École Polytechnique Fédérale de Lausanne

Co-Designing a Teacher Tool for Visualizing Self-Regulated Learning Behaviors

---

Eva Laini, January 2023

Digital Humanities Master Project at ML4ED, supervised by Pr. Tanja Käser

Abstract: Flipped classrooms, in which students engage with the materials before the class and use face-to-face time for more interactive and personalized learning activities, have become increasingly popular in recent years. While this approach has the potential to improve student learning and engagement, it also requires teachers to have effective tools for managing and supporting self-regulated learning (SRL). One such tool is the teacher dashboard, which provides teachers with real-time data on student progress and allows them to monitor and provide feedback to students learning in a flipped classroom.

In this project, we followed a teacher-centered approach to design a teacher dashboard adapted to flipped classroom context that presents university-level teachers with information on students' SRL behaviors and incorporate a novel clustering pipeline identifying learning behaviors patterns. This approach allows for a better understanding of teachers' requirements for flipped classes in terms of data and visualizations in order to design a dashboard tailored to their actual needs. We derived the requirements from 10 teacher interviews; then, ran a large-scale study with 92 teachers to test educators' visual design preferences. Next, we improved the prototype design iteratively with seven teachers and evaluated the tool with ten distinct teachers for clarity, usefulness and actionability.

Keywords: Flipped Classroom; Teacher Dashboard; AI Co-design; Clustering; Visualization Study; User Testing

1 Introduction

Nowadays, more and more educational approaches make use of blended learning, where traditional face to face instruction is blended with online educational materials. In particular, flipped classrooms offer multimedia lectures as homework for students to free up class time for discussion and student-centered learning activities [30]. Research has shown that blended classrooms require a high level of self-regulated learning from students to be effective[1]. Self-Regulated Learning (SRL) is the "self-directive process by which learners transform their mental abilities into academic skills" [34]. SRL behaviors can be analysed with the help of Learning Analytics (LA) [8]. In particular, LA can can be used to identify patterns in the students' learning behavior in a flipped classroom [14]. Moreover, student profiles with higher levels of SRL behaviors are correlated with better academic performances [4].

However, SRL can be a challenging task for many learners and while multiple SRL monitoring tools have been designed for students [7][33], these tools have often overlooked educators' role in supporting and guiding students in their learning experience. More recent research has valued teacher dashboards to monitor and support students' SRL behaviors positively [15] and multiple tools have been designed to this end. However, these tools mainly focus on fully online classes [31] or real-time classroom orchestration [13]. Moreover, teacher dashboards in general mostly show aggregated statistics on students behaviors and research found that the information provided was not enough [24].

In addition, previous work has shown that patterns of SRL behaviors can be found using ML models and are lost when showing aggregated data [20][14]. In this project, a profiles clustering method is leveraged to visualize trends in SRL behaviors patterns and show teachers more precise insights on their students SRL behaviors to better support them in course adaptation and targeted interventions. This clustering method was designed [19] and tested on university-level flipped classroom data [20] by Machine Learning for Education (ML4ED) lab members. This method uses ML to extract learning dimensions from students log data, the record of all their interactions with the online materials of a blended classroom, and cluster them into profiles with similar learning behaviors patterns.

Incorporating ML results and insights into teacher dashboards poses some challenges. Research has found that teachers may have trouble understanding and trusting insights compiled with the help of AI [23][22]. Indeed, most educators are not expected to have any knowledge of ML, AI or even statistics. Thus, especially when the LA displayed on the dashboard become more complex than simple statistics, such as ML and AI models, the data needs to be conveyed in a way that all educators, even those without data analysis knowledge, would find clear and actionable. Previous work has addressed these challenges by providing novel interaction features [27] or by introducing new kinds of prototyping methods [13]. Furthermore, these issues can be addressed by using a participatory approach when designing a teacher dashboard, which more and more researchers are using ([26], [32] among others). This co-design approach allows the researchers to take into account educators' needs and concerns from the very beginning of the design process and tailor the tool to their actual needs and concerns. While data type and tool features are designed using teacher requirements, this approach overlooks the actual data visualization. Indeed, while the type of information shown is a key aspect of a dashboard, the way it is shown can be as important. This is especially the case when displaying ML insights, considering the challenges mentioned above. General research on data visualization is extensive [9][21], but even when it focuses on insights and decision making [3], it offers only general guidelines. Thus, there is a need for research on teachers' visual design preferences and understanding, as well as actionability of the visualizations.

To address the research gaps and issues mentioned above, this project follows a a co-design approach, including educators at three different steps of the design process. At the very first step, 10 educators at university level were interviewed to better understand their interest and needs in terms of student's behaviors in the online part of a blended course. We also assessed their interest in having access to learning profiles grouping students with similar learning patterns. From their answers, requirements were compiled for a teacher dashboard showing student's learning behaviors and behavior patterns. Complementing these requirements, a study was created to systematically test graph designs conveying the information requested and a hundred educators were asked to answer various questions that allowed the visualization designs to be refined. After a dashboard prototype was designed based on the user requirements and the survey findings, the prototype design was refined through iteration with 7 teachers. Finally, the dashboard prototype was tested via semi-structured interviews with 10 potential users, university-level teachers with blended learning experience. The testing aimed to assess the general usability of the prototype, as well as the clarity and actionability of the information conveyed. The tests were analyzed both in a qualitative human-centered way and via the click-stream analysis of the users' activity during the testing.

To ensure our dashboard design was accessible to everyone, including people with disabilities, we followed the guidelines from the *Web Accessibility Initiative (WAI)* who compile accessibility guidelines and standards for accessible websites.

This methodology allowed us to answer two research questions:

1. Which SRL behaviors are teachers most interested in and can SRL behaviors clustering be leveraged in a blended classroom context?
2. How can data on students' SRL behavior be presented in a teacher dashboard in an accessible, clear and actionable way ?

2 Methods

2.1 SRL Dimensions and Profiles

In order to ensure the data displayed on the dashboard was relevant to the goal of helping teachers promote SRL behaviors in their students, the possible SRL dimensions were taken from previous work by lab members [19]. In their paper, they defined SRL dimensions based on extensive existing research on students' SRL. They then designed a clustering pipeline using these dimensions to identify patterns in learning behaviors. Finally, they adapted the pipeline and tested it on flipped classroom data [20].

The clustering pipeline is separated into three distinct steps. The first step is feature extraction from students' log data in the online part of the course. The following dimensions are used as features to represent students' behavior:

Proactivity attempts to measure the extent to which students are up-to-date or behind schedule. Up-to-date students have seen the required lecture materials for all weeks up to the current.

Effort aims to monitor the intensity of student engagement in the course, which is fundamental for learning success. It consists of the time spend on the online material and the activity, which is the number of clicks in the lecture videos (play, pause, fast-forward, etc).

Consistency is concerned with time management throughout the semester. It refers to how students distributed their effort throughout the weeks. It consists of the relative time online and activity as video clicks by week compared to the total for all weeks.

Control models the in-video behavior as a proxy of student ability to control the cognitive load of video lectures. It consists of the video playback speed and the pause frequency, which is the number of pauses compared to the video length.

Regularity is also associated with time management; capturing whether a student is regularly engaged on specific weekdays or day times, which means there will be a peak in activity.

In the next two steps, the SRL dimensions are used to classify students' learning behaviors and group them based on the learning patterns found. Different groups of students with similar patterns were identified and called **learning profiles**. Each profile has its own characteristics; for example, effort can be high or regularity can show a strong peak of activity in specific days.

The grouping of students is done in two steps. In the first step, the students are classified into learning behaviors groups for each of the dimensions using Spectral Clustering on the similarity matrices of the dimensions between students. Groups of students behaviors per dimension are created and the categories are labeled using expert knowledge. For example, one student might be categorized as "works consistently", while another might be "works more at the end of semester" for the consistency dimension. In the second step, students showing the same patterns of behaviors are grouped together to form the learning profiles. The second clustering is done using K-Modes on the previous cluster assignments. The details of the clustering pipeline can be found in the original paper [19], while its adaptation to flipped classroom data can be found in a following article [20].

This clustering pipeline results in two different types of data that are displayed in the dashboard. The first type is the grouping of SRL behaviors done for each dimension. Graphs are created using the data of each group separately and thus, the teachers can visualize patterns of behaviors for each of the SRL dimensions. The second type of data is the profiles and their characteristics. The proportion of students in each profile and its characteristics are shown to teachers in the dashboard. The characteristics of the profiles consists of the labels for the learning behavior groups that form them.

Student Behavior	**Profile A**	**Profile B**	**Profile C**	**Profile D**
Proactivity	More up-to-date	More up-to-date	Less up-to-date	Less up-to-date
Effort	Higher intensity	Lower intensity	Higher intensity	Lower intensity
Consistency	Constant work	Work before exams	Constant work	Work before exams
Control	Fast with pauses	Fast with pauses	Slow watchers	Slow watchers
Regularity	Peak before class	Peak before class	Peak before class	No peaks

Figure 1: Table of characteristics per profiles as shown in the dashboard prototype

We applied the pipeline to log data collected from 292 students, 29% of which identifying as female, of an undergraduate mathematics FC course [12]. The clustering resulted in two groups with distinct behavior patterns per dimension (e.g. higher and lower intensity of Effort). Since we did not have access to grades for the course and the link of profiles to academic performance was essential to our tool, we hand-crafted profiles using previous knowledge to keep them coherent. Each profile was linked to a fictitious average grade for all students composing it. Figure 1 shows an example of possible group labels ('More up-to-date', 'Lower intensity', 'Fast watchers' etc) for the SRL dimensions and the hand-crafted groups clustering into profiles, as shown in the dashboard prototype.

A goal of this project was to assess whether displaying the groups and profiles on a teacher dashboard would allow teachers to identify unproductive SRL behaviors in groups of students and gain insights on how to counter them. Our hypothesis was that, by offering teachers insights on their students' behavioral patterns, they could make informed decisions on their teaching methods and better support students in their learning.

2.2 User Interviews

In order to design a dashboard coherent with the actual needs of teachers, requirements were compiled both from the existing literature and from potential users interviews.

Ten teachers at university level were interviewed in a semi-structured format. The interviewees were mixed evenly between women and men. Their age ranged from 24 to 58, but skewed to the younger side, with a mean at 30 years old. The interviews lasted between 20 and 50 minutes with a mean of 32 minutes. We asked them asked questions about their teaching experience with flipped classroom, the kind of data they would be interested in knowing about their students and the way they would want this information to be displayed. They were also presented with the possibility of getting learning behavior profiles and asked how they felt about it[1].

From these interviews, a list of requirements for the dashboard was compiled and some possible data visualizations were designed. While most of the data requirements were consistent with the previously established list of SRL dimensions (see section 2.1), some suggestions were not compatible with the existing log data. Thus, not all requirements could be implemented. The out-of-scope suggestions are discussed in section 4.

2.3 User Study

The designs collected from the interviews needed to be refined to be usable and understood by teachers of all backgrounds. In addition, since ML-made insights can be complex to understand, we wanted to make sure the information displayed was comprehensible and actionable. To this end, we explored how to best communicate the data via a user study conducted on English-speaking educators from the Prolific platform[2]. The study lasted around 40 minutes on average and participants had to answer various types of questions that would help refine the designs prompted by the interviews. The survey consisted of five parts; one for the profiles, then one for each learning dimension, in randomized order. The respondents were given a short scenario and asked to answer questions related to their understanding of the graphs and the insights and actions they prompted, as well as improvement suggestions. They were also asked to rank visualization types and layouts according to their clarity and usefulness. Similar to previous research[2], we visualized time series using different designs and graph types. Figure 2 and Fig 4 show examples of the types of visualizations the participants of the study were asked to analyze and rank. In addition, following established taxonomy [11], we explored comparative designs of multiple time series including juxtaposition, superposition and explicit encoding of relationships, as shown in Figure 3.

To measure visualization literacy, we inserted four tasks adapted from a Visualization Literacy Assessment Test [17]: retrieve value (item 55), find extremum (item 2), find trend (item 38) and make comparisons with relative values (item 15).

The ranking questions of the survey were scored automatically, while text answers were analyzed using `BERTopic`[3] to find common trends in answers regarding graph analysis, insights and actions prompted. The results of this study were analyzed using the participants' demographics to try and understand whether age, gender and level of students had an impact on the kind of graph design teachers thought were clearer and more actionable. We analyzed the preferences

[1]The complete protocol of the user interviews can be found in the supplementary materials.

[2]The complete questionnaire of the user study can be found in the supplementary materials.

[3]`https://maartengr.github.io/BERTopic/index.html`

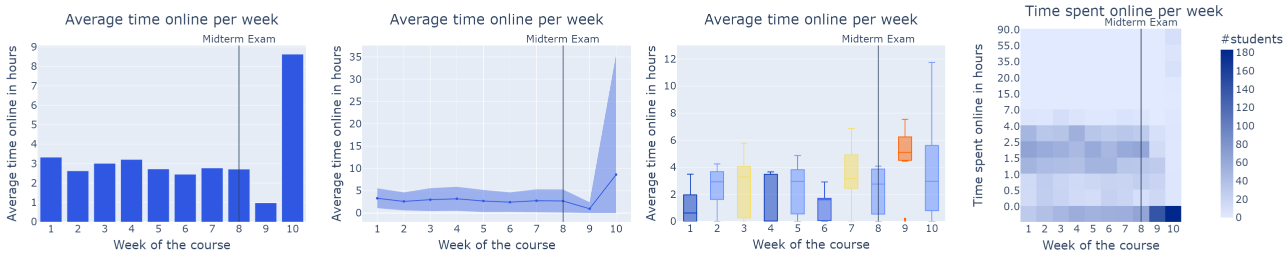

Figure 2: Examples of visual designs explored in the survey. From left to right: a bar plot, a line plot with standard deviation, a box plot with color as function of the median and a heat map, all showing the time spent online per week.

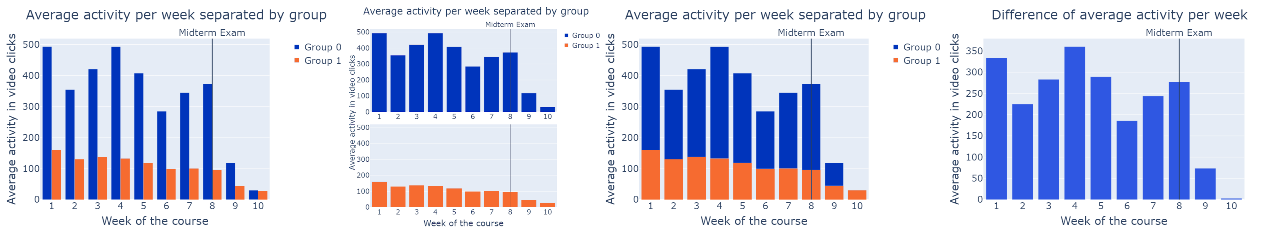

Figure 3: Examples of group comparison designs explored in the survey. From left to right: superposition (SP) side by side, juxtaposition (JP) separated, SP overlay and explicit encoding via group difference, all showing the group comparison of average activity per week.

differences based on demographics (sex and age) and personal factors (teaching level). We used the non-parametric test Kruskal–Wallis to evaluate the ordinal rankings. When significant, we performed a pairwise comparison using the Mann-Whitney U test, correcting for multiple comparisons via a Benjamini-Hochberg procedure.

2.4 User Testing

Once the dashboard prototype was created based on the compiled requirements, the prototype went through several rounds of changes and improvements with the feedback of 7 educators and researchers in the field of education. Finally, the prototype was tested on 10 university-level educators with blended learning experience using a classic Design Thinking user testing structure [25]. Before the beginning of the user test, participants were asked to fill in a short

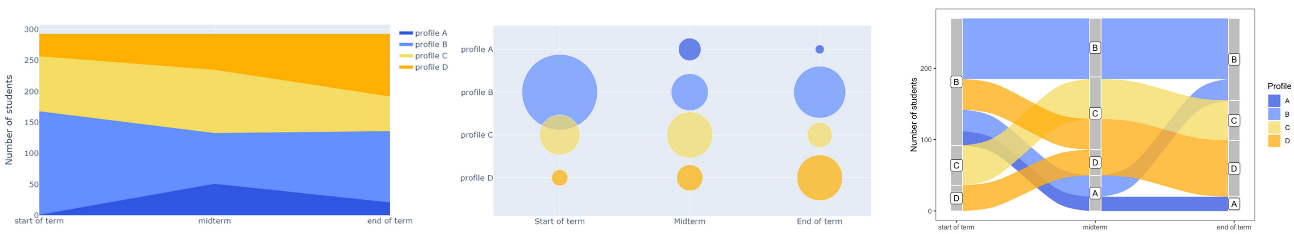

Figure 4: Examples of charts for the evolution of profiles through time explored in the survey. From left to right: an area graph, a bubble chart and a Sankey flow chart, all showing the number of students in each profiles at different time points.

questionnaire about their trust in AI[4]. The questionnaire consisted of 9 questions taken from an instrument measuring teachers' trust in AI-based EdTech [22] and hand-picked to be relevant to our study. This allowed us to get a baseline on their perception of AI in the field of education and put their potential concerns into context.

The user tests lasted 30-50 minutes and were conducted in two distinct phases. In the first phase, the participants were given a simple scenario and could explore the dashboard freely. They were asked to use a think-aloud process, where any thoughts, comments, questions or suggestions that the dashboard prompted would be voiced aloud. In addition to the think-aloud data, the screens and click-streams were also recorded to facilitate analysis. At the end of their exploration, they were asked to find specific information in the dashboard they might have missed in their exploration to assess whether the dashboard design was intuitive and important data could be found easily. In the second phase, they were asked about the way they would use the data displayed to adapt their course and what actions they prompted. They were also asked about any concerns they might have with using the tool. Finally, they were asked to share what part of the dashboard they found unintuitive or unclear and what information or feature they would use on a regular basis[5].

The user tests were analyzed in two ways, qualitatively and via computational methods. First, the think-aloud data from the exploration part of the study and the answers to the questions were studied in a qualitative way to extract all valuable comments and suggestions from each user. Next, the test transcripts were transcribed using `Whisper`[6], a state-of-the-art zero-shot model for speech recognition, and analyzed using `BERTopic`[7], a topic modeling technique that integrates the contextual information of the text by clustering embeddings generated by pre-trained transformer-based language models to find common topics mentioned during the user testings. Finally, a click-stream analysis was conducted to evaluate the usage of the prototype in terms of time spent on each page and flow through the dashboard.

3 Results

3.1 User Interviews: Requirements

The goal of the potential users interviews was to get a better understanding of educators and their data needs in a blended classroom context. We also aimed at assessing whether learning profiles could be a useful asset to present to teachers in a flipped classroom dashboard. From these interviews, we compiled requirements for the dashboard and the visualizations it would display.

All but one educator interviewed agreed having data about their students' learning behaviors would help them adapt their course and give personalized feedback to specific student profiles. The most important requirement, mentioned by 8 out of 10 interviewees, was anonymous data and aggregated statistics, which is a requirements also mentioned in the literature [15]. 6 out of 10 educators were concerned about data privacy and the intrusive nature of the information provided by the dashboard. While educators shared ethical concerns, they all

[4]The Trust in AI questionnaire can be found in the supplementary materials.
[5]The complete protocol of the user testing can be found in the supplementary materials.
[6]`https://openai.com/blog/whisper/`
[7]`https://maartengr.github.io/BERTopic/index.html`

mentioned general statistics they would be interested in seeing. The ones suggested by at least half of the interviewees were; whether students viewed the material, how much time they spent watching the videos and the amount of students keeping up with the class, all three mentioned by 6 out of 10 educators. Some other statistics mentioned by a few educators were student's consistency in watching the videos, the average video playback speed and whether students worked at the last minute before an exam, suggested by 5, 4 and 2 educators respectively. 3 interviewees said they would like to have a general overview of the students behaviors with no detailed statistics. Their reasoning was that studying strategies are too subjective, they do not give useful information or could be used to discriminate.

However, when presented with the possibility of getting learning behavior profiles for their students, all but one educator responded that, granted their anonymity, they would be very interested in having access to student profiles and their characteristics. They mentioned using them to improve the course, on the fly or for the next semester, and measure workload. As one educator mentioned: "[With access to profiles] If one year I observe this behavior, I can assume it would be similar the next year and I could adapt". Some also mentioned being interested in linking the profiles with the average grade of students forming it. In particular, one educator said that: "[Knowing profiles characteristics and the grade linked], I could try to tune the class and give more opportunities to [the lowest performing profile] to improve their grade".

When asked when they would use the dashboard, responses were mixed between weekly to adapt the class on the go and at the end of the semester to assess the success of the class and the improvement that could be done for the next semester.

Table 1 shows an overview of the dashboard features with the user requirements that prompted them and the supporting literature. These finding allowed us to refine the dashboard features and design an interface prototype. The design prototype, shown in figure 5, features an overview page, a learning profiles page and one additional page per SRL dimension, with possibility to switch between all and group view. Always visible are, on top, the name of the dashboard, later defined as FlippED, the course selector and time frame selector, and on the side, the navigation bar.

The educators were also asked to draw possible visualizations and dashboard designs. These illustrations were used as a base to design the visualizations tested in the user study, see section 3.2 for the results.

3.2 User Study: Visualizations Design

Building on the results of the potential users interviews, we created a user study to assess the clarity and actionability of the visualizations design.

We recruited 100 participants from the Prolific platform, 8 of which failed the attention checks, leaving the total of answers analyzed to 92. Demographics for the study's participants were well balanced between men and women with 51% of women. The age of respondents ranged between 21 and 76 years old, with a median at 37 years old. The participants came at 68% from Europe, with 41% from the United Kingdom and only 15% from non-Western countries. 34% of participants reported teaching at higher education level, 29% at high school, 23% at elementary school, and the remaining 14% included kindergarten, adult education, and vocational education.. While there was no significant difference in age distribution between genders or teaching levels, we found that 25% of women reported teaching at higher education

Feature	User requirements	Literature
Overview with key point from each dimension	A general overview with no detailed statistics	
	Whether students viewed the materials	
General student profiles proportion with characteristics	Anonymous student profiles	[15]
	Visualize the characteristics of each profile	
	Profiles to improve class and measure workload	[15]
Graphs and insights on general statistics for each dimension	Student's consistency in watching the videos	[28]
	Average video watch time	[5]
	Whether students worked at the last minute before an exam	[29]
	The amount of students that are keeping up	[29]
	Average video playback speed	[10]
The average grade is shown for each profile and group	Link student's behaviors and behavior patterns to grade	[4]
Possibility of choosing the time frame	See the statistics by week or for the whole semester	
Everything is anonymized	Privacy concerns about individual student profiles	[15]
	Showing individual data would be intrusive	
Accessible interface	Make an inclusive and accessible dashboard	[16]

Table 1: Features of the dashboard prototype with corresponding user requirements and supporting literature.

level compared to 52% of men. However, based on our statistical results, there does not seem to be a statistically significant difference between responses from men and women who indicated they taught at university level, suggesting that teaching level seems to be more significant in relation to visualization preference than gender.

Graph literacy: During the visualization study, participants were asked questions to test their visualization literacy level (VLL). We computed the VLL as a weighted average using as weights the items' discrimination index times the content validity ratio from [17] (0.6 for retrieving the value, 6 for finding extremum, 14 for finding trend, and 6.4 for making comparisons with relative values). The mean VLL was 0.7/1.0 ($\sigma = 0.3$) and the personal (teaching level) and demographic (gender, age) factors did not explain the variance. Nevertheless, we found that when describing the plots, more participants in the top VLL quartile reported the trends and the speed of the changes (e.g., gradual/slow). For example, for the *Consistency* time series (c_1), the trends were reported by 16% of participants in the top quartile and only by 5% in the bottom quartile.

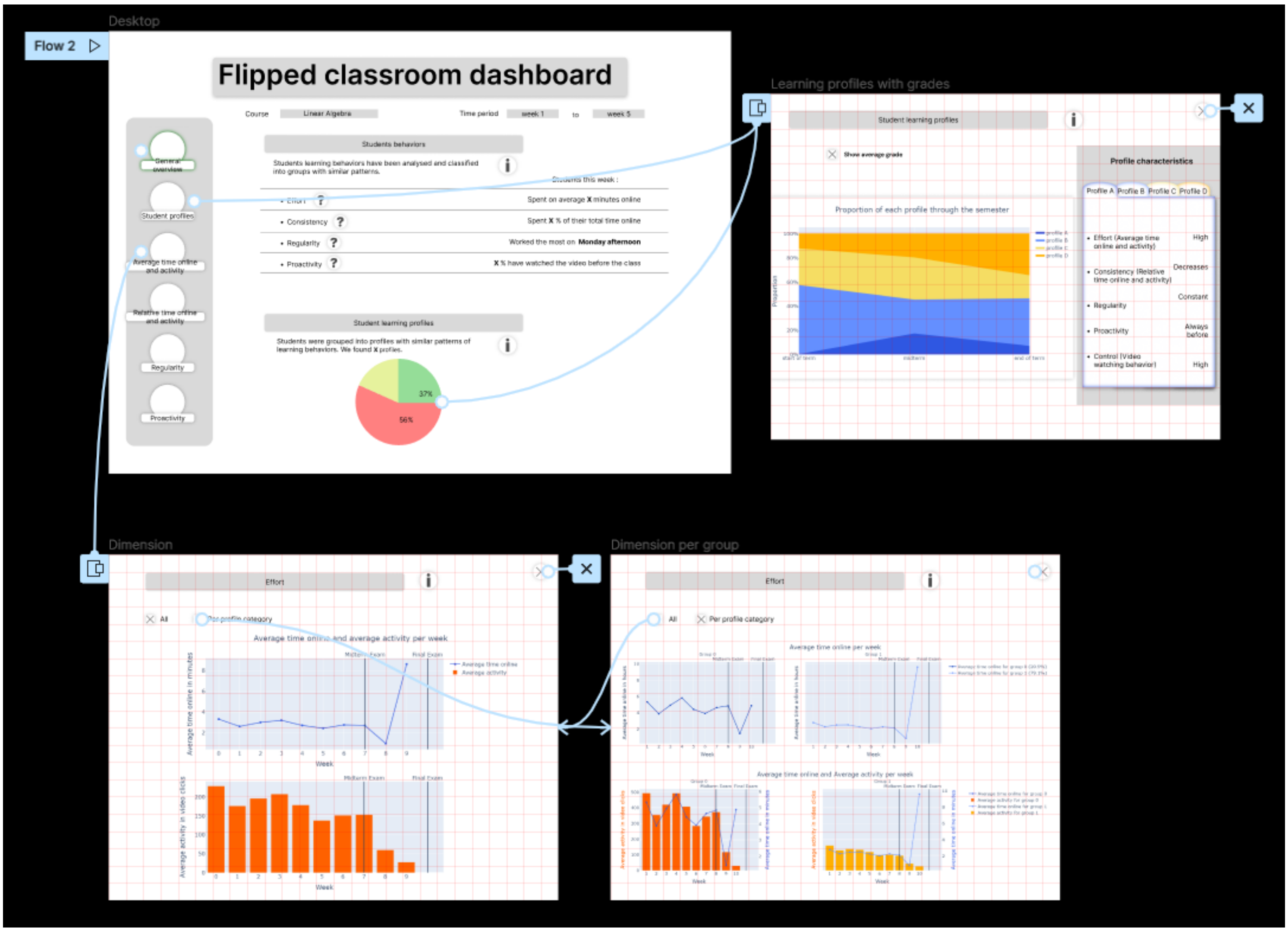

Figure 5: Dashboard design made using Figma. At the top right is the first design of the general layout and the overview page, with general statistics for all SRL dimensions and profiles proportion at one point in time. The top right shows the profiles pages with profiles proportion through time and profiles characteristics. The bottom shown a SRL dimension page with the option to switch between aggregated view and view by groups.

Graph complexity: A general observation we derived from this study was that, even though some educators found more complex visualizations to convey more information, a vast majority agreed that simpler visualizations were more clear and useful. Indeed, when we analyzed the topics in the written insights derived from the plots, we found two topics (out of 13) unique to the heat map. Participants were able to gather more specific insights with this complex visualization. Although this information could have also been inferred from the line plot with standard deviation (SD) and the box plot, it was only mentioned for the heat map. Despite this finding, people often mentioned liking simpler visualization, such as bar or line charts better in their comments, as shown by the *clarity* topic appearing in the choice explanations. Indeed, we found that participants preferred a bar plot ($\mu_{rank} = 2.1$) over a bar plot overlayed with a line plot ($\mu_{rank} = 2.7$), a line plot with standard deviation ($\mu_{rank} = 3.1$), just a line plot ($\mu_{rank} = 3.3$), a heat map ($\mu_{rank} = 4.7$), and a box plot ($\mu_{rank} = 4.9$). The graph type respondents felt was most clear and useful was a simple bar chart, which is coherent with previous findings [2], where bar plots outperformed other graph types (heat map, line and box plot) in identifying statistical properties. Interestingly, educators at university level ranked the line plot with SD higher ($\mu_{rank} = 2.7$) than the rest of the teachers ($\chi^2(2) = 4.3$, $p = .03$).

Graphs layout: Regarding the graphs layout, respondents were mixed with wanting different features, such as time online and activity, to be separated or in the same graph for easier comparison. Since a major concern mentioned by participants was clear graphs, we decided to separate the features in order to simplify the graphs layout. 75% of respondents felt that adding clear legends for each trace made the visualizations easier to understand.

Group comparison: To compare groups of students with similar behaviors for one feature, 84% of participants agreed superposition (SP, overlaying objects in the same space) allowed for easier comparison than juxtaposition (JP, showing different objects separately) or explixit encoding ($\mu_{rank} = 1.9, 2.4$ and 2.6 respectively). The preferred visualizations showed the data as bars with the groups side-by-side ($\chi^2(3) = 67$, $p < .001$), though some participants mentioned the graphs getting overwhelming when too much information was displayed. Despite the fact that explicit encodings can offload the burden of comparison from the viewer [11], almost all respondents agreed that showing the difference between the groups was the least clear and useful. Some comments called it 'unnecessary' and 'not easily actionable'. However, participants of *younger age* (younger than the 25^{th} age percentile of 30 years) ranked the difference-between-time-series graph higher ($\mu_{rank} = 2.9$) than the rest of the participants (μ_{rank}= 3.3, $\chi^2(3) = 4.6$, $p = .03$).

Others: When showing the number of students who watched the materials before, after the interactive session or did not watch at all, 87% or participants agreed that the 'did not watch' category was useful. As comments mentioned, "[the graph with all categories] brings attention to [students who did not watch]" and "otherwise, we might forget [students who did not watch]".

Participants had mixed responses when asked whether they preferred to see percentages or absolute number of students for statistics such as up-to-date students or profiles proportions. However, 71% of university teachers preferred percentages and some of them suggested adding the total number of students to the y-axis label so that it can be calculated with the percentages if desired. The final design showed percentages with total number of students in the legend.

Learning dimensions: Concerning the particular learning dimensions, we could see a strong difference in insights and action ideas variety as seen by the number of topics found via topic modeling. Indeed, the proactivity graphs prompted 15 different topic in insights and 13 in actions, whereas the medians were 2 and 10 respectively. Interestingly, while much more topics could be found in the analysis of the regularity graphs, (20/19 vs 8 median), they had a smaller variety of insights than all the other dimensions (4/3 topics vs 10 median). This illustrates that some SRL dimensions are more useful and actionable than others to educators. In particular, while participants seemed to grasp quickly the concept of learning behaviors, some reported confusion regarding consistency as relative time online and mentioned effort as average time online to be more useful.

Student profiles: Regarding the profiles, the most effective visualization of their proportion at one point in time was overwhelmingly chosen as a pie chart, which participants thoughts were simple and easy to read. As one comment said, "[pie charts] is the form I am most familiar with hence find it easiest to interpret". The network visualization of the profiles with similarity shown as links prompted mixed responses with some participants finding it would give useful additional information, while other felt it was confusing. Noteworthily, the participants of *older age* (older than the 75^{th} age percentile of 46 years) ranked the stacked area chart and the bubble chart equally ($\mu_{rank} = 1.6$), one of them noting that 'the bubble charts were easier to read and understand the focus'. To visualize the change in profiles proportion throughout the semester, we compared a stacked area chart ($\mu_{rank} = 1.5$), a bubble chart ($\mu_{rank} = 1.9$) and a Sankey flow chart ($\mu_{rank} = 2.6$). Respondents were more divided, but the graph participants felt was the most effective was the area graph. Comments mentioned it was better to compare through time and easier to read, while the flow chart was characterized as 'messy' and 'confusing'.

In summary, educators tend to value clarity over complexity of insights. Indeed, they prefer simpler and more familiar visualizations, like bar plots and pie charts over, despite getting more insights from heat maps. We also found some differences in design preferences depending on age and teaching level. The results of the study analysis allowed us to refine our visualizations design and the next step was to insert them in a usable interface that could be tested on potential users.

3.3 Interface Design

Our prototype's user interface was entirely designed from scratch based on ideas suggested during the interviews and website standards. The design went through a few iterations to decide how the interface would be structured and what amount of information would need to be displayed. The implementation of the interface (frontend) was done by an external programmer with React [8], while the content retrieval part (backend) was implemented by a lab member with FastAPI [9], using a Postgres database. Both applications were hosted on Heroku [10] using *Performance Dynos.*

Figure 6 shows the final design of the user interface for the dashboard prototype. The design includes a menu on the left with two parts, the overview (with the *Summary* and the *Student Profiles*) and the student behaviors (including *Proactivity*, *Effort*, *Consistency*, *Control* and *Regularity*). The user can select the course and the date range in the navigation panel. In addition, there are help buttons throughout the dashboard to clarify the different elements and to provide further explanations and additional information. In the *Summary* page (shown on Figure 6), weekly statistics per dimension are displayed as well as the trend in comparison to the previous week. The profiles proportion, characteristics and the associated grades are shown on the *Profiles* page. Moreover, as can be seen in Figure 7, each of the *Student Behavior* pages displays a description, relevant insights and statistics for two features of the dimension in the form of graphs, with the possibility to switch between statistics for all students and clustered by groups of similar behaviors for this dimension via the group tab (F5). In the prototype, the insights and statistics are expert-made, but could potentially be AI-generated in a further version of the dashboard.

In order to make `FlippED` accessible to people with disabilities, we followed the guidelines from the *Web Accessibility Initiative (WAI)* [6] and achieved a conformance level AAA (highest

[8] https://reactjs.org
[9] https://fastapi.tiangolo.com
[10] https://www.heroku.com

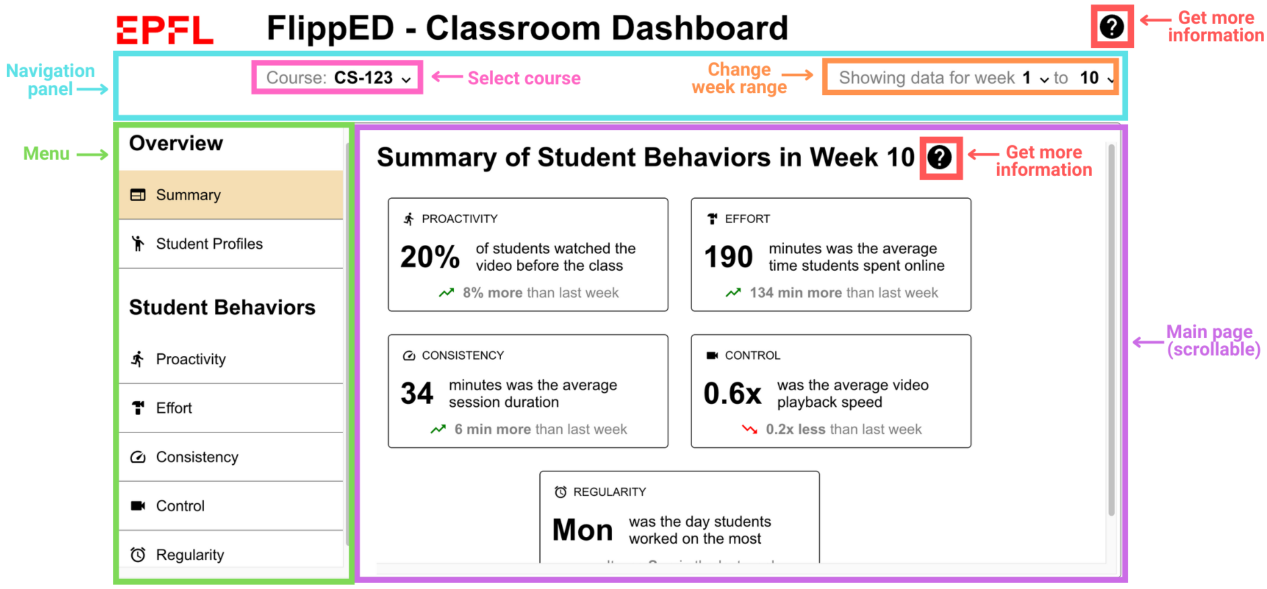

Figure 6: Screenshot of the dashboard interface with main parts highlighted. At the top is the navigation panel with course and timeframe selection. The menu showing all available pages is on the right. The main part displays the summary with statistics for all SRL dimensions for the current week. Help buttons are available to get more information at the very top and on the main page.

level)[11]. Accessibility adaptations were made in different parts of the dashboard prototype. In the visualizations, a colorblind accessible palette was used and sufficient contrast was ensured for greyscale users.

In order to be usable by users with a screen reader, all visualizations were given a detailed alt-text, describing the graphs for visually-impaired people and all UI blocks were cleary named in the code. Captions were also added for all visualizations, describing key insights from the graphs to ensure an easy understanding even for people who lacked graph literacy. In the prototype, the alt-text and captions are hand-written using alt-text standards [18], but automating them should be a goal for future research.

Regarding the general interface design, a clear navigation mechanism was ensured for keyboard navigators and support for dark mode was added. A general help button was also added, containing a description of each part of the interface.

3.4 User Testing: Dashboard usability, actionability and concerns

Following the development of a dashboard prototype, we tested our design via potential user tests in the form of semi-structured interviews. The goal of these user testings was threefold. First, we wanted to evaluate the general usability and intuitiveness of our design. Second, to assess whether the data displayed could be used by educators in a concrete and useful way. Finally, we wanted to investigate concerns educators might have when using this type of educational tools presenting machine learning results.

[11] The complete accessibility evaluation report can be found in the supplementary materials.

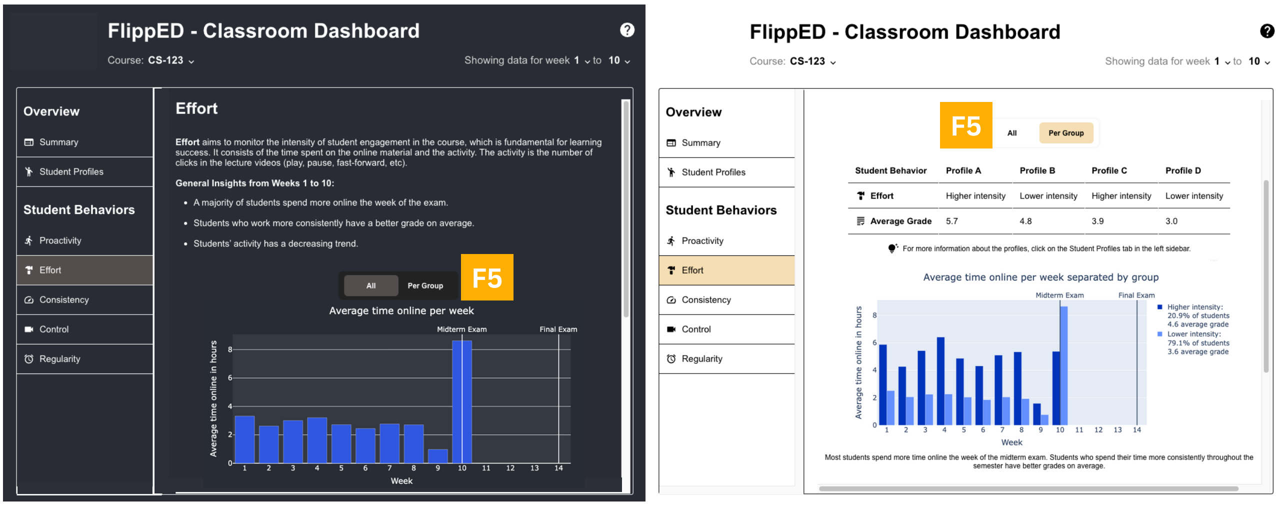

Figure 7: Screenshot of the dashboard interface showing the *Effort* page. Left is for all students and right is by groups of similar behaviors. On the main page, a short description of the dimensions is displayed with general insights on the trends. On the group page, the group assignment per profile is shown with the associated grades.

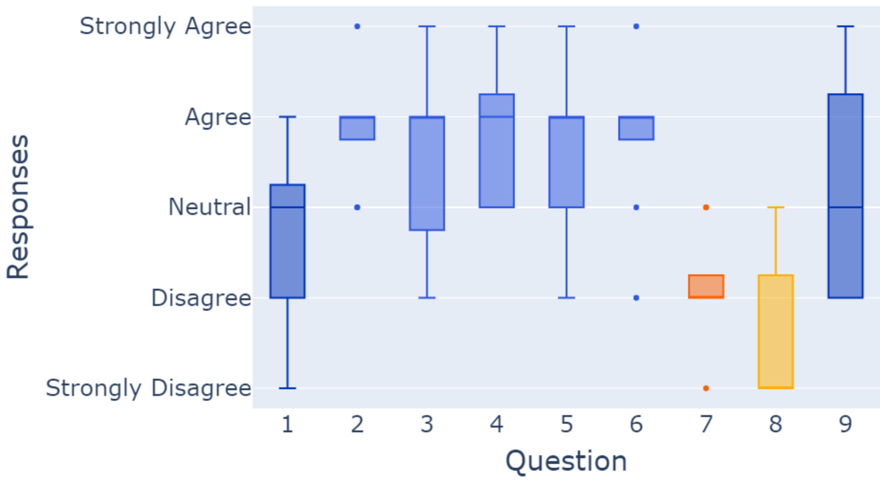

Figure 8: Responses to the Trust in AI questionnaire. The color is linked to the median. For each question, the median, interquartile range and outliers are shown on a range from strongly agree to strongly disagree.

General: 10 university-level teachers from the École Polytechnique Fédérale de Lausanne or EPFL, all with experience in blended learning, 6 with experience in flipped classrooms were presented with our prototype. The participants were generally very positive to our prototype and some mentioned being interested in getting updates on the project. Only one participant was not interested in the concept of teacher dashboard showing student behaviors. They said they did not see any information on the tool that they could not find via student surveys or direct interaction with them. They were also concerned that the teachers using this data would not take context into account and would try to make all students average.

Trust in AI: The aggregated answers to the questionnaire are shown in Figure 8. For most questions, user study participants seemed to agree on the answer, which could reflect their common position as teachers in a university-level engineering school. All participants agreed that AI can help teachers with personalization and activities management (Q2 and 4 in Fig 8). In addition, all but one think they would actually use AI personalization tools and would be successful in using them (Q5 and 6). On the other hand, most participants agree that they don't fully trust AI-based personalization tools (Q1). All participants disagree that AI-based technologies remove autonomy and control from teachers (Q7) and disagree strongly that fewer teachers would be required when AI becomes more prevalent (Q8).

Responses were more mixed considering privacy (Q9). While most participants disagree that the use of data in AI-based technologies compromises teachers' and students' privacy, 3 participants agreed with that statement, 1 strongly. Most participants agreed that AI could assist teachers in planning lessons and activities (Q3). However, 2 disagreed.

Exploration: During the exploration part, participants were all curious about the Summary data and 7 out of 10 said they enjoyed having a summary. As someone said, "Summary is very clear and summary is great because I really quickly see the [...] key parameters". Participants tended to be confused at first about the profiles, the groups and the link between them. They could not identify the (?) buttons, see Figure 6, as containing more information or were not interested in trying to get help. A better balance could be found in the future between complete explanations of the data displayed, leading to a cluttered, overwhelming dashboard filled with text and a minimalist clear design that leads to confusion about the data displayed. However, after understanding the profiles, half of the users said they would go back to the profiles regularly and use them to advise students on the best learning strategies. Participants were interested in all the learning dimensions except for consistency, which was found to be confusing or redundant with effort. When asked whether they would use any of them regularly, participants mentioned Effort and Proactivity the most. They were divided regarding the Regularity information. Most of the users said that Regularity throughout the week was about personal habits and fixed schedules and thus, would not be actionable by the teaching team. However, 4 users said they would use Regularity, for example to change when they made material available or *a posteriori* to prepare for the next year. Control data was divided as well, with some participants saying it would help them know when students had problems and adapt the videos, while others said it would not be actionable.

Actionability: After the exploration, participants were asked about how they would use the information in the dashboard. 8 out of 10 participants answered they would show part of the data directly to students. In particular, they mentioned adding one of the Proactivity graph to the course slides, showing the link between profile and grade to encourage productive behaviors or simply mentioning an issue they found at the beginning of the class. As one participant illustrated, "I would also use this information as a feedback to students on their working habits. [...] I would [show them this information] and say: 'Look, the success correlates. Why don't you try to change this?'"

7 out of 10 participants also mentioned adapting the course in some way, some examples were; changing the workload, adding activities, such as quizzes, proposing additional materials or even switching to classical teaching if students don't seem to adapt well to the flipped format.

Other possible actions teachers mentioned were sending messages to all students with some recommendations, sending automatic reminders to students who are not working regularly and advising students in the worst profile to change their learning strategies.

Concerns: The last main goal of the user testings was to assess teachers concerns with using our tool. The main concern, mentioned by more than half the participants, was privacy and the use of anonymized data. This result was expected as it was also the most required feature during the initial interviews and was mentioned in the literature on teacher dashboards, see section 2.2. Another main concern was that certain behaviors, deemed 'better', would be enforced in students and that students who don't comply with the 'correct' behavior would be penalized. Other concerns mentioned were the loss of interaction with the students and the risk of adding cognitive load to teachers.

A few participants also had concerns relating to the data, one worrying the data might not be representative of actual learning behaviors, while another thinking video clicks might not be a good proxy for activity. Other users also mentioned that time online does not reflect accurately the total time taken to study for the course and that group work was not reflected in the data displayed. One teacher in particular found it hard to trust the data displayed as they did know whether it was of good quality and reliable. Another warned us to be mindful of the metrics used as "the choices that you make in terms of saying this is the data that's relevant, [...] these are the metrics that matter. So, it's really important, as much as possible, to measure what you value, because what you measure ends up being what gets valued".

When exploring the prototype, participants mentioned potential features they would have found useful or interesting. Some of them could be used to refine the design and improve the dashboard, while others fell entirely outside of the scope of this research. However, we felt that some of these finding might be useful for future research and they are discussed in section 4.

Click-stream: The participants' exploration click-stream were analyzed by a lab member to study how much time was spent on each page and how the teachers navigated the dashboard. In Figure 9, each circle is a page and the area is time spent on each page in logarithmic scale. The dark circle is the mean time and the light circle is the average time plus the standard deviation. Participants spent the greatest time on the profiles page (10 minutes); half of the participants were at first confused about the profiles but then said they would go back to the profiles page regularly and use them to advise students on the best learning strategies. Moreover, they spent the least amount of time on the consistency groups page (2.3 minutes); they found the consistency page redundant with effort.

In addition to time spent, the transition probabilities are represented with the width and the transparency of the arrows in a linear scale. We only plotted the edges with a transition probability greater than 0.12 to visualize the most frequent transitions. The highest transition probability in Figure 9 is from *Control groups* to *Regularity* ($p = 1$) and the lowest is from *Effort* to *Effort groups* ($p = 0.14$). What can be inferred from these probabilities is that, after exploring the summary page, participants went to the profiles or proactivity pages. Then, participants went back and forth to understand the layout of the dashboard and the profiles. Once it was clear, they followed mostly an ordered exploration strategy, accessing the learning dimensions pages following the sequential order of the menu.

In the second task, when asked to use the dashboard to identify possible causes for poor

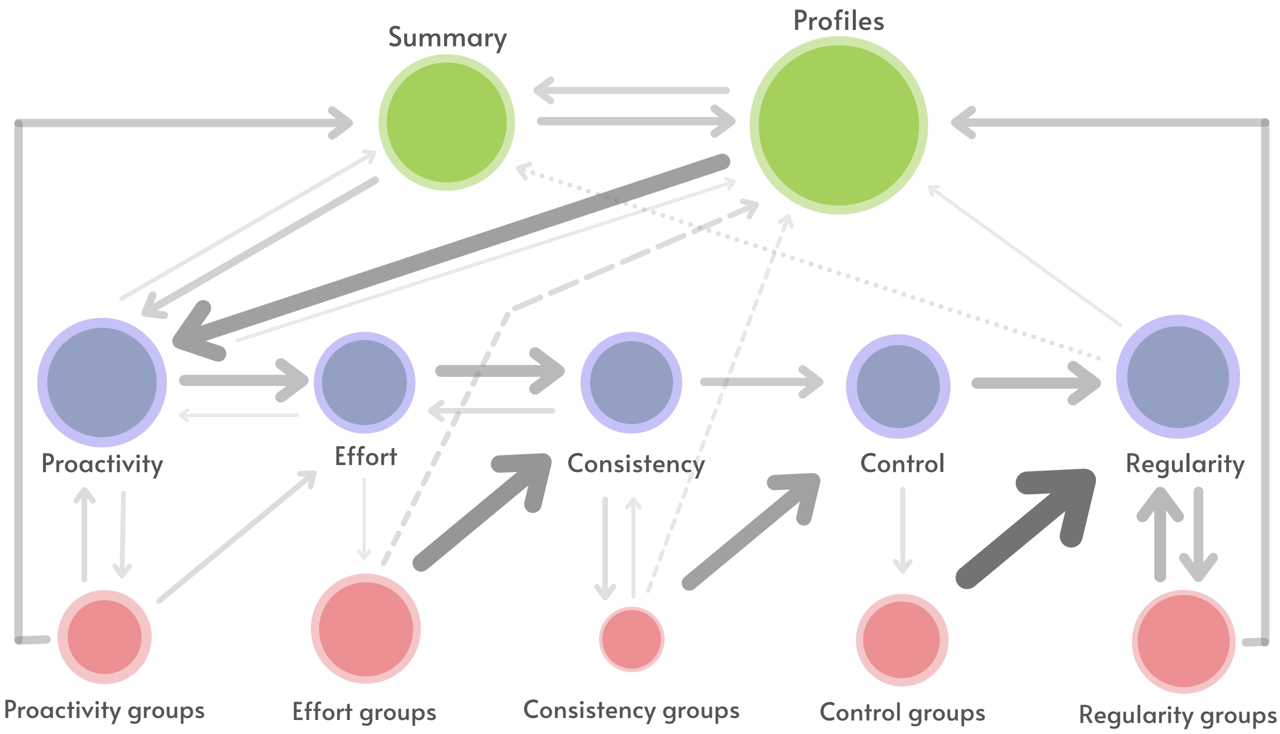

Figure 9: Click-stream of FlippED. Each circle represents a page with size logarithmically proportional to the average time spent on it. Transition probabilities are represented with the width and the transparency of the arrows.

midterm performance, 80% of the participants went to the profiles page and the remaining 20% to the summary page first and then to the profiles page. From there, participants described the qualities of the poor-performing profiles. Overall, participants knew where to find the information. However, some mentioned that there was a lot of information and would most likely just check the Summary and the Profiles page on a regular basis. On the other hand, they also mentioned being interested in the other pages if they identified some problem in the overview.

Summary: The general response to the dashboard prototype was very positive. Despite being confusing at first, the profiles page prompted many reflections from the users and was the main source of inspiration when asked about course adaptations and possible actions. The summary page was also very appreciated as a way to get a general view of the key features. While participants mentioned not seeing themselves use all the learning dimensions pages regularly, they all mentioned some dimensions they would be most interested in or wanting to have the information in case a problem was detected. The main concerns, except anonymity, related to the way the tool would be used by teachers and the data. While these aspects are mostly out of hand of the developing team, some thought will have to be put into minimizing the potential issues mentioned during the user testing.

4 Discussion

4.1 Discussion, Limitations and Future Work

This project studied teachers' student behaviors data needs in flipped classroom contexts, as well as their visual design preferences. The findings of these studies where then compiled to design a teacher dashboard presenting SRL student behaviors. Unlike prior work, the dashboard incorporated a ML-based clustering method to show patterns of SRL behaviors as student profiles. Finally, the prototype was tested with potential users to study interaction, actionability and concerns.

The results of the user interviews showed that educators are interested in learning more about their students' SRL behaviors, but are concerned about infringing on their privacy. They mentioned different data on their students behaviors they would be interested in seeing, either regularly or at the end of the semester. They were also really interested in the learning behaviors profiles and their characteristics, mentioning they would use this information to improve the class or help struggling students.

Complementing the existing research on visualizations design, the user study was able to examine more closely educators visual design preferences and assess differences in preferences regarding personal and demographic factors. We found some statistically significant differences in preferences such as university-level teachers ranking plots showing a measure of variance higher. Regarding general preferences, we found a strong tendency to find simpler graph types, such as line plots and bar graphs, easier to understand and use, despite them not showing a notion of variance and hiding outliers. This may come from educators generally not being exposed to more complex graph types, such as heat maps or box plots. Study participants also mentioned being overwhelmed when too much information was shown at once. Thus, in the dashboard prototype, only bar charts and an area graph were displayed. We acknowledge that this type of visualizations can be misleading as they show only mean values and hide variance and outliers. In future work, this could be addressed by having a clickable option to show the data distribution on top of the averaged graphs. In addition, previous work had only outlined the trade-offs between comparative designs [11] and we find that there were significant preferences for side-by-side superposition graphs over juxtaposition and explicit encoding of relationships.

The final step was testing the dashboard prototype and assessing potential uses and concerns. As described in section 3.4, participants were most interested in the profiles page. They mentioned learning dimensions they would examine regularly and others they would not find useful. Interestingly, while their answers often regrouped, we could see some divergence in the participants dimensions interest. This could be addressed by having a flexible interface, where users could decide what pages or features to hide/show.

In regards to dashboard uses and prompted actions, participants mostly went back to the summary and profiles page to assess potential problems. Most participants mentioned showing the data to their students directly to encourage productive behaviors. Possible actions also included adapting the course and messaging students directly. In future work, features could be added to the dashboard to export specific data from the tool and automatically sending emails to all, one group or one profile of student. This last point would have to be designed with no way of tracing back to the students identity in order to keep the tool fully anonymized.

The privacy concern emerged at all stages of the project, literature review, user interviews and user testings. This shows that the ethics of data privacy is a subject that educators and researchers seem to have well in mind. While all the data in the dashboard prototype was aggregated and no individual information was shown, more work needs to be done to ensure that no personal data can be extracted from the tool. Future work also needs to study possible adverse effects of the tool to minimize potential issues, as mentioned in the concerns part of section 3.4.

During the user interviews and testing, participants mentioned suggestions that were out of the scope of this project but could be interesting for future research on the subject. A feature requested by many users during the interviews was statistics per video. For example, users indicated being interested in seeing the re-watch and pauses on each video individually. Other common suggestion was to display students' weekly anonymous feedback on the dashboard, as well as student's use of secondary resources. A few participants were also interested in seeing the answers to quizzes or exercises, as well as the time taken to do them, as a way to pinpoint concepts students may have difficulties with.

Despite asking for volunteers on LinkedIn through multiple accounts, we were unsuccessful in recruiting experienced teachers from distinct universities to evaluate our design. Thus, future work should study the generalizability of our findings in different regions and cultures. Moreover, while we show a promising short-term evaluation of our design, the design and evaluation process should be iterated with our findings and be evaluated in classrooms for longer periods.

5 Summary

5.1 English

This project aimed at co-designing a teacher dashboard for flipped classroom context that displays data on their students self-regulated learning behaviors. Teachers' data needs in terms of their students behaviors and their visualization design preferences were assessed via user interviews and a visualization study. The results of the interviews and study were used to design a dashboard coherent with the context. A ML-based clustering method to show patterns of SRL behaviors as student profiles was also incorporated in the tool after confirming the interest of potential users. The dashboard prototype was tested with potential users to study interaction, actionability and concerns. The interviews and the study were analyzed in a qualitative and quantitative way, using state-of-the-art statistical and natural language processing techniques.

The preliminary interviews showed a real interest in student's behavioral data from teachers. They informed the design of a dashboard prototype comprising of a summary, student profiles and one page per SRL dimension. The visualizations displayed in the dashboard were designed following the result of the user study. Study participants showed a real preference for simpler graphs and clear layouts. The user tests demonstrated the usability and actionability of the dashboard. The layout was easy to navigate and the data displayed prompted diverse insights and actions ideas from participants. Though not all of the SRL dimension pages were deemed useful to consult regularly, participants generally agreed that the summary and profiles would help them identify problem, advise students on their learning strategies and adapt the course.

To conclude, the participatory approach allowed to assess interest and usability at all steps of the process, ensuring the relevance of the tool to a flipped classrooms context. The results of the study show a real interest from teachers in student SRL behavior data and, in particular, in SRL profiles. Building on this project, a teacher dashboard could be tested and implemented in real flipped classes at university-level, helping teachers understand their students behaviors and using these insights to adapt their course and prevent unproductive behaviors.

5.2 French

Le but de ce projet était le co-design d'un tableau de bord pour enseignants dans le contexte de classes inversées présentant des données sur les comportements d'apprentissage autorégulés (SRL) de leurs élèves. L'intérêt des enseignant sur les comportements de leurs élèves et leurs préférences en matière de design de visualisations ont été évalués au moyen d'entretiens et d'une étude utilisateur sur les visualisations. Les résultats des entretiens et de l'étude ont été utilisés pour concevoir un tableau de bord cohérent avec le contexte. Une méthode de clustering basée sur le machine learning révélant des tendances de comportements de SRL sous forme de profils d'apprentissage a également été incorporée dans l'outil après avoir confirmé l'intérêt des utilisateurs potentiels. Le prototype du tableau de bord a été testé avec des utilisateurs potentiels pour étudier l'interaction, l'utilité et les inquiétudes. Les entretiens et l'étude ont été analysés de manière qualitative et quantitative, en utilisant des techniques de pointe de statistique et d'analyse de texte.

Les entretiens préliminaires ont montré un réel intérêt pour les données comportementales des élèves de la part des enseignants. Ils ont orienté la conception d'un prototype de tableau de bord comprenant un résumé, des profils d'apprentissage et une page par dimension de SRL. Les visualisations affichées dans le tableau de bord ont été conçues en fonction du résultat de l'étude utilisateur. Les participants à l'étude ont montré une réelle préférence pour des graphiques plus simples et des mises en page claires. Les tests utilisateurs ont démontré la facilité d'utilisation et l'exploitabilité du tableau de bord. La mise en page était facile à naviguer et les données affichées ont suscité diverses idées et actions de la part des participants. Bien que les pages de dimensions de SRL n'aient pas toutes été jugées utiles à consulter régulièrement, les participants ont généralement convenu que le résumé et les profils les aideraient à cerner les problèmes, à conseiller les étudiants sur leurs stratégies d'apprentissage et à adapter le cours.

Pour conclure, l'approche participative a permis d'évaluer l'intérêt et l'exploitabilité à toutes les étapes du processus, assurant ainsi la pertinence de l'outil dans un contexte de classes inversées. Les résultats de l'étude montrent un réel intérêt des enseignants pour les données comportementales de SRL des élèves et, en particulier, pour les profils de SRL. S'appuyant sur ce projet, un tableau de bord pour enseignants pourrait être testé et mis en œuvre dans de véritables classes inversées au niveau universitaire, aidant les enseignants à comprendre les comportements de leurs élèves et à utiliser ces informations pour adapter leur cours et prévenir les comportements improductifs.

References

[1] Noor Latiffah Adam et al. "Self-Regulated Learning and Online Learning: A Systematic Review". In: *Advances in Visual Informatics*. Ed. by Halimah Badioze Zaman et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017, pp. 143–154. ISBN: 978-3-319-70010-6. DOI: 10.1007/978-3-319-70010-6_14.

[2] Danielle Albers, Michael Correll, and Michael Gleicher. "Task-driven evaluation of aggregation in time series visualization". In: *CHI '14*. 2014, pp. 551–560.

[3] Sakinah Alhadad. "Visualizing Data to Support Judgement, Inference, and Decision Making in Learning Analytics: Insights from Cognitive Psychology and Visualization Science". In: *Journal of Learning Analytics* 5 (Aug. 5, 2018), pp. 60–85. DOI: 10.18608/jla.2018.52.5.

[4] Lucy Barnard-Brak, William Lan, and Valerie Paton. "Profiles in Self-Regulated Learning in the Online Learning Environment". In: *International Review of Research in Open and Distance Learning* 11 (Mar. 1, 2010). DOI: 10.19173/irrodl.v11i1.769.

[5] Moon-Heum Cho and Demei Shen. "Self-regulation in online learning". In: *Distance Education* 34.3 (Nov. 1, 2013), pp. 290–301. ISSN: 0158-7919. DOI: 10.1080/01587919.2013.835770. URL: https://doi.org/10.1080/01587919.2013.835770 (visited on 01/10/2023).

[6] World Wide Web Consortium et al. "Web Accessibility Initiative (WAI)". In: (1998). URL: http://www.w3.org/WAI/.

[7] Prajakta Diwanji, Knut Hinkelmann, and Hansfriedrich Witschel. "Enhance Classroom Preparation for Flipped Classroom using AI and Analytics:" in: *Proceedings of the 20th International Conference on Enterprise Information Systems*. 20th International Conference on Enterprise Information Systems. Funchal, Madeira, Portugal: SCITEPRESS - Science and Technology Publications, 2018, pp. 477–483. ISBN: 978-989-758-298-1. DOI: 10.5220/0006807604770483. URL: http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0006807604770483 (visited on 10/05/2022).

[8] Amr Elsayed et al. "Research in Learning Analytics and Educational Data Mining to Measure Self-Regulated Learning: A Systematic Review". In: Sept. 25, 2019.

[9] Stephanie Evergreen and Chris Metzner. "Design Principles for Data Visualization in Evaluation". In: *New Directions for Evaluation* 2013.140 (2013), pp. 5–20. ISSN: 1534-875X. DOI: 10.1002/ev.20071. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ev.20071 (visited on 10/26/2022).

[10] Michail N. Giannakos, Konstantinos Chorianopoulos, and Nikos Chrisochoides. "Making Sense of Video Analytics: Lessons Learned from Clickstream Interactions, Attitudes, and Learning Outcome in a Video-Assisted Course". In: *International Review of Research in Open and Distributed Learning* 16.1 (Feb. 2015). Publisher: Athabasca University ERIC Number: EJ1061073, pp. 260–283. URL: https://eric.ed.gov/?id=EJ1061073 (visited on 01/10/2023).

[11] Michael Gleicher et al. "Visual comparison for information visualization". In: *Information Visualization* 10.4 (2011), pp. 289–309.

[12] Cécile Hardebolle et al. “Gender, prior knowledge, and the impact of a flipped linear algebra course for engineers over multiple years”. In: *J. Eng. Educ.* 111.3 (2022), pp. 554–574.

[13] Kenneth Holstein, Bruce M. McLaren, and Vincent Aleven. “Co-Designing a Real-Time Classroom Orchestration Tool to Support Teacher–AI Complementarity”. In: *Journal of Learning Analytics* 6.2 (July 22, 2019). ISSN: 1929-7750. DOI: 10.18608/jla.2019.62.3. URL: https://learning-analytics.info/index.php/JLA/article/view/6336 (visited on 09/20/2022).

[14] Jelena Jovanović et al. “Learning analytics to unveil learning strategies in a flipped classroom”. In: *The Internet and Higher Education* 33 (Apr. 2017), pp. 74–85. ISSN: 10967516. DOI: 10.1016/j.iheduc.2017.02.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S1096751617300684 (visited on 10/05/2022).

[15] Rogers Kaliisa et al. “Teachers’ Perspectives on the Promises, Needs and Challenges of Learning Analytics Dashboards: Insights from Institutions Offering Blended and Distance Learning”. In: Dec. 17, 2021, pp. 351–370. ISBN: 978-3-030-81221-8. DOI: 10.1007/978-3-030-81222-5_16.

[16] A. Kavcic. “Software Accessibility: Recommendations and Guidelines”. In: *EUROCON 2005 - The International Conference on ”Computer as a Tool”*. Vol. 2. 2005, pp. 1024–1027. DOI: 10.1109/EURCON.2005.1630123.

[17] Sukwon Lee, Sung-Hee Kim, and Bum Chul Kwon. “Vlat: Development of a visualization literacy assessment test”. In: *IEEE transactions on visualization and computer graphics* 23.1 (2016), pp. 551–560.

[18] Alan Lundgard and Arvind Satyanarayan. “Accessible visualization via natural language descriptions: A four-level model of semantic content”. In: *IEEE transactions on visualization and computer graphics* 28.1 (2021), pp. 1073–1083.

[19] Paola Mejia-Domenzain et al. “Evolutionary Clustering of Apprentices’ Self- Regulated Learning Behavior in Learning Journals”. In: *IEEE Transactions on Learning Technologies* (2022). Conference Name: IEEE Transactions on Learning Technologies, pp. 1–14. ISSN: 1939-1382. DOI: 10.1109/TLT.2022.3195881.

[20] Paola Mejia-Domenzain et al. “Identifying and Comparing Multi-dimensional Student Profiles Across Flipped Classrooms”. In: *Artificial Intelligence in Education*. Ed. by Maria Mercedes Rodrigo et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2022, pp. 90–102. ISBN: 978-3-031-11644-5. DOI: 10.1007/978-3-031-11644-5_8.

[21] Stephen R. Midway. “Principles of Effective Data Visualization”. In: *Patterns* 1.9 (Dec. 11, 2020), p. 100141. ISSN: 2666-3899. DOI: 10.1016/j.patter.2020.100141. URL: https://www.sciencedirect.com/science/article/pii/S2666389920301896 (visited on 10/26/2022).

[22] Tanya Nazaretsky, Mutlu Cukurova, and Giora Alexandron. "An Instrument for Measuring Teachers' Trust in AI-Based Educational Technology". In: *LAK22: 12th International Learning Analytics and Knowledge Conference*. LAK22. New York, NY, USA: Association for Computing Machinery, Mar. 21, 2022, pp. 56–66. ISBN: 978-1-4503-9573-1. DOI: 10.1145/3506860.3506866. URL: https://doi.org/10.1145/3506860.3506866 (visited on 11/11/2022).

[23] Tanya Nazaretsky et al. "Empowering Teachers with AI: Co-Designing a Learning Analytics Tool for Personalized Instruction in the Science Classroom". In: *LAK22: 12th International Learning Analytics and Knowledge Conference*. LAK22: 12th International Learning Analytics and Knowledge Conference. Online USA: ACM, Mar. 21, 2022, pp. 1–12. ISBN: 978-1-4503-9573-1. DOI: 10.1145/3506860.3506861. URL: https://dl.acm.org/doi/10.1145/3506860.3506861 (visited on 09/20/2022).

[24] Mar Pérez-Sanagustín et al. "Designing a Moodle Plugin for Promoting Learners' Self-regulated Learning in Blended Learning". In: *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption*. Ed. by Isabel Hilliger et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2022, pp. 324–339. ISBN: 978-3-031-16290-9. DOI: 10.1007/978-3-031-16290-9_24.

[25] Rim Razzouk and Valerie Shute. "What Is Design Thinking and Why Is It Important?" In: *Review of Educational Research* 82.3 (Sept. 2012), pp. 330–348. ISSN: 0034-6543, 1935-1046. DOI: 10.3102/0034654312457429. URL: http://journals.sagepub.com/doi/10.3102/0034654312457429 (visited on 10/05/2022).

[26] Carolyn P. Rosé et al. "Explanatory learner models: Why machine learning (alone) is not the answer". In: *British Journal of Educational Technology* 50.6 (2019). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/bjet.12858, pp. 2943–2958. ISSN: 1467-8535. DOI: 10.1111/bjet.12858. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/bjet.12858 (visited on 11/23/2022).

[27] Shiva Shabaninejad et al. "Recommending insightful drill-downs based on learning processes for learning analytics dashboards". In: *AIED*. 2020.

[28] Varshita Sher, Marek Hatala, and Dragan Gašević. "Analyzing the consistency in within-activity learning patterns in blended learning". In: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. LAK '20. New York, NY, USA: Association for Computing Machinery, Mar. 23, 2020, pp. 1–10. ISBN: 978-1-4503-7712-6. DOI: 10.1145/3375462.3375470. URL: https://doi.org/10.1145/3375462.3375470 (visited on 01/10/2023).

[29] Mina Shirvani Boroujeni et al. "How to quantify student's regularity?" In: Sept. 13, 2016.

[30] "The use of flipped classrooms in higher education: A scoping review". In: *The Internet and Higher Education* 25 (Apr. 1, 2015). Publisher: JAI, pp. 85–95. ISSN: 1096-7516. DOI: 10.1016/j.iheduc.2015.02.002. URL: https://www.sciencedirect.com/science/article/pii/S1096751615000056 (visited on 09/30/2022).

[31] Jacqueline Wong et al. "Supporting Self-Regulated Learning in Online Learning Environments and MOOCs: A Systematic Review". In: *International Journal of Human–Computer Interaction* 35.4 (Mar. 16, 2019), pp. 356–373. ISSN: 1044-7318. DOI: 10.1080/10447318.2018.1543084. URL: https://doi.org/10.1080/10447318.2018.1543084 (visited on 09/20/2022).

[32] Françeska Xhakaj, Vincent Aleven, and Bruce M. McLaren. "How Teachers Use Data to Help Students Learn: Contextual Inquiry for the Design of a Dashboard". In: *Adaptive and Adaptable Learning*. Ed. by Katrien Verbert, Mike Sharples, and Tomaž Klobučar. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016, pp. 340–354. ISBN: 978-3-319-45153-4. DOI: 10.1007/978-3-319-45153-4_26.

[33] Meehyun Yoon, Janette Hill, and Dongho Kim. "Designing supports for promoting self-regulated learning in the flipped classroom". In: *Journal of Computing in Higher Education* 33.2 (Aug. 1, 2021), pp. 398–418. ISSN: 1867-1233. DOI: 10.1007/s12528-021-09269-z. URL: https://doi.org/10.1007/s12528-021-09269-z (visited on 10/05/2022).

[34] Barry Zimmerman. "Becoming a Self-Regulated Learner: An Overview". In: *Theory Into Practice* 41 (June 1, 2002), pp. 64–70. DOI: 10.1207/s15430421tip4102_2.

A Supplementary materials

List of supplementary materials available at
https://anonymous.4open.science/r/flipped

- User Interview Protocol
- Accessibility Report
- User Study Questionnaire
- Code to make dashboard visualizations
- All the visualizations tested in the study
- User Testing Protocol
- Trust in AI Questionnaire

Also available are the code for the frontend (dashboard app) and the backend (database).