper understanding of the sense-making process of different audiences. This deeper understanding helped us and can also aid the LAD researchers and practitioners in designing more effective and intuitive dashboards.

RQ3 studied the dashboard's impacts on students' course performance, learning behaviors, and emotions. To the best of our knowledge, this study is the first one that evaluated the dashboard's impacts from those three perspectives. Our results show a statistically significant correlation between students' use of the dashboard and their course performance. Furthermore, the correlation strengthens over time. Our results also reveal a statistically significant correlation between students' use of the dashboard and their homework submission behavior change. More importantly, we found that students who used the dashboard submitted homework earlier than they did before the dashboard was released. Finally, our evaluation of the dashboard's impact on students' emotions showed mixed results. While some students reported they were motivated to improve by the dashboard, others reported anxiety. It was interesting to learn some students felt both pride and anxiety simultaneously. One interviewee helped us understand why. Her comment was "*I am proud of the progress I has achieved while feel anxious about maintaining the high performance moving forward*". This reveals the complexities of the emotional impact of technology-mediated feedback.

6 LIMITATIONS AND FUTURE WORK

We admit that there exist a few limitations in this study. First, it was conducted in a single course with 55 participants. This relatively small sample size limits the generality of the findings. Second, all the participants were business majors. It's unclear what impacts the dashboard would have on students in other majors. Third, the users of the dashboard were voluntarily signed-up but not randomly assigned, which could result in a biased evaluation of the dashboard's impacts on students. We did not use a randomized controlled trial in this study because the instructors were not comfortable randomly deciding who has access to the dashboard and who does not. The dashboard could potentially benefit all students, and they consider it unethical to withhold it from any students who want to access it.

To overcome the limitations mentioned above, we will expand the study to include multiple large gateway courses offered in three different institutions in the future. With a large group of diversified participants, we can gain more insights on the dashboard design and further evaluate the dashboard's impact on students with different demographics and academic backgrounds. In addition, the large diversified sample will allow us to experiment with releasing the dashboard to different groups at different times and study the impact of when to release the dashboard on students. We will also explore the ethical options of randomly assigning students as dashboard users.

7 CONCLUSIONS

This study presents the design process of a dashboard that provides all students actionable feedback to improve their SRL skills. It also shares the dashboard's impacts on students' course performance, learning behaviors, and emotions. To the best of our knowledge, this study is the first one that evaluated the dashboard's impacts from those three perspectives. The results reveal new perspectives of the dashboard's impacts on students and open the door for future studies to gain more insights.

ACKNOWLEDGEMENTS

This research was supported in part by the U.S. National Science Foundation through grant IIS-1955395. We would like to thank everyone in our team for their feedback and suggestions, especially Alex Ambrose, Frederick Nwanganga, and Steve Sporinsky. We would also like to thank the students who participated in this study.

REFERENCES

Akpinar, N.-J., Ramdas, A., and Acar, U. (2020). Analyzing student strategies in blended courses using clickstream data. *arXiv preprint arXiv:2006.00421*.

Ali, L., Hatala, M., Gašević, D., and Jovanović, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education*, 58(1):470–489.

Arnold, K. E. and Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of International Conference on Learning Analytics and Knowledge*, pages 267–270.

Bakharia, A. and Dawson, S. (2011). SNAPP: A bird's-eye view of temporal participant interaction. In *Proceedings of International Conference on Learning Analytics and Knowledge*, pages 168–173.

Baradwaj, B. and Pal, S. (2011). Mining educational data to analyze student's performance. *International Journal of Advanced Computer Science and Applications*, 6(2):63–69.

Bennett, L. and Folley, S. (2021). Students' emotional reactions to social comparison via a learner dashboard. In *Visualizations and Dashboards for Learning Analytics*, pages 233–249. Springer.

Bjork, R. A., Dunlosky, J., and Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, 64:417–444.

Bodily, R. and Verbert, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies*, 10(4):405–418.

Choi, S. P., Lam, S. S., Li, K. C., and Wong, B. T. (2018). Learning analytics at low cost: At-risk student prediction with clicker data and systematic proactive interventions. *Journal of Educational Technology & Society*, 21(2):273–290.

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education*, 18(6):683–695.

Deng, H., Wang, X., Guo, Z., Decker, A., Duan, X., Wang, C., Ambrose, G. A., and Abbott, K. (2019). PerformanceVis: Visual analytics of student performance data from an introductory chemistry course. *Visual Informatics*, 3(4):166–176.

Essa, A. and Ayad, H. (2012). Student success system: Risk analytics and data visualization using ensembles of predictive models. In *Proceedings of International Conference on Learning Analytics and Knowledge*, pages 517–520.

Feild, J. (2015). Improving student performance using nudge analytics. In *Proceedings of International Conference on Educational Data Mining*, pages 464–467.

Jovanovic, J., Mirriahi, N., Gašević, D., Dawson, S., and Pardo, A. (2019). Predictive power of regularity of preclass activities in a flipped classroom. *Computers & Education*, 134:156–168.

Kim, J., Jo, I.-H., and Park, Y. (2016). Effects of learning analytics dashboard: Analyzing the relations among dashboard utilization, satisfaction, and learning achievement. *Asia Pacific Education Review*, 17(1):13–24.

Malmberg, J., Järvelä, S., and Kirschner, P. A. (2014). Elementary school students' strategic learning: Does tasktype matter? *Metacognition and Learning*, 9(2):113–136.

Matcha, W., Gašević, D., Pardo, A., et al. (2019). A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE Transactions on Learning Technologies*, 13(2):226–245.

Moreno-Marcos, P. M., Muñoz-Merino, P. J., Alario-Hoyos, C., Estévez-Ayres, I., and Kloos, C. D. (2018). Analysing the predictive power for anticipating assignment grades in a massive open online course. *Behaviour & Information Technology*, 38(10-11):1021–1036.

Muldner, K., Wixon, M., Rai, D., Burleson, W., Woolf, B., and Arroyo, I. (2015). Exploring the impact of a learning dashboard on student affect. In *Proceedings of International Conference on Artificial Intelligence in Education*, pages 307–317.

Okubo, F., Yamashita, T., Shimada, A., and Ogata, H. (2017). A neural network approach for students' performance prediction. In *Proceedings of International Conference on Learning Analytics and Knowledge*, pages 598–599.

Ott, C., Robins, A., Haden, P., and Shephard, K. (2015). Illustrating performance indicators and course characteristics to support students' self-regulated learning in CS1. *Computer Science Education*, 25(2):174–198.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., and Perry, R. P. (2011). Measuring emotions in students' learning and performance: The achievement emotions questionnaire (AEQ). *Contemporary Educational Psychology*, 36(1):36–48.

Podgorelec, V. and Kuhar, S. (2011). Taking advantage of education data: Advanced data analysis and reporting in virtual learning environments. *Elektronika ir Elektrotechnika*, 114:111–116.

Santos, J. L., Verbert, K., Govaerts, S., and Duval, E. (2013). Addressing learner issues with StepUp!: An evaluation. In *Proceedings of International Conference on Learning Analytics and Knowledge*, pages 14–22.

Sawyer, R. K. (2014). The future of learning: Grounding educational innovation in the learning sciences. In *The Cambridge Handbook of the Learning Sciences*, pages 726–746. Cambridge University Press.

Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., and Kirschner, P. A. (2020). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior*, 107:105512.

Siemens, G. and Baker, R. S. J. d. (2012). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of International Conference on Learning Analytics and Knowledge*, pages 252–254.

Syed, M., Anggara, T., Lanski, A., Duan, X., Ambrose, G. A., and Chawla, N. V. (2019). Integrated closed-loop learning analytics scheme in a first year experience course. In *Proceedings of International Conference on Learning Analytics and Knowledge*, pages 521–530.

Wan, H., Liu, K., Yu, Q., and Gao, X. (2019). Pedagogical intervention practices: Improving learning engagement based on early prediction. *IEEE Transactions on Learning Technologies*, 12(2):278–289.

Winne, P. H. and Hadwin, A. F. (1998). Studying as self-regulated learning. In *Metacognition in Educational Theory and Practice*, pages 277–304. Lawrence Erlbaum Associates Publishers.

Winne, P. H. and Jamieson-Noel, D. (2003). Self-regulating studying by objectives for learning: Students' reports compared to a model. *Contemporary Educational Psychology*, 28(3):259–276.

Wong, B. T.-M. and Li, K. C. (2020). A review of learning analytics intervention in higher education (2011–2018). *Journal of Computers in Education*, 7(1):7–28.

Zhou, Y., Huang, C., Hua, Q., Zhu, J., and Tang, Y. (2018). Personalized learning full-path recommendation model based on LSTM neural networks. *Information Sciences*, 444:135–152.