

Detect, Explain, Act: How Teachers Trust and Use an Explainable Real-Time Monitoring Dashboard to Detect Student Outliers in Class

Jeroen Ooge*

Department of Information and
Computing Sciences
Utrecht University
Utrecht, Netherlands
Department of Computer Science
KU Leuven
Leuven, Belgium
j.ooge@uu.nl

Anissa Faik*

Department of Computer Science
KU Leuven
Leuven, Belgium
anissa.faik@odisee.be

Katrien Verbert

Department of Computer Science
KU Leuven
Leuven, Belgium
Augment, an imec research group at
KU Leuven
Leuven, Belgium
katrien.verbert@kuleuven.be

Abstract

Adaptive learning platforms can personalise learning in classrooms, yet teachers need tools to monitor student progress and rapidly identify students who need additional attention. Monitoring dashboards that detect outlier students can fulfil this need, but insufficient explanation of the detection logic may undermine trust and lead to inappropriate use. To address this challenge, we iteratively designed and implemented a real-time monitoring dashboard with model-centric and data-centric explanations, informed by teacher input. We then conducted a counterbalanced within-group experiment with follow-up interviews to study how 11 teachers used the dashboard in a real classroom and how they engaged with the explanations to calibrate their trust. We found that teachers successfully integrated the dashboard into their classes, and that their trust was shaped by many dispositional, situational, and learnt factors. Crucially, the data-centric explanations enabled teachers to validate the accuracy of outlier predictions, check alignment with their prior knowledge of students, and identify suitable interventions. Based on these findings, we present design recommendations for explainable outlier detection systems in education.

CCS Concepts

• **Applied computing** → **E-learning**; • **Human-centered computing** → **Human computer interaction (HCI)**; **Visualization systems and tools**.

Keywords

XAI, outlier detection, education, teacher-facing dashboard

ACM Reference Format:

Jeroen Ooge, Anissa Faik, and Katrien Verbert. 2026. Detect, Explain, Act: How Teachers Trust and Use an Explainable Real-Time Monitoring Dashboard to Detect Student Outliers in Class. In *LAK26: 16th International*

*Jeroen Ooge and Anissa Faik contributed equally to this research, which is based on Anissa Faik's master's thesis [15] supervised by Jeroen Ooge and Katrien Verbert.



This work is licensed under a Creative Commons Attribution 4.0 International License. *LAK 2026, Bergen, Norway*

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2066-6/26/04

<https://doi.org/10.1145/3785022.3785087>

Learning Analytics and Knowledge Conference (LAK 2026), April 27–May 01, 2026, Bergen, Norway. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3785022.3785087>

1 Introduction

As in many other domains, artificial intelligence (AI) offers opportunities to improve decision-making and support in education [19, 27, 31]. One such example in the context of adaptive learning platforms is data-driven detection of outlier students, who may need additional support from teachers. The key challenge, however, is that teachers may not trust or effectively use such outlier detection systems if they do not understand how they work or why specific students are flagged as outliers [10, 20]. A potential remedy is explainable AI (XAI), which encompasses techniques for explaining the outputs of opaque AI systems [1, 4]. These explanations should be actionable for teachers to make informed classroom interventions. In education, the demand for XAI is particularly high, as it can address education-specific challenges, such as handling noisy data and supporting learning processes [20]. Instructors often lack sufficient understanding of AI, raising concerns and hesitation about incorporating AI into teaching [10].

We address three key challenges in the context of XAI in education. First, we aim to improve teachers' understanding of outlier detection models [10, 20]. Second, we study the complex nature of trust and its impact on AI use in education. Third, we operationalise model-centric and data-centric explanations in an educational outlier detection system. Existing literature primarily emphasises detection accuracy, neglecting the importance of explainability.

Concretely, we extended an adaptive learning platform with a teacher dashboard that detects outlier students and exercises. During an iterative design process involving teachers, we also designed explanations for these detections. We then conducted a between-group experiment and follow-up interviews with 11 teachers in a real classroom setting to evaluate our explainable dashboard and examine how they trusted and used it. Our research questions were as follows:

RQ1. Which factors affect teachers' trust in an explainable outlier detection system?

RQ2. How do teachers assess model-centric and data-centric explanations in terms of effectiveness, satisfaction, and trust?

RQ3. How do teachers (intend to) use our outlier detection dashboard in a classroom setting?

Our findings show that teachers effectively integrated the dashboard into their classroom in multiple ways, with their trust shaped by dispositional, situational, and learned factors. The data-centric explanations enabled teachers to verify the accuracy of detected outliers, compare results with their own knowledge of students, and identify possible interventions. We discuss these insights to inform the design of explainable outlier detection systems in education.

2 Background and Related Work

2.1 Explainable Artificial Intelligence

AI models often lack transparency in that they do not disclose their internal workings and are essentially “black boxes” [1, 4]. This raises questions about how much trust can be placed in them. To address this issue, AI systems should be both interpretable and explainable. Doshi-Velez and Kim [12] define interpretability as “the ability to explain or present in understandable terms to a human.” Barredo Arrieta et al. [4] define explainability as “the details and reasons a model gives to make its functioning clear or easy to understand, given a certain audience.” Considering these definitions, *explainable artificial intelligence* (XAI) aims to provide explanations for the outcomes of opaque AI systems.

In education specifically, concerns are emerging around educational interventions supported by AI [20]. Instructors often possess an insufficient understanding of AI, leading to concerns about its use and questioning its effectiveness and trustworthiness in carrying out tasks without error [10]. There is a need for XAI in education to effectively communicate AI outcomes to instructors, enhance their perceptions of AI, and support them in addressing the challenges they encounter when integrating AI into teaching practices.

2.2 Different Types of Explanations

Many different XAI techniques have been proposed and can be categorised in various ways. For example, explanations can be classified by the scope and level of detail they convey. Mohseni et al. [26] differentiate between *global* explanations, which provide insights into the overall functioning of a model, and *local* explanations, which focus on explaining individual instances. Alternatively, Mohseni et al. [26] and Liao et al. [22] distinguish between explanations based on the questions they answer. For example, *how* explanations provide insights into the internal functioning of the model, while *why* explanations offer reasoning behind specific predictions. Other approaches include contrastive or similar examples, simplification, and performance measures. Finally, explanations can be *data-centric* [3], describing the data used for model training, or *model-centric*, explaining the overall model, similar to global explanations.

Explanations can have a visual, verbal, or analytical format [26]. Visual explanations use visual elements, while verbal explanations rely on natural language, and both can be combined [35]. Analytical explanations are similar to visual explanations but use more advanced visualisation techniques to explore the data and model representations [33]. When designing explanations, the target audience matters, as needs differ between AI novices, data experts, and AI experts [26]. For AI novices like most teachers, XAI should prioritise transparency, trust, fairness, and privacy awareness [26].

2.3 Trust in Automated Systems and Analytics

Trust in automated systems is a slippery concept that is particularly hard to measure because it evolves and is influenced by dispositional, situational, and learnt factors [17]. *Dispositional trust* depends on people’s enduring tendency to trust arising from biological and environmental influences such as culture, age, and personality traits. *Situational trust* depends on the specific interaction context, encompassing both internal factors (e.g., self-confidence) and external factors (e.g., system type). *Learnt trust* is based on past or current experiences with the automated system and is influenced by people’s pre-existing knowledge and the system’s performance. In the context of automated systems and XAI evaluations, trust is best considered as a multi-dimensional construct [25, 31].

Specifically, trust in learning analytics systems and AI-based educational technology is equally multifaceted, including factors such as perceived benefits and lack of transparency [28]. To ensure usable and useful systems, it is critical to involve stakeholders throughout the design process [19]. Moreover, privacy and transparency require careful consideration for trust-building [2]. While explanations can enhance transparency, experts themselves vary in which explanations they find important and trustworthy [34]. Consulting teachers to include pedagogically meaningful [27] and actionable [20] explanations is therefore key. Furthermore, visual design alternatives to traditional dashboards, such as data storytelling and narratives, may improve teachers’ reflection and understanding [16].

3 E-Learning Platform With Outlier Detection Dashboard

We extended an e-learning platform for secondary school mathematics practice [31] with an explainable learning analytics dashboard. Resulting from iterative design with teachers, the dashboard enables: (1) monitoring student performance and outliers based on speed and attempts; (2) intervening when students struggle or excel; and (3) identifying outlier exercises. Some teachers participated in multiple iterations.

3.1 Design Process

We first interviewed 10 teachers (semi-structured) about their use of and wishes for digital tools to support students. Teachers reported they rarely use such tools due to their technical limitations and scattered functionalities. Discussing monitoring dashboards from the literature [9, 11, 13, 23], teachers stressed the importance of focusing on individual students, rather than comparing them to class averages. They also suggested personalising support with a warning system for underperforming students and recommended actions. Most teachers identified completion time as a key factor in determining which students require support. These findings inspired us to create an outlier detection dashboard with performance timelines for all students. Our low-fidelity prototype (version 1 in Figure 1) included distribution plots inspired by Campen and Moleenaar [9] and visual per-question performance indicators inspired by Dourado et al. [13].

In a co-design session with 2 human-computer interaction experts, we refined the visualisations and interactions in our prototype (version 2 in Figure 1). We added histograms for the time and attempt parameters used by the outlier detection algorithm (see



Figure 1: Our design process with an initial version based on semi-structured interviews (version 1), a low-fidelity prototype based on a co-design session (version 2), and a high-fidelity prototype (version 3). We conducted usability tests for the latter two versions.

Section 3.3) and enabled highlighting students' position therein. We also included which exercises students were struggling with to provide actionable insights for teachers.

Next, we conducted task-based think-aloud studies with 7 teachers to identify usability issues in our prototype. Teachers also suggested various functionalities and interventions, including filtering out outlier students, providing general feedback, pairing strong and weaker students, and providing extra practice materials for weaker students. We incorporated these suggestions in a high-fidelity prototype (version 3 in Figure 1), which we again evaluated and refined through task-based think-aloud studies with 5 teachers. Further details about our design process can be found in [15].

3.2 Monitoring Dashboard and Explanations

Our final dashboard in Figure 2 provides a real-time overview of student performance, highlighting outliers and suggesting interventions. The timelines visualise each student's progress, with colour-coded indicators for correct and incorrect answers. Clicking on a specific exercise in a timeline aligns all students to it, allowing teachers to compare performance (see Figure 5). Clicking on outlier students reveals suggested interventions: teachers could choose to offer help or assign a buddy to 'weak' students, and assign harder exercises to 'strong' students. In both cases, teachers could also choose not to intervene and ignore the outlier warning. Moreover, the distribution plots and timelines are linked: clicking or hovering over either of them highlights the corresponding elements in the other plot. Finally, teachers can filter on specific exercises and track exercises that are generally difficult for students.

To give teachers insights into the system's outlier detection, we incorporated two types of explanations in our platform. First, the timelines and distribution plots in Figure 2, discussed above, serve as **data-centric why explanations**. Indeed, they visualise the data used by the outlier algorithm and contextualise individual students' position within the distribution of speed and attempt scores. Second, we added a **model-centric how explanation** that uses illustrated text to clarify the model's internal functioning and parameters (see Figure 7). Teachers were referred to this model-centric explanation upon logging in and could revisit it on a page titled "Explanation about detection system."

3.3 Outlier Detection and Interventions

Outlier students were detected with the Local Outlier Factor (LOF) algorithm, which flags students whose local density is lower than

that of their k nearest neighbours [6, 36]. We set k to half the class size and marked students with an LOF score above 1 as outliers. As input to the LOF algorithm, each student had two scores per set of exercises, which were updated after each attempt. First, a speed score increased by 1 if the attempt was as fast or faster than the class average, and decreased by 1 otherwise. Second, an attempt score was updated as follows:

$$\text{new score} = \begin{cases} \text{old score} + 1/N & \text{for correct attempts,} \\ \text{old score} - N/(N+1) & \text{for incorrect attempts,} \end{cases}$$

where N is the number of attempts. Outliers with both scores above average were labelled 'strong', and those below average were labelled 'weak.' Teachers could assign a buddy or provide help themselves to struggling students. We also implemented a multivariate Elo rating system [32] to estimate student proficiency and exercise difficulty, enabling teachers to assign harder exercises to students who progress quickly. Note that the focus of our study was not to optimise the LOF and Elo algorithms, but to evaluate our dashboard.

4 Methods for In-Class Experiment

We recruited 11 mathematics teachers via email, social media, and personal contacts to use our platform during regular classes. None of these teachers participated in our design process, described in Section 3.1. To account for the small sample, we conducted a within-group experiment to assess how our explainable dashboard impacts teachers' trust, perceived accuracy, satisfaction, and perceived effectiveness of explanations. The study was approved by KU Leuven's ethical committee (reference: G-2022-6129).

Before the experiment, teachers composed series of fixed exercises for all their students and read the model-centric explanation of the outlier detection system. Adaptation based on Elo ratings only kicked in when teachers later assigned harder exercises to outlier 'strong' students. To study the impact of our data-centric explanations, we created a baseline dashboard that displayed only student names, without visualisations (see Figure 4).

During the experiment, teachers used both dashboards in a counterbalanced order to account for ordering effects. Concretely, 6 teachers first saw the full dashboard for 25 minutes and then saw the baseline dashboard for 15 minutes (**Group A**; P3, P5, P7–9, P11), and 5 teachers used the dashboards in the opposite order (**Group B**; P1–2, P4, P6, P10). Teachers could spend more time with the full dashboard because it was more complex, and its visual explanations were the main focus of our research. While students were solving exercises, teachers freely interacted with the monitoring dashboard. We logged hovering, clicking, and filtering interactions with suggested interventions, outlier students, and the explainable visualisations. Whenever teachers registered an intervention for detected outliers, they had to indicate whether they agreed with the detection. As our study focused on teachers' perceptions, we did not collect student data.

After the experiment, teachers shared initial impressions via email. Except for P11, all then participated in semi-structured interviews for more in-depth discussions based on selected questions from existing questionnaires (see Table 1). We covered trust based on constructs from the Madsen-Gregor trust scale [25], including trustworthiness, perceived reliability, perceived understandability,

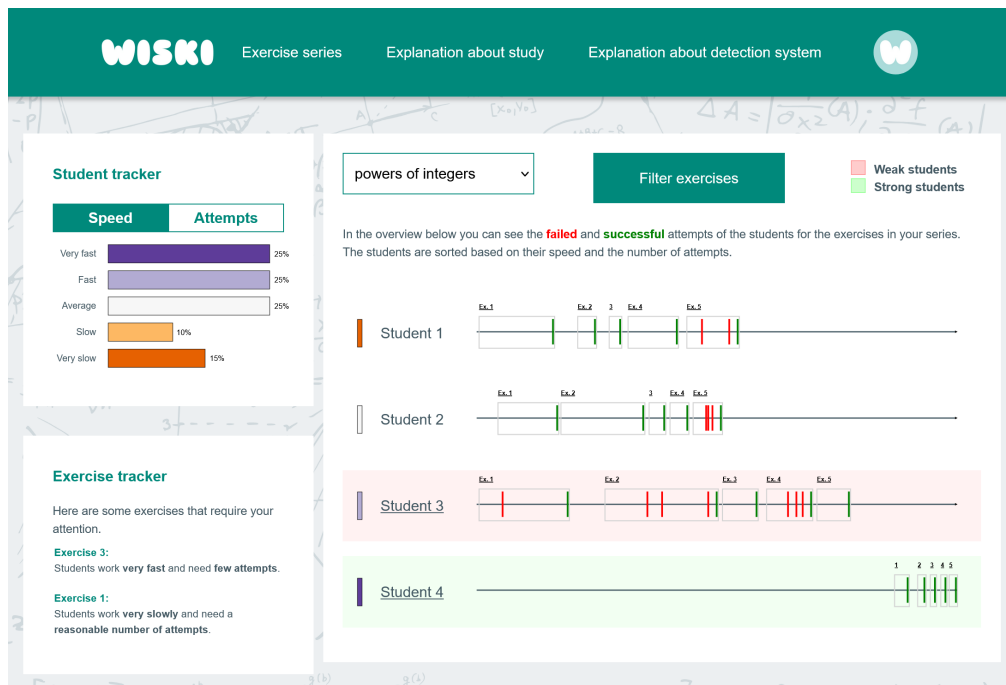


Figure 2: Our monitoring dashboard with highlighted student outliers. In the timelines, green and red vertical lines indicate correct and incorrect attempts, respectively, and boxes indicate which exercise each attempt belongs to.

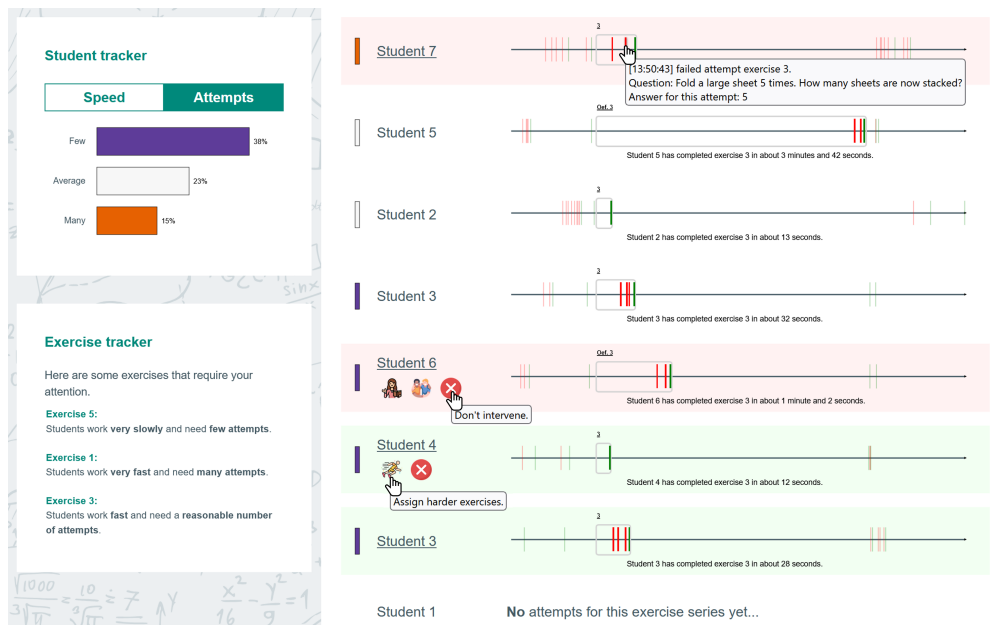


Figure 3: Our monitoring dashboard with aligned timelines and suggested interventions for outlier students. Hovering over attempts in the timelines shows the exercise and students' answers.

and faith. Perceived accuracy questions explored teachers' views on the system's accuracy. Satisfaction questions were adapted from the explanation satisfaction scale by Hoffman et al. [18], assessing understandability, satisfaction, sufficiency of detail, and usefulness of the explanations. Finally, effectiveness questions evaluated whether explanations supported teachers in making interventions.

The semi-structured interviews were transcribed and thematically analysed, following guidelines by Braun and Clarke [5]. We started with deductive coding and then grouped codes about **trust**, **usefulness and satisfaction**, and **integration in classrooms**. One leading author coded all transcripts and discussed initial themes with the other leading author to optimise their focus. The final themes result from three major iterations. Afterwards, we inductively mapped themes to the three-layered trust framework by Hoff and Bashir [17].

5 Results of In-Class Experiment

5.1 Teachers' Interactions with the Dashboard

The logged hover and click counts in Figure 5 show that teachers most often interacted with detected student outliers. When data-centric explanations were available, they often examined timelines and distribution plots, most frequently focusing on successful and failed attempts, followed by very slow or very fast students. Group A engaged with a broader range of components than group B, but there seem to be no further major differences in interactions between the groups. Thus, the order in which teachers used the dashboards did not significantly impact their interactions. Most teachers registered 4 or fewer interventions for outlier students; only two teachers in group B registered 8 and 9 interventions, respectively, on the baseline dashboard. While this makes further analysis preliminary, one interesting observation is that teachers interacted more with student outlier detections on the dashboard with data-centric explanations (68%) versus the baseline (32%), and also agreed more with detections (52% vs 16%).

5.2 Trust

Some teachers placed **blind trust** in the outlier detection system because they believed it to be objective and well-designed, while others were more cautious.

Some teachers expressed blind trust in the system, believing in its objectivity and setup. For example, **P4** stated: *"I trust it [...] because there's an objective system that generates judgments based on data."* Similarly, **P6** assumed the detection system was well-designed despite not having a precise understanding of the system's inner workings: *"I suppose if the detection system is well-made, you can trust it. But of course, you don't know exactly how it works."* In contrast, other teachers such as **P9** took a more careful stance towards fully relying upon automated technologies for outlier detection, mentioning *"I'm not naturally inclined to trust the computer for that."*

To assess the accuracy of the outlier detection, teachers heavily relied upon **alignment with their own findings** and knowledge of students. Alignment sparked trust; disalignment sparked distrust.

Some teachers expressed confidence in the system because its detected outliers aligned with their expectations. For example, **P5** stated: *"I think the detections [of outlier exercises] were accurate. I expected them because the exercises were a bit more challenging."* Similarly, for detecting outlier students, teachers appreciated the system's ability to identify students who were guessing, as it aligned with their observations. **P1** remarked: *"Yes, everything matched and made sense except for one student, just as I expected. I knew that one student would guess."*

In addition to observations, several teachers emphasised the importance of their prior knowledge about students when assessing the system's accuracy and trustworthiness. **P10** agreed that the system was accurate and working as expected: *"There were no surprises, really. The students whom I expected to perform well actually did so."* Multiple teachers seconded this remark. For example, **P7** explained: *"Because I've known the students for a while now, I know which ones are strong in mathematics and which ones are slightly weaker. The system also indicated that the high-performing students were doing well, and those who were a bit weaker aligned with how I see them in class."* Similarly, **P3** explained that having no prior knowledge of students might influence how strongly they would be able to trust the system: *"Because I really know my students, I trust the system. But if I were to start the school year with a completely new class or, for any other reason, find myself in an unfamiliar setting, I can't be sure if it would be as reliable. It's all about knowing my students and recognising their weaknesses."*

However, some teachers expressed scepticism and lacked trust in the system due to inconsistencies with their perceptions and observations. For example, **P8** remarked: *"It didn't always correspond with my observations. When distinguishing between weak and strong learners, I interpret it differently in my lessons."* **P9** also noticed a notable difference between what they observed in the students and what the system's dashboard indicated. They expressed distrust and elaborated: *"There was quite a difference, I think, between what I observed in the students and what the dashboard said. It wasn't always 100% accurate. [...] The dashboard primarily highlighted which students were fast or slow. I noticed that some students were engaged in guessing, which might have led them to be labelled as fast. Even in terms of strong and weak students, it didn't fully correspond to what I personally observed."* Also **P10** echoed the sentiment of mistrust, stating: *"According to my knowledge of the students, [I do] not really [trust it]. [...] Some students scored poorly, while in class, I had the impression that they were actually stronger."* **P10** did acknowledge that there might be some reasoning as to why some of their students were detected, but still did not trust it.

Analysing the **data-centric explanations** revealed the **accuracy** of detected outliers, fostering trust in the system.

P1 used our monitoring dashboard in multiple classes and beautifully described how they were initially sceptical but gained trust in

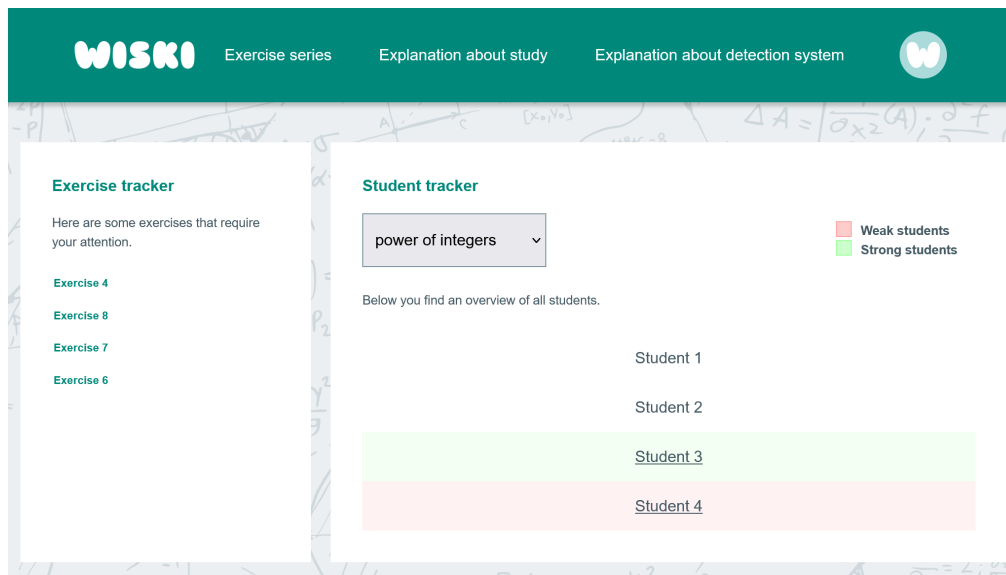


Figure 4: Our baseline dashboard showed outlier students and exercises without data-centric explanations.

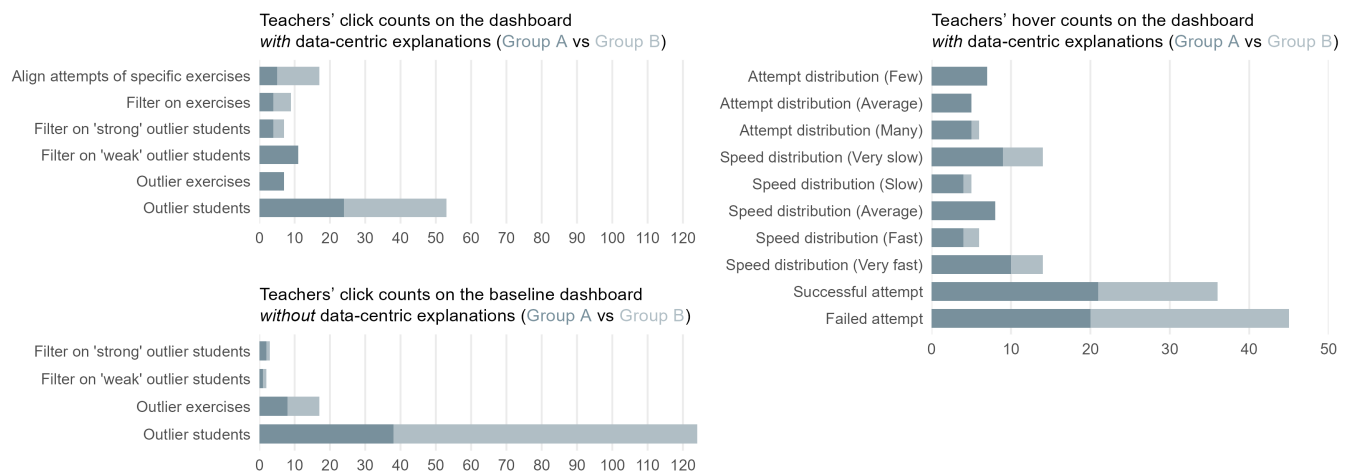


Figure 5: Click and hover counts for the dashboards with and without data-centric explanations.

the system by engaging with the data-centric explanations and observing consistent patterns. *"In the beginning, [...] using the timeline, I wanted to verify it was accurate. Like, is it being displayed correctly, showing me the time taken and the number of correct attempts? Does it match up? [...] You could effectively see that the students who made only one or two attempts and those who took less time were also chosen as better learners."* Especially cases where the outlier results initially contradicted their expectations were pivotal for P1: *"I also noticed that students whom I know are weaker were being highlighted as stronger. And I thought, 'Really? Can that be?' But there were instances where it was true: the students who diligently completed their exercises on paper without attempting them multiple times before moving on. I wasn't surprised because I knew that you receive more points for providing the correct answer quickly or getting it right on*

the first try. [...] It's funny to see that the weaker students also rank higher. It's nice for them to see that they're also earning points or being ranked higher because of their good approach to solving problems. I found that quite enjoyable."

P3 also described how initial disagreement with certain detections was resolved after analysing the students' behaviour through the timeline, leading to a better understanding of the algorithm's rationale and increased trust: *"I have a few students who are strong but were rated as less strong [...]. One of my students is extremely slow but very strong in mathematics, so he just needs a lot of response time to react. And I could follow that on the timeline. And yes, that can give a distorted picture of his abilities, since he was often seen as a weaker student. [...] But it's because I know the underlying idea, the reasoning behind the classifications, that I [trust the system]."*

In addition, **P2** gained trust by noticing that a student who usually performed well was marked as weaker due to guessing: *“I could see on the timeline that they were guessing because they genuinely believed they always had a one-in-four chance and would eventually select the correct answer. [...] So yeah, I trusted that the system was identifying the students correctly.”*

Teachers **built trust over time** by being exposed to the dashboard and monitoring the system’s outlier detection performance.

P1 highlighted the need for time when building trust and working with the system: *“The more I work with the platform, the more my trust in it grows. You learn that the system responds correctly. [...] I had looked into [the system’s detection rationale] thoroughly with the first group, [...], so I trusted it would work in the next lessons as well.”* Yet, **P1** also acknowledged the importance of ongoing evaluation: *“It seemed to be accurate [...], but to assess it better, I need to try it a few more times.”* Other teachers seconded that they found it challenging to assess the system’s accuracy. For example, **P8** noted they *“haven’t engaged enough with it to give a proper answer”* and **P2** *“would need to play around with it more to extract valuable insights.”*

Another reason for requiring time to build trust was the need to familiarise themselves with the dashboard and the data-centric explanations. For example, **P8** remarked that giving teachers more time to familiarise themselves with the dashboard in advance could help: *“There are so many things [on the dashboard]. Especially when I first saw it, I found it really difficult to grasp everything at once.”*

Teachers could start **doubting the system’s reliability as detections fluctuated** frequently.

One participant, **P6**, expressed less or mixed trust in the system due to its fluctuating detections: *“Sometimes [I trusted it], but sometimes not. [...] It would say ‘weak,’ and then suddenly that student would be labelled ‘strong,’ and then they would become ‘weak’ again. I think it’s related to the constantly running algorithm, but it does cause confusion at times.”* **P6**’s lacking trust stemmed from the inconsistency they observed in the system’s detections.

5.3 Satisfaction and Usefulness

Teachers found the outlier detections and timelines **satisfactory and useful** for monitoring students and intervening promptly.

P2 pointed out the question tracker feature of the dashboard as a suitable support tool to intervene promptly: *“I found it particularly useful when a question required extra attention.”*

In addition, most teachers strongly appreciated the timelines in the data-centric explanations. For example, **P4** stated: *“It provides a clear snapshot of when students start and complete their exercises, allowing me to assess their progress and identify areas that need attention.”* **P5** seconded this: *“I can easily track students’ completion time, performance, and problem-solving approach. The dashboard [presents] this information concisely and accessibly.”* Many teachers highlighted the effective colour-coding. For instance, **P7** remarked: *“The red and green indicators are so clear. They quickly catch my*

attention and prompt me to take action.” As highlighted by **P10**, the timeline data could inform teachers about new students’ mastery level: *“If you can already see this in the first trimester, then you have an idea about your students before you fully know them.”*

Furthermore, most teachers elaborated on how the visual timelines helped them with decision-making and optimising classroom management, particularly because of the time aspect and real-time updates. This explains why they preferred the dashboard with timelines over the baseline dashboard without them. For example, **P1** shared that the timeline helped them decide when to intervene: *“I did want to help the person who had been working on [an exercise] for so long when I looked at the timeline.”* **P7** added: *“Especially if a student had been working on an exercise for a long time [...] you notice they have tried it multiple times and failed, then you think, ‘Maybe I should take a look myself.’”* Moreover, **P2** explained: *“[The timeline] allowed me to spot when someone was guessing [...] and address the student on it.”* Finally, **P7** highlighted that real-time monitoring at a distance could prevent student discomfort: *“I really enjoyed being able to track in real-time which exercise students were working on. Otherwise, you have to literally look over their shoulders, and they don’t really like that.”* In response, **P2** nuanced that observing students’ behaviour in person could still be valuable: *“Observing their movements helped me understand that an exercise took longer because they also went to get scrap paper.”*

The data-centric and model-centric **explanations improved understanding** of the algorithm and its parameters.

Teachers found the model-centric explanations clear and sufficient for a basic understanding of the algorithm, even for those with a limited technical background. Some could correctly recall and explain the algorithm’s overall logic. Two teachers even asked for more detailed model-centric explanations because they were unsure how the system made its detections. **P6** emphasized that the ‘strong’ and ‘weak’ labels for students were not properly explained. **P1** suggested that more details about the precise allocation of points per attempt could be interesting, although not essential.

However, some teachers admitted they had not focused much on the model-centric explanation and had limited knowledge of it. For example, **P9** rushed through it due to time constraints, **P2** said they had no idea how students were detected, and **P10** had a hard time explaining the model’s parameters.

Some teachers also seemed to have mainly deduced information about the algorithm’s functioning and parameters based on the data-centric explanations. For example, **P4** speculated that *“detections were probably based on the accuracy of the answers and speed.”* **P7** and **P8** offered a similar perspectives, stating: *“I suppose fast and accurate answers represent strong students.”*

Teachers expressed a **need for control** over the outlier detection by tuning parameters and customising visual components.

Inspired by observed cases of slow yet capable students, **P3** recognised the potential value of changing the weights of model parameters based on individual student needs. They suggested: *“Maybe it would be an idea to make adjustments for such a student,*

so that less emphasis is placed on slowness in some learners because you're aware of it." Thus, by adjusting the focus on aspects such as slowness and speed, teachers could tailor their approach to cater to students' unique learning styles.

P10 suggested another form of control: personalising the dashboard by altering its visual components. Acknowledging that different teachers have varying preferences and teaching styles, they suggested letting teachers switch between the baseline and full dashboard layouts: *"I should be able to switch between the two. I find that quite interesting. Or even just choosing for yourself what you want to see on your screen as the teacher."*

5.4 Dashboard Integration in Classrooms

The dashboard can help teachers decide whether students need **remediation or differentiation** based on their performance.

P2 shared their thoughts on using the dashboard for remediation and for task differentiation. They said: *"I would love to use it to differentiate tasks. For example, if a student performed poorly on a particular test section, I could assign them additional exercises."* Similarly, **P5** described how they would use the bar charts, categorising students according to speed and accuracy: *"I think it is really useful for situations where I need to assess which students need remediation. Just looking at the dashboard, I would pick out the students who make a lot of mistakes. When differentiating, the bar charts really help group students and see how much time or attempts they need."* Finally, **P10** highlighted the feature to assign extra exercises to "strong" outliers: *"This allows you to assign more tasks to some students within the same time frame as others."*

Some teachers **shared data insights with students** to stimulate them, while others cautioned about potential negative impacts.

Some teachers chose to share dashboard information directly with students to give a glimpse into their performance. For example, **P1** shares their positive experience: *"I could say to them, 'Hey, the system recognises you as a strong student', and you could see them light up with pride."* Similarly, **P10** reflected on the impact of the dashboard on her dynamic and energetic class, noting: *"It was indeed different from usual. The class was quite lively, and they were definitely active, especially when you could say, 'Come on, you're in the red zone, let's get back to green.' They would make an effort, although some more than others."* In addition, **P5** described students' positive reactions: *"They found it fun to see if they were considered good or average students. It even sparked some laughter."* In contrast, **P2** expressed their reservations about sharing the dashboard with students: *"If they see other students who didn't score as well, the boys in the class can become quite competitive and start mocking others. [...] I would like to avoid that."*

In physical classrooms, **balance technology use** with classroom management and personal connections with students.

Multiple teachers indicated it was challenging to strike a balance between offering assistance and leveraging the dashboard's

insights. For example, **P8** shared their initial struggles, remarking: *"I had to figure out how it worked and what it exactly meant. It was a lot to process on the computer while also attending to the students."* **P9** further emphasised the delicate nature of motivating students while monitoring the dashboard: *"It's difficult to monitor the dashboard and walk around at the same time. You would need to be in two places at once."* **P2** agreed that observing students' progress on the dashboard and physically interacting with them posed logistical challenges: *"Identifying which exercises required attention was very useful. However, I had to physically turn around to see who was actually working on them. [...] I found myself shouting in two directions from the centre of the room. It didn't make much sense."*

Interestingly, **P10** reported greater reliance on face-to-face engagement when using the baseline dashboard. In contrast, the data-centric visualisations in the full dashboard made them less inclined to approach individual students, as they already had a clear understanding of the situation.

Notably, half of the teachers believed that personal connection remains essential for catering to students' unique needs, often citing the importance of students' facial expressions and individual learning circumstances. For example, **P1** highlighted that outlier classification based on performance alone may overlook individual circumstances and growth. Furthermore, **P8** described how technology cannot fully replace *"the human factor"*, which encompasses the impact of the teacher's presence, their ability to explain concepts, and their observations of students' facial expressions. **P9** indicated this might be *"especially [relevant] for [students] who haven't always had good experiences with maths."*

Moving forward, **P7** recognised the potential benefits of technology in distance learning, but considered personal interaction more logical and practical in traditional classroom settings. **P1** added: *"It's the combination of a good teacher and system that together ensure a good lesson."*

6 Discussion

Comparing groups A and B, we did not observe major ordering effects between the full and baseline dashboards, neither regarding trust nor intention to use. As a basis for our discussion, we could therefore reliably map all themes to the three-layered trust framework by Hoff and Bashir [17] (see Section 2.3).

6.1 Teachers Already Start Calibrating Trust Before Dashboard Use

Addressing RQ1, Figure 6 shows that teachers raised themes related to dispositional, situational, and learned trust, spanning all layers of the framework by Hoff and Bashir [17]. This illustrates the complexity of how teachers calibrate their trust in an outlier detection dashboard. It is interesting to realise that trust calibration starts even before teachers interact with monitoring dashboards. For example, teachers' initial trust was influenced by their disposition towards technology. Some teachers alluded to blindly trusting the detection system, believing that its objectivity and data-driven nature would yield accurate outlier predictions. This blind trust has been documented previously: people may place greater trust in automated systems than in their own judgments [21, 24] and may be less forgiving when they make mistakes [14]. Other teachers,

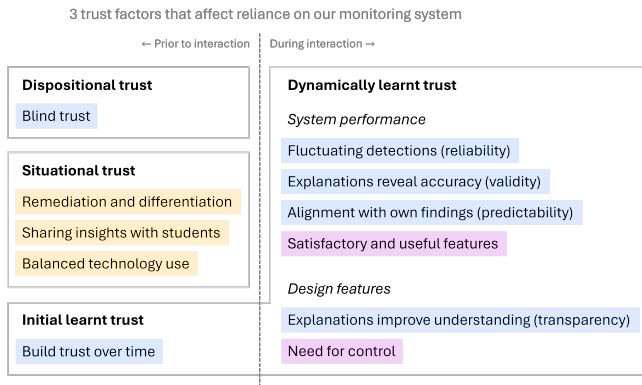


Figure 6: Mapping the themes from our thematic analysis in Section 5 to the three-layered trust framework by Hoff and Bashir [17].

however, expressed a more cautious attitude towards technology, which predisposed them to distrust the system.

The possibilities teachers described for integrating the monitoring dashboard into their classrooms also hinted at situational trust. Concretely, trust could depend on the context of use. For example, teachers could trust the system for remediation and differentiation, but not for sharing its insights with students.

Finally, some teachers indicated they believed the system would remain effective in subsequent lessons, which is a good example of initial learnt trust. Exposure to the dashboard and observing its performance over time helps build this trust, underscoring the importance of sufficient use time and training.

6.2 Data-Centric Explanations Strongly Influence Trust During Interaction

Continuing to address RQ1, our findings show the variety in dynamically learnt trust, both in terms of system performance and design features. An interesting observation was that teachers typically grew (dis)trust when suggested outliers (dis)aligned with their own observations and knowledge of exercises and students. This resonates with the work of Nourani et al. [30], which emphasises the significance of domain experience in shaping teachers' trust in a system. It also links to the idea that perceived accuracy is a crucial driver of trust calibration [38]. Studies have shown that people adjust their trust levels based on estimated accuracy, increasing trust when model accuracy exceeds their own [37].

Regarding RQ2, our thematic analysis clearly showed that many participants primarily paid attention to the data-centric explanations on the monitoring dashboard, considering them the most vivid and memorable aspect of their experience during the experiment. In contrast, the model-centric explanations received considerably less attention from teachers, even though all were expected to read them to gain an intuitive understanding of the outlier algorithm's inner workings. The data-centric explanations played a crucial role in shaping teachers' trust. By analysing the visualisations and seeing what the system based its detections on, teachers better understood its rationale, even in cases where they realised their initial

expectations regarding student performance were wrong. This is reassuring as the dashboard should not just confirm teachers' beliefs. These findings align with previous research about the importance of meaningful explanations [29] and the positive effect of richer explanations on trust [8]. Even though our quantitative logging results are preliminary, given the small sample, it is interesting to see how this is reflected in larger agreement with outlier detections when data-centric explanations were present.

6.3 Classroom Integration Requires Balancing Technology Use and Personal Connections

Answering RQ3, teachers successfully integrated the monitoring dashboard into their classrooms and actively used its visual components and outlier detections. Many reported intentions to use it for remediation, differentiation, and sharing insights with students about their learning process. However, while teachers acknowledged the dashboard's potential, they also admitted to face challenges in simultaneously monitoring it and providing necessary assistance in the classroom. One particular concern was the potential loss of personal connection when using a monitoring dashboard. This is a good reminder that data-centric explanations should not come at the cost of undermining teacher-student relationships, especially because teachers consider a lack of human characteristics as a big limitation of AI-supported educational technologies [28]. However, teachers emphasised that effective lessons can be achieved by combining their skills with reliable systems.

6.4 Design Implications for Outlier Detection Dashboards in Education

Based on our findings, we propose several design implications that could increase reliance on outlier detection dashboards in educational settings. First, to stimulate situational trust, it is important to ensure the dashboard aligns with teachers' goals and the educational context in which they deploy it. Involving teachers in a human-centred design process can help achieve this alignment. Second, given that many teachers indicated using the dashboard while walking around the classroom, monitoring could be supported with portable devices such as tablets or AR devices [19]. Third, incorporating data-centric explanations, such as visual timelines, seems crucial for fostering dynamically learnt trust among teachers and their students. Sharing (parts of) students' raw interaction data, rather than just showing outliers, allows teachers to verify whether outlier detections align with their own findings. Fourth, meeting teachers' need for control can boost their dynamically learnt trust. Thus, we recommend extending outlier detection dashboards with control mechanisms that enable teachers to adjust model parameters and customise visual components. Finally, as fluctuating detections can raise doubts about the system's reliability, smoothing techniques could be applied to flag students as outliers only after the algorithm has detected them for some time. Uncertainty visualisations could also help convey the confidence levels of detections.

6.5 Limitations and Future Work

Our findings are subject to several limitations. First, teachers used our dashboard for a limited time. Conducting similar studies over a

longer period would give teachers more time to assess the system's accuracy and to compare the baseline more extensively with the full dashboard. Second, the study's small sample size limits the generalizability of our quantitative findings. While our logged data showed some interesting initial patterns concerning feature use and agreement with detected outliers, increasing the sample size would allow for more reliable conclusions. Third, the technical simplicity of our outlier model based on speed and attempts does not fully capture the complexity of student performance. Future work could incorporate additional variables and more carefully adjust the speed score and flag outliers, e.g., by considering how far the completion time deviates from the average. Using completion time for outlier detection might also be controversial; our study included it at the teachers' request during design (see Section 3.1). Fourth, there likely was a discrepancy between the recorded and actual interventions, given that teachers had to manually register them while actively moving around the room. Future studies could apply observational or automated approaches to record interventions.

For future work, we recommend providing teachers with control over the system's parameters and visual components. Given the growing prevalence of remote education, investigating the effectiveness of the explainable dashboard in distance learning scenarios would be interesting as well. Additionally, considering long-term performance in the detection system would provide insights into how explanations can support sustained learning and intervention strategies. Lastly, aligned with the endeavour of open learner models [7], it would be relevant to explore how sharing the monitoring dashboard with students could impact their learning outcomes, motivation, and self-regulation.

7 Conclusion

This study extended an adaptive learning platform with a teacher-oriented dashboard for explainable outlier detection of students. We conducted a within-group experiment with 11 teachers to evaluate it in terms of perceived accuracy, trust, usefulness, satisfaction, and classroom integration. Our findings highlighted the multifaceted nature of trust, tapping into dispositional, situational, and learned factors. Crucially, the data-centric *why* explanations, consisting of visual timelines, were teachers' most vivid aspect of their experience, helping them understand outlier detections, fostering trust in the system, and enabling them to verify whether outlier detections aligned with their prior experience with students. Our recommendations for designing outlier detection dashboards and for future research can guide the development of effective and trustworthy educational monitoring tools.

Acknowledgments

We thank all participating teachers and students for using our platform during the in-class experiment. We used GitHub Copilot for minor typesetting and rephrasing support, and Grammarly for minor language improvements.

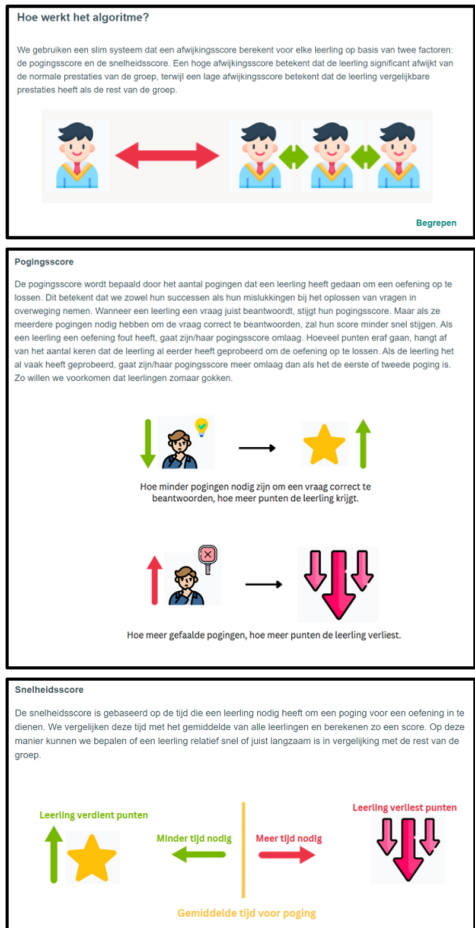
References

- [1] Amina Adadi and Mohammed Berrada. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6 (2018), 52138–52160. doi:10.1109/ACCESS.2018.2870052
- [2] June Ahn, Fabio Campos, Ha Nguyen, Maria Hays, and Jan Morrison. 2021. Co-Designing for Privacy, Transparency, and Trust in K-12 Learning Analytics. In *LAK21: 11th International Learning Analytics and Knowledge Conference*. ACM, Irvine CA USA, 55–65. doi:10.1145/3448139.3448145
- [3] Ariful Islam Anik and Andrea Bunt. 2021. Data-Centric Explanations: Explaining Training Data of Machine Learning Systems to Promote Transparency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3411764.3445736
- [4] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Benetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. 2020. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion* 58 (June 2020), 82–115. doi:10.1016/j.inffus.2019.12.012
- [5] Virginia Braun, Victoria Clarke, Nikki Hayfield, and Gareth Terry. 2018. Thematic Analysis. In *Handbook of Research Methods in Health Social Sciences*, Pranee Liamputtong (Ed.). Springer Singapore, Singapore, 1–18. doi:10.1007/978-981-10-2779-6_103-1
- [6] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. 2000. LOF: Identifying Density-Based Local Outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data (SIGMOD '00)*. Association for Computing Machinery, New York, NY, USA, 93–104. doi:10.1145/342009.335388
- [7] Susan Bull. 2020. There Are Open Learner Models About! *IEEE Transactions on Learning Technologies* 13, 2 (April 2020), 425–448. doi:10.1109/TLT.2020.2978473
- [8] Adrian Bussone, Simone Stumpf, and Dymna O'Sullivan. 2015. The Role of Explanations on Trust and Reliance in Clinical Decision Support Systems. In *2015 International Conference on Healthcare Informatics*. IEEE, Dallas, TX, USA, 160–169. doi:10.1109/ICHL.2015.26
- [9] Carolien Knoop-van Campen and Inge Molenaar. 2020. How Teachers Integrate Dashboards into Their Feedback Practices. *Frontline Learning Research* 8, 4 (July 2020), 37–51. doi:10.14786/flr.v8i4.641
- [10] Irene-Angelica Chounta, Emanuele Bardone, Aet Raudsep, and Margus Pedaste. 2022. Exploring Teachers' Perceptions of Artificial Intelligence as a Tool to Support Their Practice in Estonian K-12 Education. *International Journal of Artificial Intelligence in Education* 32, 3 (Sept. 2022), 725–755. doi:10.1007/s40593-021-00243-5
- [11] Nicholas Diana, Michael Eagle, John Stamper, Shuchi Grover, Marie Bienkowski, and Satabdi Basu. 2017. An Instructor Dashboard for Real-Time Analytics in Interactive Programming Assignments. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. ACM, Vancouver British Columbia Canada, 272–279. doi:10.1145/3027385.3027441
- [12] Finale Doshi-Velez and Been Kim. 2017. Towards A Rigorous Science of Interpretable Machine Learning. *arXiv:1702.08608 [cs, stat]* (March 2017). arXiv:1702.08608 [cs, stat]
- [13] Raphael A. Dourado, Rodrigo Lins Rodrigues, Nivan Ferreira, Rafael Ferreira Mello, Alex Sandro Gomes, and Katrien Verbert. 2021. A Teacher-facing Learning Analytics Dashboard for Process-oriented Feedback in Online Learning. In *LAK21: 11th International Learning Analytics and Knowledge Conference (LAK21)*. Association for Computing Machinery, New York, NY, USA, 482–489. doi:10.1145/3448139.3448187
- [14] Mary T. Dzindolet, Linda G. Pierce, Hall P. Beck, and Lloyd A. Dawe. 2002. The Perceived Utility of Human and Automated Aids in a Visual Detection Task. *Human Factors* 44, 1 (March 2002), 79–94. doi:10.1518/0018720024494856
- [15] Anissa Faik. 2023. *Bringing a New Perspective to the Classroom: Detecting and Explaining Student Outliers*. Master's thesis. KU Leuven. Faculteit Wetenschappen, Leuven.
- [16] Gloria Milena Fernandez Nieto, Kirsty Kitto, Simon Buckingham Shum, and Roberto Martinez-Maldonado. 2022. Beyond the Learning Analytics Dashboard: Alternative Ways to Communicate Student Data Insights Combining Visualisation, Narrative and Storytelling. In *LAK22: 12th International Learning Analytics and Knowledge Conference (LAK22)*. Association for Computing Machinery, New York, NY, USA, 219–229. doi:10.1145/3506860.3506895
- [17] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 57, 3 (May 2015), 407–434. doi:10.1177/0018720814547570
- [18] Robert R. Hoffman, Shane T. Mueller, Gary Klein, and Jordan Litman. 2019. Metrics for Explainable AI: Challenges and Prospects. *arXiv:1812.04608 [cs]* (Feb. 2019). arXiv:1812.04608 [cs]
- [19] Kenneth Holstein, Bruce M. McLaren, and Vincent Aleven. 2019. Co-Designing a Real-Time Classroom Orchestration Tool to Support Teacher-AI Complementarity. *Journal of Learning Analytics* 6, 2 (July 2019). doi:10.18608/jla.2019.62.3
- [20] Hassan Khosravi, Simon Buckingham Shum, Guanliang Chen, Cristina Conati, Yi-Shan Tsai, Judy Kay, Simon Knight, Roberto Martinez-Maldonado, Shazia Sadiq, and Dragan Gašević. 2022. Explainable Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence* 3 (Jan. 2022), 100074. doi:10.1016/

- j.caeai.2022.100074
- [21] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* 46, 1 (March 2004), 50–80. doi:10.1518/hfes.46.1.50_30392
 - [22] Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–15. doi:10.1145/3313831.3376590
 - [23] Steven Lonn, Stephen J. Aguilar, and Stephanie D. Teasley. 2015. Investigating Student Motivation in the Context of a Learning Analytics Intervention during a Summer Bridge Program. *Computers in Human Behavior* 47 (June 2015), 90–97. doi:10.1016/j.chb.2014.07.013
 - [24] P. Madhavan and D. A. Wiegmann. 2007. Similarities and Differences between Human–Human and Human–Automation Trust: An Integrative Review. *Theoretical Issues in Ergonomics Science* 8, 4 (July 2007), 277–301. doi:10.1080/14639220500337708
 - [25] Maria Madsen and Shirley Gregor. 2000. Measuring Human-Computer Trust. In *Proceedings of the 11th Australasian Conference on Information Systems*, Vol. 53. Australasian Association for Information Systems, Brisbane, Australia, 6–8.
 - [26] Sina Mohseni, Niloofar Zarei, and Eric D. Ragan. 2021. A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. *ACM Transactions on Interactive Intelligent Systems* 11, 3-4 (Aug. 2021), 24:1–24:45. doi:10.1145/3387166
 - [27] Tanya Nazaretsky, Carmel Bar, Michal Walter, and Giora Alexandron. 2022. 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*. ACM, Online USA, 1–12. doi:10.1145/3506860.3506861
 - [28] Tanya Nazaretsky, Mutlu Cukurova, and Giora Alexandron. 2022. An Instrument for Measuring Teachers’ Trust in AI-Based Educational Technology. In *LAK22: 12th International Learning Analytics and Knowledge Conference*. ACM, Online USA, 56–66. doi:10.1145/3506860.3506866
 - [29] Mahsan Nourani, Samia Kabir, Sina Mohseni, and Eric D. Ragan. 2019. The Effects of Meaningful and Meaningless Explanations on Trust and Perceived System Accuracy in Intelligent Systems. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 7 (Oct. 2019), 97–105.
 - [30] Mahsan Nourani, Joanie King, and Eric Ragan. 2020. The Role of Domain Expertise in User Trust and the Impact of First Impressions with Intelligent Systems. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 8 (Oct. 2020), 112–121.
 - [31] Jeroen Ooge, Shotallo Kato, and Katrien Verbert. 2022. Explaining Recommendations in E-Learning: Effects on Adolescents’ Trust. In *27th International Conference on Intelligent User Interfaces (IUI ’22)*. Association for Computing Machinery, New York, NY, USA, 93–105. doi:10.1145/3490099.3511140
 - [32] Radek Pelánek. 2016. Applications of the Elo Rating System in Adaptive Educational Systems. *Computers & Education* 98 (July 2016), 169–179. doi:10.1016/j.compedu.2016.03.017
 - [33] T. Spinner, U. Schlegel, H. Schäfer, and M. El-Assady. 2020. Explainer: A Visual Analytics Framework for Interactive and Explainable Machine Learning. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2020), 1064–1074. doi:10.1109/TVCG.2019.2934629
 - [34] Vinitra Swamy, Sijia Du, Mirko Marras, and Tanja Kaser. 2023. Trusting the Explainers: Teacher Validation of Explainable Artificial Intelligence for Course Design. In *LAK23: 13th International Learning Analytics and Knowledge Conference*. ACM, Arlington TX USA, 345–356. doi:10.1145/3576050.3576147
 - [35] Maxwell Szymanski, Martijn Millecamp, and Katrien Verbert. 2021. Visual, Textual or Hybrid: The Effect of User Expertise on Different Explanations. In *26th International Conference on Intelligent User Interfaces*. ACM, College Station TX USA, 109–119. doi:10.1145/3397481.3450662
 - [36] Bo Tang and Haibo He. 2017. A Local Density-Based Approach for Outlier Detection. *Neurocomputing* 241 (June 2017), 171–180. doi:10.1016/j.neucom.2017.02.039
 - [37] Ming Yin, Jennifer Wortman Vaughan, and Hanna Wallach. 2019. Understanding the Effect of Accuracy on Trust in Machine Learning Models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–12. doi:10.1145/3290605.3300509
 - [38] Kun Yu, Shlomo Berkovsky, Ronnie Taib, Jianlong Zhou, and Fang Chen. 2019. Do I Trust My Machine Teammate? An Investigation from Perception to Decision. In *Proceedings of the 24th International Conference on Intelligent User Interfaces (IUI ’19)*. Association for Computing Machinery, New York, NY, USA, 460–468. doi:10.1145/3301275.3302277

A Appendix

Figure 7 shows our model-centric explanations, and Table 1 contains the questions we asked teachers after they used our dashboard.

Figure 7: The model-centric explanations on our learning platform with translated text.

How does the algorithm work? We use an intelligent system that calculates a deviation score for each student based on two factors: the attempt score and the speed score. A high deviation score indicates that the student significantly deviates from the group's normal performance, while a low deviation score indicates that the student performs similarly to the rest of the group.

Attempt score. The attempt score is determined by the number of attempts a student has made to solve an exercise. This means that we consider both their successes and failures when answering questions. When a student answers a question correctly, their attempt score increases. However, if they require multiple attempts to answer the question correctly, their score will increase more slowly. If a student answers an exercise incorrectly, their attempt score decreases. The amount deducted depends on how many times the student has previously attempted to solve the exercise. If the student has attempted it many times before, their attempt score will decrease more than if it is their first or second attempt. This approach aims to prevent students from simply guessing.

Speed score. The speed score is based on the time a student takes to submit an attempt for an exercise. We compare this time with the average of all students and calculate a score. This allows us to determine whether a student is relatively fast or slow compared to the rest of the group.

Table 1: Questions asked in the semi-structured interviews with teachers.

English Translation	Dutch Original
<i>General Introductory Questions</i>	
In what order did you see the dashboard?	In welke volgorde heb jij het dashboard gezien?
What were your first impressions of the detection system?	Wat waren jouw eerste indrukken van het detectiesysteem?
How did you use it during the practice session?	Hoe heb je het gebruikt tijdens de oefensessie?
Did you receive different feedback from students compared to a typical lesson?	Kreeg je van leerlingen andere respons dan tijdens een gebruikelijke les?
Which insights did you find most interesting and why?	Welke inzichten vond je het interessantst en waarom?
Which insights encourage you to use interventions?	Welke inzichten moedigen jou aan om interventies in te zetten?
Which insights do you feel were missing?	Welke inzichten miste volgens jou nog?
<i>Trust</i>	
Do you feel that you can trust the detection system?	Heb je het gevoel dat je het detectiesysteem kunt vertrouwen?
<i>Perceived Reliability</i>	
Can you describe how reliable you find the detection system?	Kun je beschrijven hoe betrouwbaar je het detectiesysteem vindt?
According to you, what are the important indicators of the system's reliability?	Wat zijn volgens jou belangrijke indicatoren van de betrouwbaarheid van het systeem?
To what extent do you trust that the system functions correctly?	In hoeverre vertrouwt u erop dat het systeem correct functioneert?
<i>Perceived Understandability</i>	
Can you tell me how the detection system works?	Kun je me vertellen hoe het detectiesysteem werkt?
Where did you find information about how the detection system works?	Waar heb je informatie gevonden over hoe het detectiesysteem werkt?
Would you like more detailed explanations about how the detection system works? If so, what kind of information would you like?	Zou je graag meer gedetailleerde uitleg hebben over hoe het detectiesysteem werkt? Zo ja, wat voor soort informatie zou je graag willen hebben?
Can you describe how the detection system helps you make decisions and whether this is understandable to you?	Kun je beschrijven hoe het detectiesysteem je helpt bij het nemen van beslissingen en of dit voor jou begrijpelijk is?
What do you find easy or difficult to understand about the behaviour of the detection system?	Wat vind je gemakkelijk of moeilijk te begrijpen aan het gedrag van het detectiesysteem?
<i>Perceived Accuracy</i>	
How accurate do you find the system's detections?	Hoe nauwkeurig vind je de detecties van het systeem?
To what extent do you rely on the accuracy of the system and why?	In hoeverre vertrouwt u op de nauwkeurigheid van het systeem en waarom?
<i>Faith</i>	
To what extent do you have faith in the detections of the system, even if you are not sure whether they are correct?	In hoeverre heb je vertrouwen in de detecties van het systeem, zelfs als je niet zeker weet of ze correct zijn?
<i>Satisfaction</i>	
Do you understand how the detection system works based on the explanations?	Begrijp je hoe het detectiesysteem werkt op basis van de uitleg?
Are you satisfied with the explanations about the detection system?	Ben je tevreden met de uitleg over het detectiesysteem?
Does the explanation help you evaluate when you can trust the detection system?	Helpt de uitleg je om te beoordelen wanneer je het detectiesysteem kunt vertrouwen?
<i>Effectivity</i>	
Does the explanation help you know how to use the detection system to achieve your goals?	Helpt de uitleg je om te weten hoe je het detectiesysteem kunt gebruiken om je doelen te bereiken?
<i>General Concluding Questions</i>	
Have you read the general explanation about the detection system? Does this help you understand the detection system?	Heb je de algemene uitleg over het detectiesysteem gelezen? Helpt dit bij het begrijpen van het detectiesysteem?
What did you generally think of the dashboard?	Wat vond je in het algemeen van het dashboard?