Designing Actionable and Interpretable Analytics Indicators for Improving Feedback in AI-Based Systems

Esther Félix[1][a], Elaine De Oliveira[2][b], Ilmara M. M. Ramos[2], Mar Perez-Sanagustin[1][c], Esteban Villalobos[1][d], Isabel Hilliger[3][e], Rafael Ferreira Mello[4][f] and Julien Broisin[1][g]

[1] *Universite de Toulouse - IRIT - CNRS, France*

[2] *Universidade Federal do Amazonas, Brazil*

[3] *Pontificia Universidad Católica de Chile, Chile*

[4] *CESAR, Brazil*

{esther.felix, mar.perez-sanagustin, esteban.villalobos, julien.broisin}@irit.fr,

{elaine frpe.br

Keywords: Feedback, Learning Analytics, Card Sorting, Human-Computer Interaction, User-Centered Approach.

Abstract: In AI-based educational systems, transparency and understandability are particularly important to ensure reliable human-AI interaction. This paper contributes to the ongoing research on developing analytics for AI-based educational systems by delivering feedback throughout indicators that learners can easily interpret and act upon during their studies. Specifically, this paper introduces a mixed methods study that examines the types of indicators that ought to be incorporated into the feedback offered by an AI-based system designed to help students develop competencies in programming. Building upon prior work in Human-Centered Design, the card sorting technique was used to collect both qualitative and quantitative data from 31 Computer Science students. We created 16 cards that presented students with different indicators to explain the reasoning behind the system's decisions and feedback. These indicators were displayed in different formats (visual and textual representations; temporal vs. non-temporal and social vs. non-social reference frames). Our goal was to discover the most interpretable and actionable method for delivering feedback to learners. Our study found low consensus among students. Overall, students found indicators based on social comparison to be less actionable and interpretable compared to those without; and textual indicators were perceived as less actionable and interpretable than visual ones.

1 INTRODUCTION

1.1 Designing Effective Feedback for Learning Analytics in AI-Based Educational Systems

The growing availability of educational data has led to increasingly complex statistical models, promising significant improvements in learning. However, their effectiveness relies on how clearly and actionable their outputs are for students. In Learning Analytics, a key challenge is designing feedback indicators that are both interpretable and actionable, enabling students to adjust their behaviors effectively to improve learning outcomes (Álvarez et al., 2022).

These goals align with the field of Explainable Artificial Intelligence (XAI), which focuses on fostering user trust and understanding in educational AI systems (Khosravi et al., 2022). Specifically, within the technology-enhanced learning (TEL) community, explainability is seen as a critical area for exploration and development of solutions aimed at enhancing the transparency of AI-based educational systems, thus fostering transparent and trustworthy interactions between humans and AI in learning environments (Sharples, 2023). However, beyond explainability, it is essential that indicators are also actionable: that is, they enable students to take concrete steps to improve their learning.

[a] https://orcid.org/0009-0007-6905-8939
[b] https://orcid.org/0000-0003-2884-9359
[c] https://orcid.org/0000-0001-9854-9963
[d] https://orcid.org/0000-0002-6026-3756
[e] https://orcid.org/0000-0001-5270-7655
[f] https://orcid.org/0000-0003-3548-9670
[g] https://orcid.org/0000-0001-8713-6282

Félix, E., De Oliveira, E., Ramos, I. M. M., Perez-Sanagustin, M., Villalobos, E., Hilliger, I., Mello, R. F. and Broisin, J.
Designing Actionable and Interpretable Analytics Indicators for Improving Feedback in AI-Based Systems.
DOI: 10.5220/0013294300003932

In *Proceedings of the 17th International Conference on Computer Supported Education (CSEDU 2025) - Volume 1*, pages 428-435
ISBN: 978-989-758-746-7; ISSN: 2184-5026

1.2 Towards Actionable and Interpretable Feedback

Across previous research in TEL, incorporating explainability into Artificial Intelligence in Education (AIED) systems means offering ”actionable” explanations (Khosravi et al., 2022). Actionability refers to the extent to which information supports and encourages students in modifying their behavior effectively. Therefore, the explanations delivered to users should facilitate the initiation of corrective actions or feedback loops in response to their activities (Jørnø and Gynther, 2018). In line with this, as proposed by Winne (2021), explanations should not only clarify how results are derived, but also be coupled with feedback that motivates individuals to reflect or take action. Actionability is inevitably related to the students' capacity to interpret the indicators they are provided with, as interpretability is defined as a high understandability of the information presented.

This is were the concepts of actionability and interpretability are key for providing explainable AI-based systems for end users. In the field of Learning Analytics, research have extensively focus on dashboards to effectively transform trace data into ”actionable insights” able to change students' behaviour (Álvarez et al., 2022). Still, previous research shows that the indicators proposed in this objective of promoting actionable insight often fail to induce a change in the students' behaviour (Jørnø and Gynther, 2018; Villalobos et al., 2023), emphasizing the need to investigate how to make indicators more interpretable and actionable (van Leeuwen et al., 2022).

2 CONTEXT, OBJECTIVES AND RESEARCH QUESTION

Our study was conducted in a Brazilian university, as part of a joint research project between Brazil and France. Both French and Brazilian universities use tools for computer science students, in which AI-based feedback can be implemented. This study builds on a prior investigation conducted in a French university, where students interacted with an AI-based programming tool that used unsupervised machine learning to classify their programming behaviors into distinct profiles. These profiles were determined based on engagement and performance metrics such as code submission frequency, error rates, and time spent between submissions. Feedback provided by the system included textual explanations: (1) suggestions for improvement based on past behavior, and (2) explanations of how the system classified their profiles using the algorithm's features (Félix et al., 2022).

The findings from this prior study revealed key limitations. Students reported that the textual feedback provided was unattractive, difficult to interpret, and insufficient to build trust in the system. Several participants expressed a strong preference for graphical explanations over textual ones. These observations suggest that while students value transparency, the format and delivery of feedback play a critical role in its effectiveness.

Considering these previous results and building upon other research highlighting the role of explanations in increasing system trust (Conati et al., 2021), the present study aims to push our research forward by creating actionable and interpretable explanations to help build trust among students in our AI-based programming learning environment. We employed the card sorting technique as our approach to evaluate the design for this type of feedback, so that it can be implemented in future works in both France and Brazil. The card sorting method is commonly used in Human-Computer Interaction (HCI) to gather insights into user-centered design practices, card sorting aids in making informed decisions about designing indicators (Spencer and Garrett, 2009). It examines how individuals categorize various items, seeking to identify common patterns in their thought processes. Although this method has been previously applied with similar goals (Villalobos et al., 2023), prior works often proposed abstract indicators without practical applications.

Our main objective is to improve the feedback provided to students in our programming tool by empirically evaluating how different designs of indicators impact students' ability to engage with and interpret AI-driven feedback. Specifically, we designed a collection of indicators of diverse types—visual/textual, temporal/non-temporal, social/non-social, high-performance/low-performance. We then applied the card sorting method used in previous studies (Villalobos et al., 2023) to assess and compare the various indicator designs that could be implemented in our AI-based system for programming education. Students were asked to arrange the cards along a two-dimensional axis of interpretability and actionability. Therefore, the main research question addressed is the following: To what extent are the proposed indicator designs understandable (interpretable and actionable) depending on their type (visual/textual, temporal/non-temporal, social/non-social, high-performance/low-performance)?

3 METHODS: CARDS DESIGN, PARTICIPANTS, PROCEDURE AND DATA COLLECTION

For our study, we chose a concurrent mixed-methods approach that combines both qualitative and quantitative data. We designed and conducted our card sorting based on guidelines by Spencer and Garrett (2009), to ensure we collected clear, useful insights. The indicators and scenario selected are based on our previous study regarding explainability in a technology-enhanced environment where programming student learn coding skills (Félix et al., 2022).

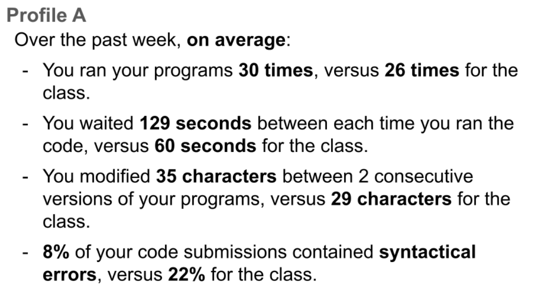
3.1 Cards Design

We developed 16 cards representing feedback indicators, classified into key categories based on prior research (Jivet et al., 2020; Molenaar and Wise, 2022; Villalobos et al., 2023). First, **temporal versus non-temporal**. As emphasized by Molenaar and Wise (2022), including temporality helps contextualize feedback. Temporal indicators provided weekly data alongside comparisons to previous weeks (represented with line graphs), while non-temporal indicators only displayed current data (using radar graphs). Textual versions included or omitted references to prior weeks accordingly. (2) Then, **social versus non-social**. Based on Jivet et al. (2020), social indicators included comparisons with the class, while non-social indicators presented individual performance only. Thirdly, following Vytasek et al. (2020), indicators were designed to reflect how feedback would appear for both **high-achieving and low-achieving** students.

Each indicator was created in both textual and visual formats, as suggested by Villalobos et al. (2023), resulting in 16 cards covering all combinations of the studied dimensions. Table 1 outlines the card classifications, and figures 1 and 2 provide examples. The full card set is available upon request.

Table 1: Types of cards designed for the study, according to the dimensions investigated (+ corresponds to high-achiever, and - to low-achiver, V to Visual, T to Text, S to Social and NS to Non-Social).

	Non Temporal				Temporal			
	S+	S-	NS+	NS-	S+	S-	NS+	NS-
V	C1	C3	C5	C7	C9	C11	C13	C15
T	C2	C4	C6	C8	C10	C12	C14	C16


Profile A

Over the past week, **on average**:

- You ran your programs **30 times**, versus **26 times** for the class.
- You waited **129 seconds** between each time you ran the code, versus **60 seconds** for the class.
- You modified **35 characters** between 2 consecutive versions of your programs, versus **29 characters** for the class.
- **8%** of your code submissions contained **syntactical errors**, versus **22%** for the class.

Figure 1: Card number 2 translated in English (C2): textual, high-achiever, social comparison, non temporal.

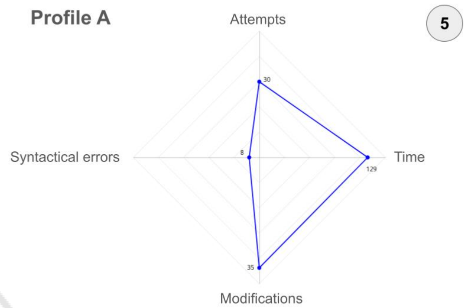


Figure 2: Card number 5 (C5): graph, high-achiever, no social comparison, non temporal.

3.2 Participants

A total of 31 computing students (18 men, 13 women) from a public Brazilian university participated in the study. This population was selected due to the programming-specific nature of the indicators. Students were recruited via email invitations, with participation being voluntary and rewarded with extra "research" credits. Participants selected convenient time slots and provided informed consent before joining the study.

3.3 Procedure and Data Collection

We conducted four one-hour workshop sessions in English and Portuguese. One researcher, fluent in English, led the session with support from a Portuguese-speaking researcher to ensure students fully understood the material and could ask questions. A third researcher was in charge of taking pictures of the students' productions. All workshop sessions were structured into different phases.

First, students were welcomed and thanked for their participation, after which the researcher briefly outlined the study's objectives and explained the workshop structure, presenting a scenario in which participants, as programming students, interacted with an AI-based platform providing feedback on their programming profiles; this introduction included

16 indicator cards alongside three explanatory cards detailing the scenario ("You are a computer science or software engineering student enrolled in a programming course. The practical work takes place over a period of several weeks on an online platform on which you are required to solve exercises and write short programs. The platform collects data which is then processed by an artificial intelligence algorithm that identifies your programming profile. This profile corresponds to the strengths and weaknesses of your programming style, and is accompanied by advice on how to improve. As researchers, we want to know what information is relevant to you, and what presentation of information is most understandable and can lead you to take action to improve your way of programming. The indicators we ask you to evaluate could be provided to students corresponding to one of two profiles"), an example of feedback for profile A ("Congratulations! Over the past week, we've seen that you've implemented some good strategies for making effective progress. In particular, it seems you took enough time to think about your programs before writing them. Keep it up!"), and an example of feedback for profile B ("We noticed that last week, you have used a "trial and error" strategy in your programming. This means that you probably submitted a lot of code with errors without taking the time to reflect on it. Try to spend more time to think about your code before running it!").

Then, each student reviewed the 16 cards over a 10-minute period, with an opportunity to ask clarifying questions, before individually ranking the cards on a two-dimensional grid, with the horizontal axis representing actionability ("Not at all actionable" to "Very actionable") and the vertical axis representing interpretability ("Not at all interpretable" to "Very interpretable"), ensuring that all participants evaluated the cards uniformly (irrespective of their own academic performance, which was not considered in the study).

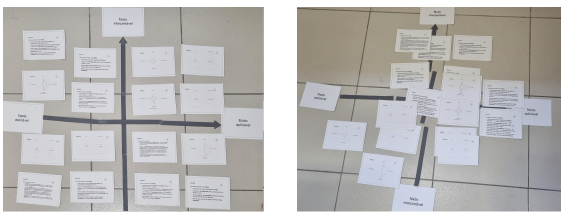


Figure 3: Examples of grids sorted by students.

Finally, after completing the card sorting, students were asked to write a paragraph explaining the rationale behind their card placements, and researchers collected both photographs of the final grids and the written justifications for further analysis.

3.4 Analytical Methods

Quantitative Analysis. Photographs of the students' grids were used to calculate two key metrics: the average ranking of different indicator types and the level of agreement between students. Each card was assigned a score from 1 (lowest preference) to 16 (highest preference) based on its placement on the grid. The rank scores allowed for statistical analysis, with the Kendall Tau-b correlation coefficient employed to measure agreement levels for interpretability and actionability across all pairs of students. The coefficient values ranged from -1 (complete disagreement) to 1 (complete agreement).

Qualitative Analysis. Following the guidelines of Nowell et al. (2017), the rationale paragraphs written by students were analyzed using a six-step process. First, the texts were translated from Portuguese to English, enabling all researchers to review them. Initial codes were then generated based on the design dimensions of the cards (text vs. visual, high-performance vs. low-performance, social vs. non-social, and temporal vs. non-temporal). Researchers collaborated to identify three main themes: actionability, interpretability, and preferred indicators. Relevant quotes were selected to support the themes, and the frequency of codes and themes in the students' responses was quantified to provide further insights.

4 RESULTS

Quantitative Results. The analysis of actionability rankings revealed a weak level of agreement among students, with an average Kendall Tau-b correlation coefficient of 0.09. About 34.02% of student pairs showed negative correlations, indicating reversed preferences. Visual indicators were preferred over textual ones (mean scores of 9.14 and 7.90, respectively), while temporal and non-temporal cards scored similarly for actionability (both around 8.5). Non-social indicators were favored over social ones (9.95 vs. 7.05), and high-performance cards were slightly preferred over low-performance ones (8.62 vs. 8.38).

For interpretability, the agreement among students was also weak, with a mean Kendall Tau-b coefficient of 0.04. Visual indicators (9.04) were again preferred over textual ones (7.96), and temporal indicators were found more interpretable (9.20) than non-temporal ones (7.80). Non-social cards (9.21) were rated higher than social cards (7.79). High-performance and low-performance cards were rated similarly for interpretability (8.46 vs. 8.54).

The most actionable and interpretable card was Card 5, representing a visual, non-social, non-temporal indicator for a high-achieving student. In contrast, Card 4 (its textual counterpart) was rated the least actionable, and Card 2 the least interpretable. Figure 4 is a graphical representation of the cards repartition, with for instance Card 5 in the top right corner and Card 2 and 4 in the bottom left corner.

Figure 4: Repartition of the 16 cards in terms of actionability and interpretability.

Qualitative Results. In the justifications provided by students for their card ranking choices based on actionability and interpretability, we observed that 61% of the participants (19 students) explicitly mentioned the textual or visual aspect as a criteria used for their choices. Among them, 3 students showed preference towards texts (e.g. Student 15: "I found the written feedback better to understand, especially because many people have difficulty interpreting graphics") while 7 students were more positive towards the graphs. The other students either only mentioned those aspects as a criteria without providing details on their reasoning (Student 5: "I analysed the difference between the graphs and texts to distinguish between those that were very or not at all interpretable"), or were more nuanced (for example, Student 6 finds graphs more actionable but less interpretable).

Additionally, 61% (19 students) referenced the social comparison aspect. The vast majority (18) of them were in favour of social comparison, with statements such as "compared to the average of the class is easier to understand what I need to improve or I'm doing good" (Student 24). Only one student stated that "data that compare your performance with the class can be discouraging" (Student 23).

Temporality was mentioned by 26% (8) of the students. 5 students were in favour of information about the evolution through time, while 3 students were more nuanced or preferred no temporal information (e.g. Student 24: "The details of the weeks seemed too much and hard to understand but I think is actionable to some extent because if the student knows what they did differently in that week they can improve/change").

Finally, 13% (4 students) commented on the two profiles, whether high- or low-performance. Among those, 2 preferred cards represented high-achieving profiles (e.g. Student 3: "I chose the cards with the highest number of errors as not actionable and interpretable"), and 2 students cited the profiles as a criteria when ranking the cards without explaining the details.

5 DISCUSSION OF THE RESULTS

5.1 Comparison of the Dimensions

Textual vs Visual. Our study shows that there is a preference of students' for visual elements over textual elements in terms of both actionability and interpretability. This findings aligns with the results reported by Clark et al. (2004), Kühl et al. (2011) and Kuhlmann and Fiorella (2022), who showed that explanatory visuals are usually more effective than text alone. However, these results are not entirely consistent with the preferences students expressed in their rationale. Some students found text easier to interpret. Despite both representations conveying the same information, some students perceived the texts as more detailed. This aversion to graphs may stem from varying levels of data literacy among students. Indeed, research by Park and Jo (2015) indicates that the ease of interpreting graphs can depend on a student's data literacy skills, explaining why some students exhibit mixed feelings about graphs and prefer textual information. In AI-based systems, were the models are difficult to explain, adding text could potentially help students with less data literacy competencies in trusting the system. While the study confirms a general preference for visual elements regarding actionability and interpretability, it also uncovers nuanced and contrasting preferences through qualitative data analysis. This suggests that promoting students' trust on AI-based feedback systems would require further exploration on students' preferences. This also confirms the call from other researchers about the need of personalizing AI-based educational systems feedback (Ouyang and Jiao, 2021; Khosravi et al., 2022). In summary, the study not only reaffirms the general preference for visual explainability elements in terms of actionability and interpretability but also delves into the nuanced and contrasting preferences that emerge from the qualitative data. The exploration of these preferences within the context of data literacy provides valuable insights into the complexities of how individuals interact with and interpret different types of information representations.

Social Comparison and Profiles. Contrary to prior findings (Villalobos et al., 2023), our study revealed that non-social cards were generally perceived as more actionable than those incorporating social comparison. However, qualitative data showed that many students viewed social comparison positively for its ability to provide context and clarity. This duality highlights both the motivational potential of social comparison and its risks, as some students found it discouraging or intimidating. For example, one student remarked that comparing performance with peers could reduce confidence.

These mixed responses align with existing research, which identifies social comparison as a contentious element in feedback design (Bayrak et al., 2021; Vytasek et al., 2020). While some studies report a preference for social comparison in dashboards (Bodily et al., 2018; Schumacher and Ifenthaler, 2018), others note its potential to generate negative emotions (Guerra et al., 2016). Preferences often depend on students' goals, such as mastering a subject versus merely passing a course (Jivet et al., 2020; Villalobos et al., 2023). Our analysis found that high-performance profiles with social comparison received slightly higher scores than low-performance profiles, suggesting that social comparison may be more beneficial for high-achieving students. To mitigate its negative effects, feedback design should account for individual characteristics, such as academic performance and personal goals (Vieira et al., 2018). Offering options to personalize the inclusion or type of social comparison could foster a more supportive and effective learning experience, reducing the risk of discouragement.

Temporality and Self-Regulation. Unlike previous findings where temporal cards were preferred (Villalobos et al., 2023), our study found no global preference for temporality in terms of actionability. This discrepancy may be due to differences in graph types (line vs. radar), as students tend to prefer visualizations they find more familiar (Kuosa et al., 2016; Clark et al., 2004; Sahin and Ifenthaler, 2021). While some students appreciated temporal information for tracking progress, others found it confusing, as noted in one comment: "although useful to see the evolution, it is not much interpretable." Temporal indicators were rated slightly higher for interpretability, particularly for high-performing profiles, suggesting that students prefer comparing progress when their performance is strong.

Previous findings indicated that framing indicators in a temporal context facilitated self-regulated learning by providing benchmarks for students to track their progress (Villalobos et al., 2023). However, our study concentrates on shorter scenarios that offer task-specific guidance, alongside detailed feedback and advice. This approach presents a nuanced view of the impact of temporality, suggesting that direct, task-oriented support can modify the perceived significance of temporal indicators. This nuanced approach emphasizes the need to consider the specific educational context and support mechanisms when evaluating the role of temporal indicators in enhancing students' trust in the system and therfore, on their learning experiences.

5.2 Implications for Designing Actionable and Interpretable Indicators in AI-Based Systems

Our study provides valuable insights for designing actionable and interpretable feedback indicators in AI-based educational systems, with implications for both researchers and practitioners. These implications address critical considerations such as the format of indicators (visual vs. textual), the inclusion of social comparisons, and the balance between short-term and long-term feedback.

Visual vs. Textual Indicators. The results indicate a clear preference for visual indicators over textual ones in terms of both actionability and interpretability. Visual formats, such as graphs, enable students to quickly grasp key insights, making them particularly effective for fostering actionable learning behaviors. However, the study also highlights the diversity of preferences, with some students finding textual indicators easier to understand due to their perceived detail and clarity. This suggests that a combined approach is optimal: using visuals as the primary medium to convey information, complemented by textual explanations to support students with lower data literacy.

Social vs. Individual Feedback. The study reveals nuanced perspectives on social comparisons. While non-social indicators were generally perceived as more actionable, qualitative feedback shows that many students appreciate social comparisons for their ability to contextualize individual performance within a group setting. However, the potential for discouragement among low-performing students underscores the need for caution when incorporating social elements. Designers should consider offering personalized options, allowing students to toggle between social and individual views or tailoring the level of social comparison based on the learner's performance and goals. For example, high-achieving students may benefit more from comparative metrics, while low-performing students may respond better to individual progress indicators.

Temporal Feedback: Single Session vs. Multi-Session. Temporal indicators, which provide insights into progress across sessions, were found to enhance interpretability but elicited mixed reactions regarding actionability. Some students valued the ability to track long-term progress, while others found it overwhelming or confusing. This highlights the importance of designing temporal feedback with user preferences in mind. Systems should offer flexible temporal views, enabling learners to focus on session-specific data when needed while also being able to access historical trends for broader self-regulation and reflection. For example, toggles or filters can provide a seamless way to customize temporal feedback according to individual learning needs.

Recommendations for Dashboard Design. Given that this dashboard will be implemented in a real-world tool, the following recommendations can guide its development: (1) prioritize visual representations: use graphs and charts to present key insights, ensuring they are intuitive and easy to interpret; (2) provide textual explanations: complement visual feedback with short textual summaries to enhance understanding, especially for users with varying data literacy; (3) enable personalization: allow users to customize their dashboards by toggling between social comparisons, temporal views, and feedback formats, as personalization can improve user engagement and align feedback with individual learning goals (Smith, 2019); and (4) adapt feedback to user profiles: consider learner-specific factors, such as performance levels and preferences, to design indicators that increase engagement.

Broader Implications for System Designers. From a broader design perspective, our findings highlight the need for feedback indicators in learning analytics systems to (1) prioritize **clarity and usability**, ensuring they are accessible to diverse users; (2) to address **data literacy** by combining visuals and text with explanatory resources to build trust; and (3) to balance **motivation and transparency**, using social comparisons thoughtfully and tailoring temporal feedback to the learning context. The mixed reactions to social comparison and preference for non-social indicators emphasize the importance of personalization, allowing learners to customize feedback based on their preferences and goals. Adapting temporal and comparison features to individual needs should enhance engagement and learning outcomes. By incorporating these principles, AI-based systems should provide actionable, interpretable feedback that fosters trust and transparency, aligning with the goals of Explainable AI (XAI) in education to support all stakeholders (Khosravi et al., 2022).

6 CONCLUSION AND FUTURE WORK

This study is a contribution towards the understanding on how AI-based educational systems can be empowered with mechanisms to make their results more explainable through interpretable and actionable indicators. Our findings reveal a nuanced preference among learners for visual over textual indicators, highlight the mixed responses to social comparison, and underscore the importance of considering temporality and learner profiles in the design of educational AI systems. Importantly, our research underscores the need for designing AI-based educational tools that are not only technically effective but are also able to make informed use of available data to encourage students to change their behaviour for the better. Moreover, this research goes towards making AI-based systems transparent to students. This is necessary if students are to have enough trust to follow the recommendations and advice provided by these systems, as without trust, ”analytics can have no influence on the learning activity” (Wise et al., 2016).

Future work should explore the integration of personalized learning analytics that take into account individual learner characteristics, such as data literacy levels and learning goals. The next step is to integrate actionable and interpretable AI-based feedback into tools used by students during a real course, in order to assess the impact of this type of feedback on students' trust, understanding of the system, and on their pedagogical outcomes. Moreover, there is a need for studies to assess the long-term impact of different indicator types on learning outcomes and student engagement. As AI continues to be used in educational contexts, ensuring that these systems are transparent, understandable, and aligned with human learning processes is essential for maximizing their educational value and fostering an environment of trust and effective learning, and ourish the global dialogue on AI regulation.

ACKNOWLEDGEMENTS

This work was supported by the STIC-AmSud LAFE project.

REFERENCES

Bayrak, F., Nuhoğlu Kibar, P., and Kocadere, S. A. (2021). Powerful Student-Facing Dashboard Design Through Effective Feedback, Visualization, and Gamification. In *Visualizations and Dashboards for Learning Analytics*. Cham.

Bodily, R., Ikahihifo, T. K., Mackley, B., and Graham, C. R. (2018). The design, development, and implementation of student-facing learning analytics dashboards. *Journal of Computing in Higher Education*.

Clark, R., Lyons, C., and Hoover, L. (2004). Graphics for learning: Proven guidelines for planning, designing, and evaluating visuals in training materials. *Performance Improvement*.

Conati, C., Barral, O., Putnam, V., and Rieger, L. (2021). Toward personalized XAI: A case study in intelligent tutoring systems. *Artificial Intelligence*.

Félix, E., Amadieu, F., Venant, R., and Broisin, J. (2022). Process and Self-regulation Explainable Feedback for Novice Programmers Appears Ineffectual. Lecture Notes in Computer Science, Cham.

Guerra, J., Hosseini, R., Somyurek, S., and Brusilovsky, P. (2016). An Intelligent Interface for Learning Content: Combining an Open Learner Model and Social Comparison to Support Self-Regulated Learning and Engagement. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*.

Jivet, I., Scheffel, M., Schmitz, M., Robbers, S., Specht, M., and Drachsler, H. (2020). From students with love: An empirical study on learner goals, self-regulated learning and sense-making of learning analytics in higher education. *The Internet and Higher Education*.

Jørnø, R. L. and Gynther, K. (2018). What Constitutes an 'Actionable Insight' in Learning Analytics? *Journal of Learning Analytics*.

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., and Gašević, D. (2022). Explainable Artificial Intelligence in education. *Computers and Education: Artificial Intelligence*.

Kuhlmann, S. and Fiorella, L. (2022). Effects of instructor-provided visuals on learner-generated explanations. *Educational Psychology*.

Kuosa, K., Distante, D., Tervakari, A., Cerulo, L., Fernández, A., Koro, J., and Kailanto, M. (2016). Interactive Visualization Tools to Improve Learning and Teaching in Online Learning Environments:. *International Journal of Distance Education Technologies*.

Kühl, T., Scheiter, K., Gerjets, P., and Gemballa, S. (2011). Can differences in learning strategies explain the benefits of learning from static and dynamic visualizations? *Computers & Education*.

Molenaar, I. and Wise, A. (2022). Temporal Aspects of Learning Analytics - Grounding Analyses in Concepts of Time. In *The Handbook of Learning Analytics*.

Nowell, L., Norris, J., White, D., and Moules, N. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative*.

Ouyang, F. and Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*.

Park, Y. and Jo, I.-H. (2015). Development of the Learning Analytics Dashboard to Support Students' Learning Performance. *Journal of Universal Computer Science*.

Sahin, M. and Ifenthaler, D. (2021). Visualizations and Dashboards for Learning Analytics: A Systematic Literature Review. In *Visualizations and Dashboards for Learning Analytics*, Advances in Analytics for Learning and Teaching.

Schumacher, C. and Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior*.

Sharples, M. (2023). Towards social generative AI for education: theory, practices and ethics. *Learning: Research and Practice*.

Smith, P. (2019). Engaging online students through peer-comparison progress dashboards. *Journal of Applied Research in Higher Education*.

Spencer, D. and Garrett, J. J. (2009). *Card sorting: designing usable categories*. Brooklyn, N.Y.

van Leeuwen, A., Teasley, S., and Wise, A. (2022). Teacher and Student Facing Learning Analytics.

Vieira, C., Parsons, P., and Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education*.

Villalobos, E., Hilliger, I., Pérez-Sanagustín, M., González, C., Celis, S., and Broisin, J. (2023). Analyzing Learners' Perception of Indicators in Student-Facing Analytics: A Card Sorting Approach. In *Responsive and Sustainable Educational Futures*, Lecture Notes in Computer Science, Cham.

Vytasek, J. M., Patzak, A., and Winne, P. H. (2020). Analytics for Student Engagement. In *Machine Learning Paradigms: Advances in Learning Analytics*, Intelligent Systems Reference Library. Cham.

Winne, P. H. (2021). Open Learner Models Working in Symbiosis With Self-Regulating Learners: A Research Agenda. *International Journal of Artificial Intelligence in Education*.

Wise, A. F., Vytasek, J. M., Hausknecht, S., and Zhao, Y. (2016). Developing Learning Analytics Design Knowledge in the "Middle Space": The Student Tuning Model and Align Design Framework for Learning Analytics Use. *Online Learning*.

Álvarez, R. P., Jivet, I., Pérez-Sanagustín, M., Scheffel, M., and Verbert, K. (2022). Tools Designed to Support Self-Regulated Learning in Online Learning Environments: A Systematic Review. *IEEE Transactions on Learning Technologies*.