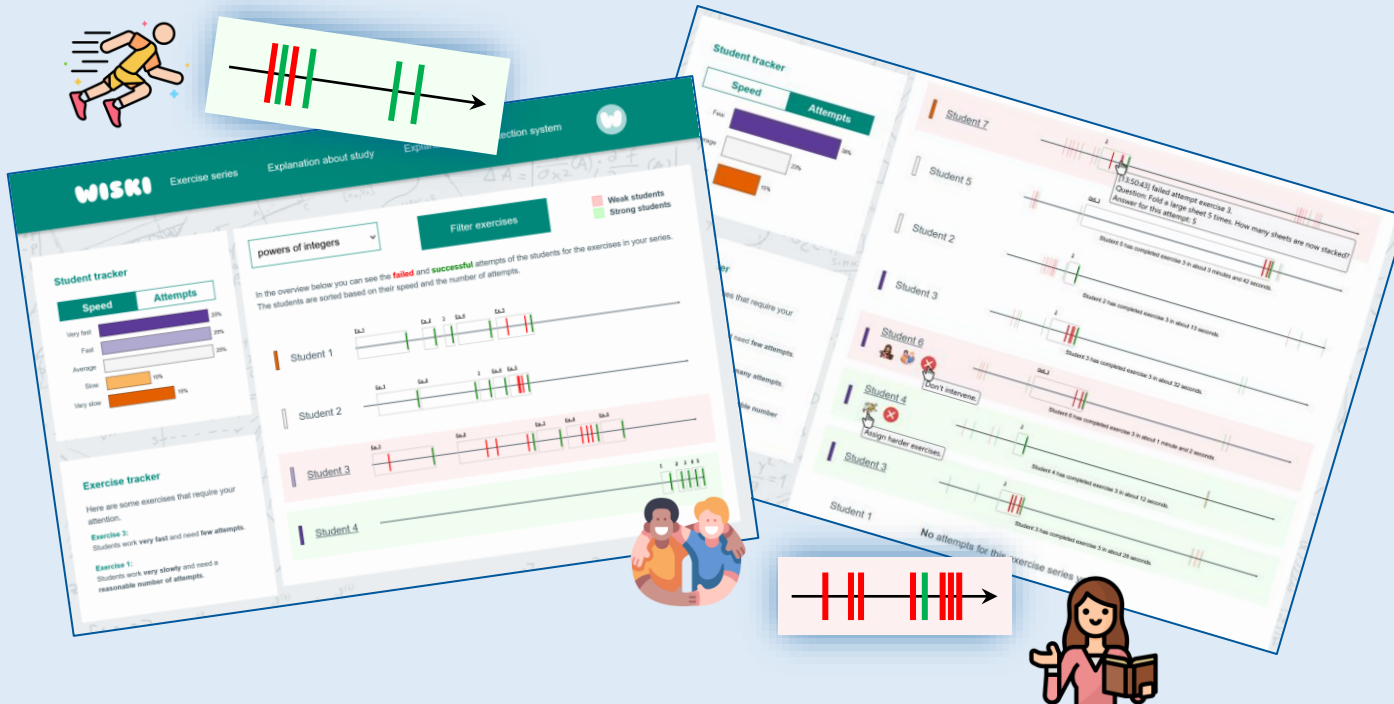


# Detect, Explain, Act

## How Teachers Trust and Use an Explainable Real-Time Monitoring Dashboard to Detect Student Outliers in Class



Jeroen Ooge\*  
[jeroenooge.be](http://jeroenooge.be)



Anissa Faik\*



Katrien Verbert  
[augment.cs.kuleuven.be](http://augment.cs.kuleuven.be)



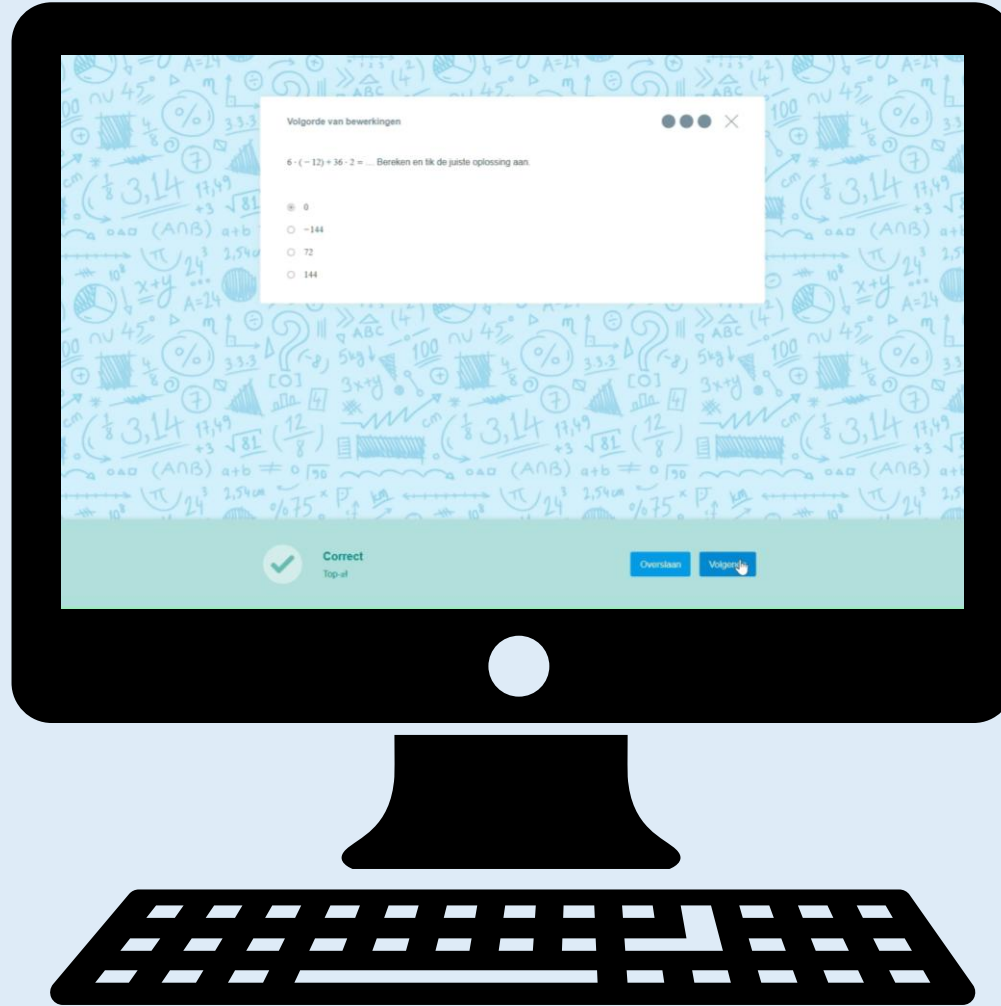
expert



competent



beginner



Learning platform for secondary school maths  
that recommends exercise series



# Our earlier research ...

**Hoe is je nieuw niveau bepaald?**

Wiski schat jouw niveau en de moeilijkheid van oefeningen in. Beide veranderen bij het oplossen van oefeningen. Je niveau is gestegen na het maken van de reeks oefeningen. Daarna is het nog extra gestegen door je feedback.

Maak meer oefeningen over dit onderwerp    Ga terug naar oefenpagina

Trust effects of explaining recommended exercises

<https://doi.org/10.1145/3490099.3511140>

Trust effects of steering recommended exercises

<https://doi.org/10.1145/3581641.3584046>

**Goed gewerkt!**

**Maak een aangeraden oefening van hetzelfde hoofdstuk**

Aangeraden

- Oefening 4
- Oefening 5
- Oefening 7

**Waarom deze oefening?** Wiski denkt dat jouw huidige niveau past bij dat van deze oefening!

Wiski verwacht dat je **1 of 2 pogingen** nodig gaat hebben om oefening 4 juist te maken, gebaseerd op de resultaten van jou en je medeleerlingen.

Aantal pogingen medeleerlingen nodig hadden om oefening 4 juist op te lossen

Maak oefening 4

... of kies zelf je volgende oefening

Naar het oefeningenoverzicht

Trust effects of explaining recommended exercises

<https://doi.org/10.1145/3490099.3511140>

**gevorderde beginner** Volgens mij is dit nu je level voor het onderwerp *Volgorde van bewerkingen*

Welke moeilijkheidsgraad wil je voor de volgende oefeningenreeks?

Heel makkelijk    Makkelijk    Gewoon    Moeilijk    Heel moeilijk

Als je alle oefeningen in de reeks juist oplost, dan stijgt je level:

- Expert
- Bedreven
- Competent
- Gevorderde beginner
- Beginner

Je level na de reeks

Je level nu

Start de reeks

Combining explanations and control over difficulty

<https://doi.org/10.1145/3706468.3706470>

# ... did not focus on teachers

### Hoe is je nieuw niveau bepaald?

Wiski schat jouw niveau en de moeilijkheid van oefeningen in. Beide veranderen bij het oplossen van oefeningen. Je niveau is gestegen na het maken van de reeks oefeningen. Daarna is het nog extra gestegen door je feedback.

Expert  
Bedreven  
Competent  
Gevorderde beginner  
Beginner

Voor reeks Na reeks Na feedback

oefeningen over dit onderwerp

### Goed gewerkt!

#### Maak een aangeraden oefening van hetzelfde hoofdstuk

Aangeraden  
Oefening 4  
Oefening 5  
Oefening 7

Waarom deze oefening? Wiski denkt dat jouw huidige niveau past bij dat van deze oefening! Wiski verwacht dat je 1 of 2 pogingen nodig gaat hebben om oefening 4 juist te maken, gebaseerd op de resultaten van jou en je medeleerlingen.

Oefening	Aantal pogingen
1	1
2	2
3	3
4	4
5	5
6	6
7	7

Max oefening 1

... of kies zelf je volgende oefening

Naar het oefeningenoverzicht

gevorderde beginner Volgens mij is dit nu je level voor het onderwerp. Volgde van bewerkingen

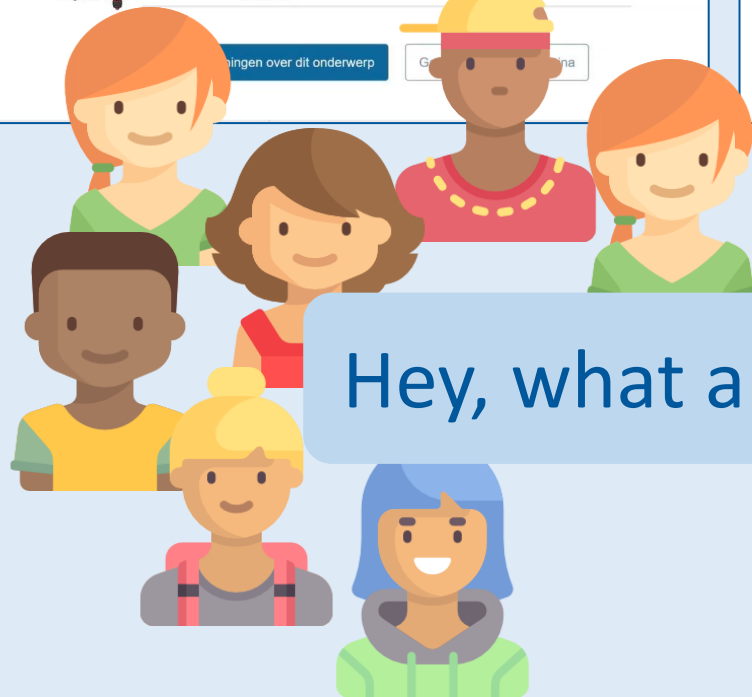
Welke moeilijkheidsgraad wil je voor de volgende oefeningenreeks?

Heel makkelijk Makkelijk Gewoon Moeilijk Heel moeilijk

Als je alle oefeningen in de reeks juist oplost, dan stijgt je level:

Expert

Na de reeks

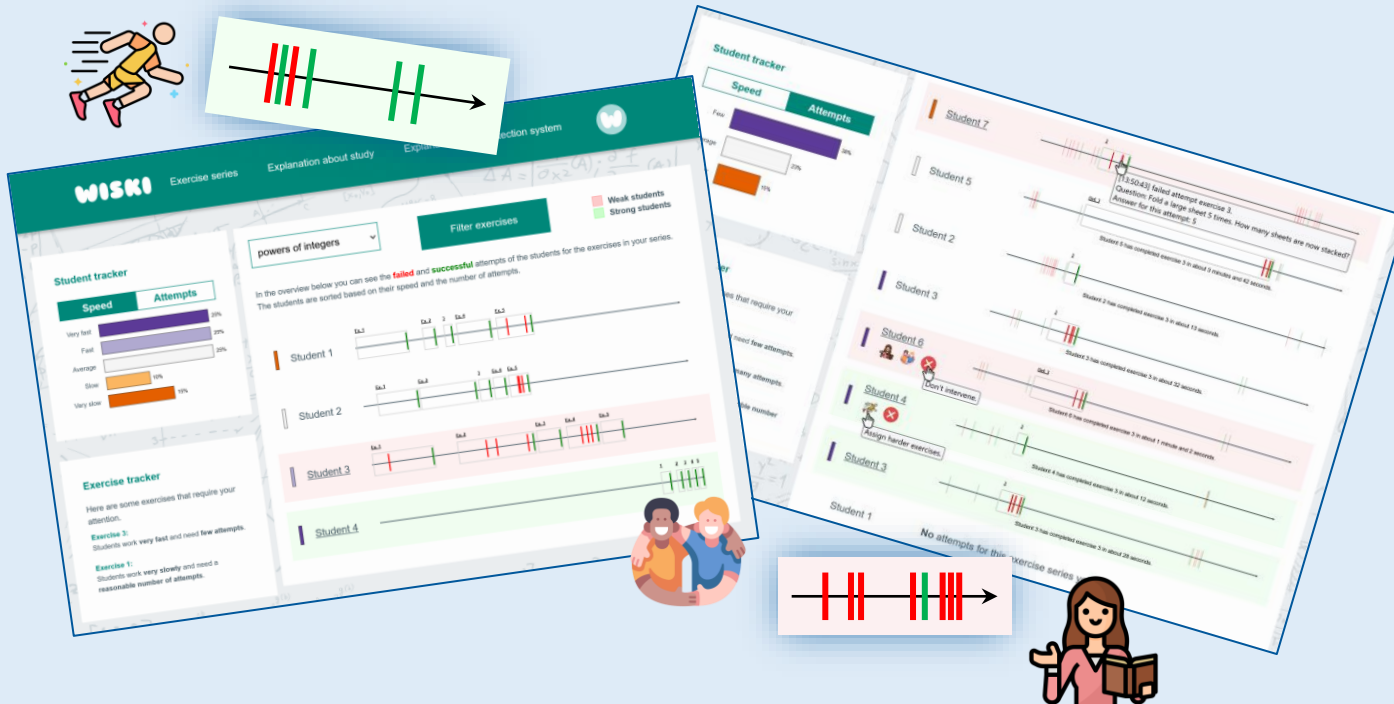


Hey, what about us?



# Detect, Explain, Act

## How Teachers Trust and Use an Explainable Real-Time Monitoring Dashboard to Detect Student Outliers in Class



Jeroen Ooge\*  
[jeroenooge.be](http://jeroenooge.be)



Anissa Faik\*



Katrien Verbert  
[augment.cs.kuleuven.be](http://augment.cs.kuleuven.be)

# Co-design with teachers ...

Detect

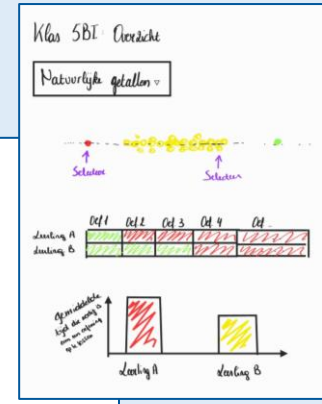
Teachers want a **warning system** that focuses on individual students and recommends **actions**

Act

The collage displays various educational analytics dashboards. Key components include:

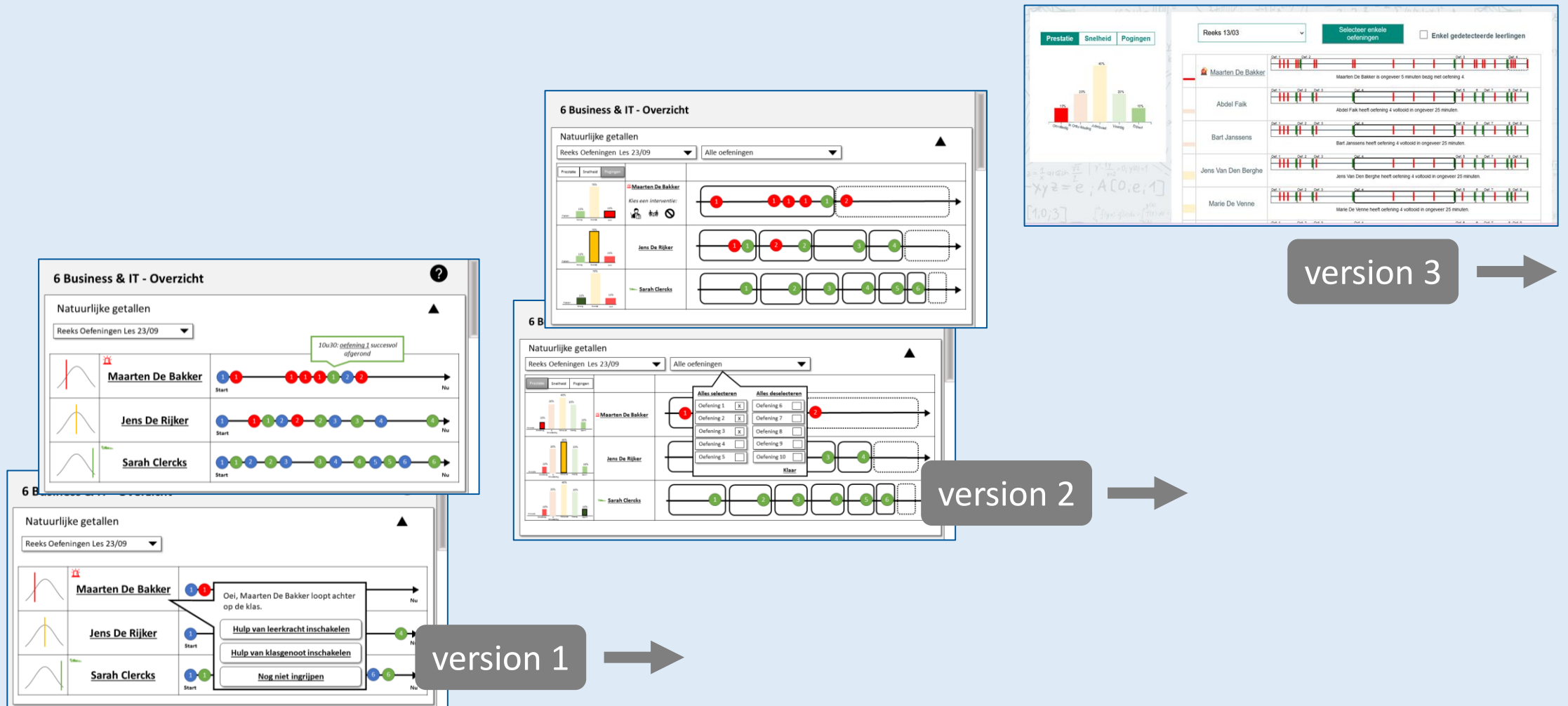
- Performance Overview:** A bar chart showing student performance across different categories.
- Pattern Discovery:** A funnel chart showing the flow of students through different stages of a course.
- Student List:** A table listing students with their names, IDs, and status indicators (e.g., 'NO', 'YES').
- Classroom (n=30):** A matrix showing predicted scores and accuracy for a group of 30 students.
- Selected Student (3010):** A detailed view of a specific student's performance, including a timeline and code state.
- Summary for McIntosh II, Barton:** A bar chart comparing student and class percentages for different courses.
- Class Performance Table:** A table showing class performance over time, with columns for Class, Week, Student Percentage, Class Percentage, and Action.

Class	Week	Student Percentage	Class Percentage	Action
CSP 100 Su13	8/24/2013	82 %	90 %	Encourage
MATH 103 Su13	8/24/2013	30 %	68 %	Engage
WRITING 100 Su13	8/24/2013	56 %	71 %	Engage

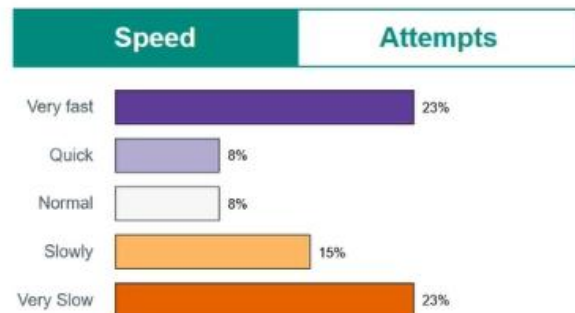


Knoop-van Campen et al. 2020. How Teachers integrate Dashboards into their Feedback Practices. <https://doi.org/10.14786/flr.v8i4.641>  
 Diana et al. 2017. An instructor dashboard for real-time analytics in interactive programming assignments. <https://doi.org/10.1145/3027385.3027441>  
 Dourado et al. 2021. A Teacher-facing Learning Analytics Dashboard for Process-oriented Feedback in Online Learning. <https://doi.org/10.1145/3448139.3448187>  
 Lonn et al. 2015. Investigating student motivation in the context of a learning analytics intervention... <https://doi.org/10.1016/j.chb.2014.07.013>

# ... and HCI experts in multiple iterations



## Student Tracker



## Exercise Tracker

Here are some exercises that require your attention.

### Exercise 5:

Students work **very slowly** and need **few attempts**.

### Exercise 1:

Students work **very fast** and need **a lot of attempts**.

### Exercise 3:

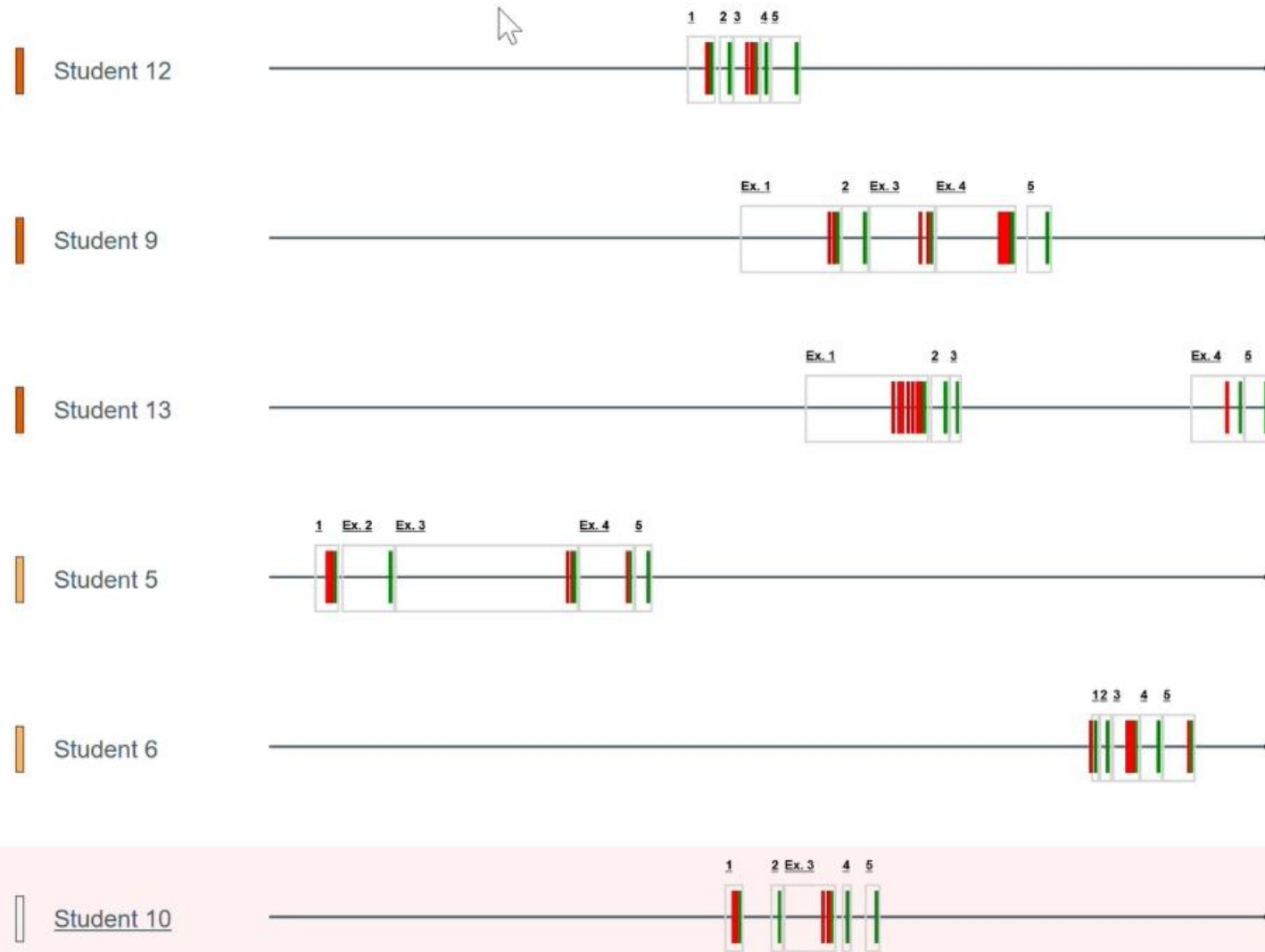
Students work **quickly** and need a **reasonable number of attempts**.

powers of integers

Filter exercises

Weak students  
Strong students

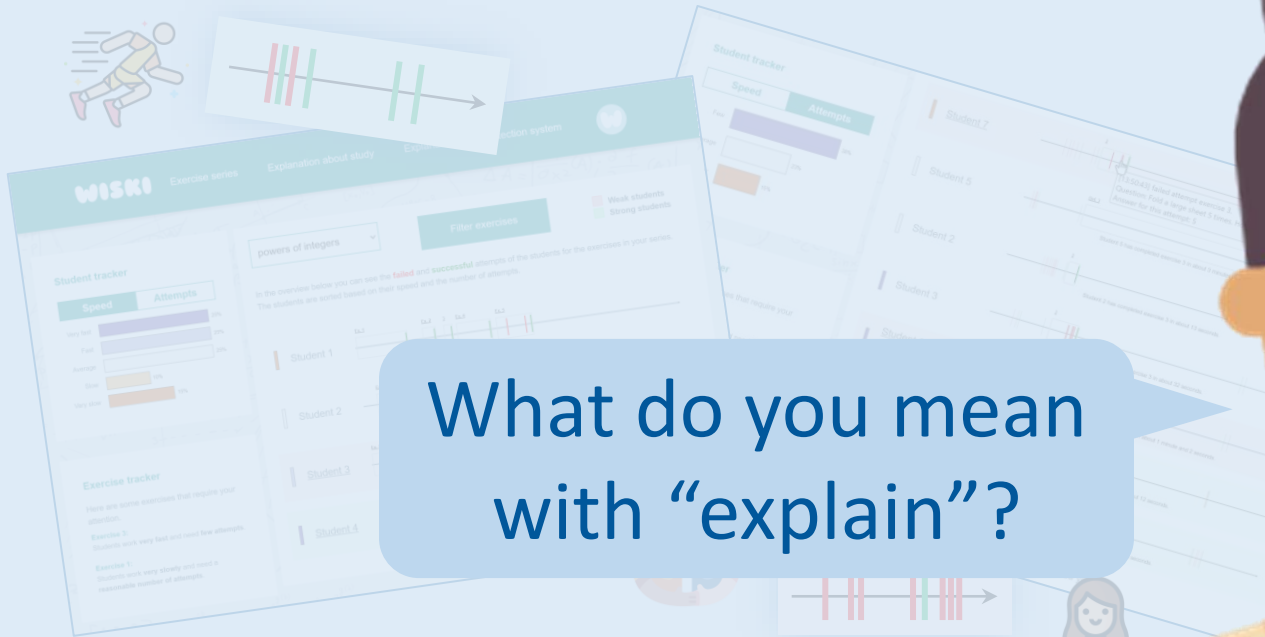
In the overview below you can see the **failed** and **successful** attempts of the students for the exercises from your series. The students are sorted according to their speed and the number of attempts.



final version

# Detect, Explain, Act

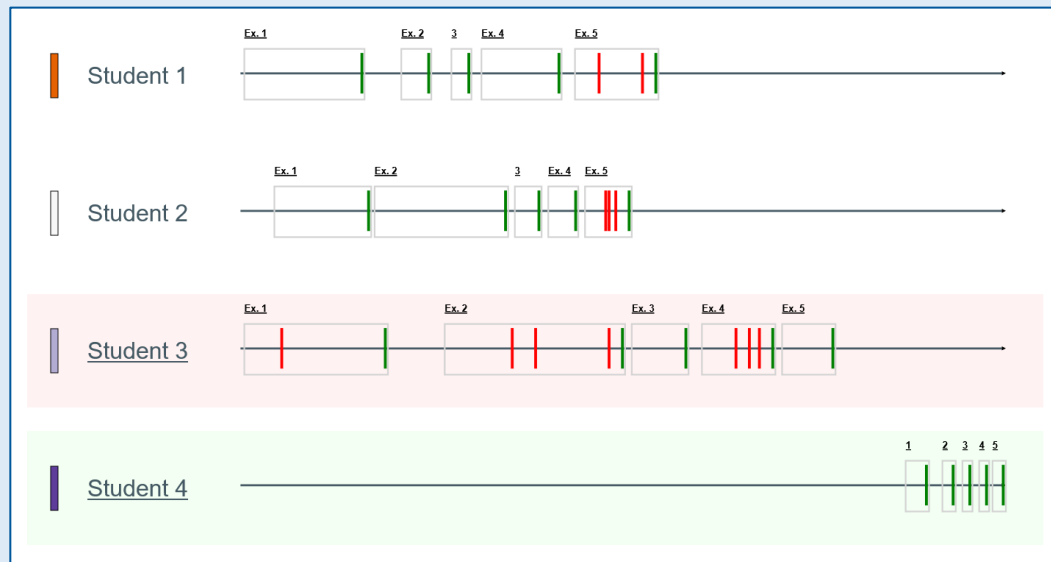
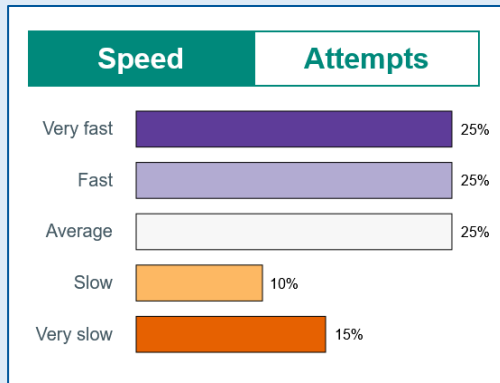
How Teachers Trust and Use an Explainable Real-Time Monitoring Dashboard to Detect Student Outliers in Class



What do you mean with “explain”?



## Data-centric *why* explanations visualise speed and attempt scores



## Model-centric *how* explanations clarify the model's internal functioning

**Hoe werkt het algoritme?**

We gebruiken een slim systeem dat een afwijkingsscore berekent voor elke leerling op basis van twee factoren: de pogingsscore en de snelheidsscore. Een hoge afwijkingsscore betekent dat de leerling significant afwijkt van de normale prestaties van de groep, terwijl een lage afwijkingsscore betekent dat de leerling vergelijkbare prestaties heeft als de rest van de groep.

**Snelheidsscore**

De snelheidsscore is gebaseerd op de tijd die een leerling nodig heeft om een poging voor een oefening in te dienen. We vergelijken deze tijd met het gemiddelde van alle leerlingen en berekenen zo een score. Op deze manier kunnen we bepalen of een leerling relatief snel of juist langzaam is in vergelijking met de rest van de groep.

aantal pogingen dat een leerling heeft gedaan om een oefening op te lossen als hun mislukkingen bij het oplossen van vragen in een vraag juist beantwoordt, stijgt hun pogingsscore. Maar als ze een vraag correct te beantwoorden, zal hun score minder snel stijgen. Als zijn/haar pogingsscore omlaag. Hoeveel punten eraf gaan, hangt af der heeft geprobeerd om de oefening op te lossen. Als de leerling het al vaak heeft geprobeerd, gaat zijn/haar pogingsscore meer omlaag dan als het de eerste of tweede poging is. Zo willen we voorkomen dat leerlingen zomaar gokken.

Hoe minder pogingen nodig zijn om een vraag correct te beantwoorden, hoe meer punten de leerling krijgt.

Hoe meer gefaalde pogingen, hoe meer punten de leerling verliest.

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?

# In-class experiment with 11 teachers

**Student tracker**

Speed Attempts

Very fast 25%  
Fast 25%  
Average 25%  
Slow 10%  
Very slow 15%

**Exercise tracker**

Here are some exercises that require your attention.

**Exercise 3:**  
Students work very fast and need few attempts.

**full**

powers of integers

Filter exercises

Weak students  
Strong students

In the overview below you can see the failed and successful attempts of the students for the exercises in your series. The students are sorted based on their speed and the number of attempts.

Student 1

Student 2

Student 3

Student 4

Use full dashboard (25 min) and baseline dashboard (15 min) (counterbalanced)

Semi-structured interviews about trust and effectiveness of use

**Exercise tracker**

Here are some exercises that require your attention.

**Exercise 4**  
**Exercise 8**  
**Exercise 7**  
**Exercise 6**

**baseline**

**Student tracker**

power of integers

Weak students  
Strong students

Below you find an overview of all students.

Student 1

Student 2

Student 3

Student 4

# Themes in the interviews

## Trust

Blind trust

Fluctuating detections (reliability)

Explanations reveal accuracy (validity)

Alignment with own findings (predictability)

Build trust over time

## Satisfaction and Usefulness

Satisfactory and useful features

Need for control

Explanations improve understanding (transparency)

## Dashboard Integration in Classrooms

Remediation and differentiation

Sharing insights with students

Balanced technology use

# Themes mapped to trust framework

## Dispositional trust

Blind trust

## Situational trust

Remediation and differentiation

Sharing insights with students

Balanced technology use

## Initial learnt trust

Build trust over time

## Dynamically learnt trust

*System performance*

Fluctuating detections (reliability)

Explanations reveal accuracy (validity)

Alignment with own findings (predictability)

Satisfactory and useful features

*Design features*

Explanations improve understanding (transparency)

Need for control

# RQ1: Which factors affect teachers' trust?

← Prior to interaction

During interaction →

## Dispositional trust

Blind trust

## Situational trust

Remediation and differentiation

Sharing insights with students

Balanced technology use

## Initial learnt trust

Build trust over time

## Dynamically learnt trust

*System performance*

Fluctuating detections (reliability)

Explanations reveal accuracy (validity)

Alignment with own findings (predictability)

Satisfactory and useful features

*Design features*

Explanations improve understanding (transparency)

Need for control

# RQ1: Which factors affect teachers' trust?

← Prior to interaction

**Dispositional trust**

Blind trust

Situational trust

Remediation and differentiation

Sharing insights with students

Balanced technology use

Initial learnt trust

Build trust over time

During interaction →

Dynamically learnt trust

System performance

Fluctuating trust

Explanation

Alignment with own findings (predictability)

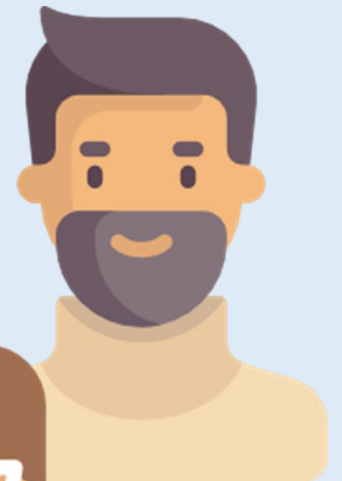
Satisfactory and useful features

Explanations improve understanding (transparency)

Need for control

"I trust it [...] because there's an objective system that generates judgments based on data."

"I'm not naturally inclined to trust the computer for [detecting outliers]."



# RQ1: Which factors affect teachers' trust?

← Prior to interaction

Dispositional trust

Blind trust

**Situational trust**

Remediation and differentiation

Sharing insights with students

Balanced technology use

Initial learnt trust

Build trust over time

During interaction →

Dynamically learnt trust

*System performance*

Fluctuating detections (reliability)

Explanations reveal accuracy (validity)

Alignment with own findings (predictability)

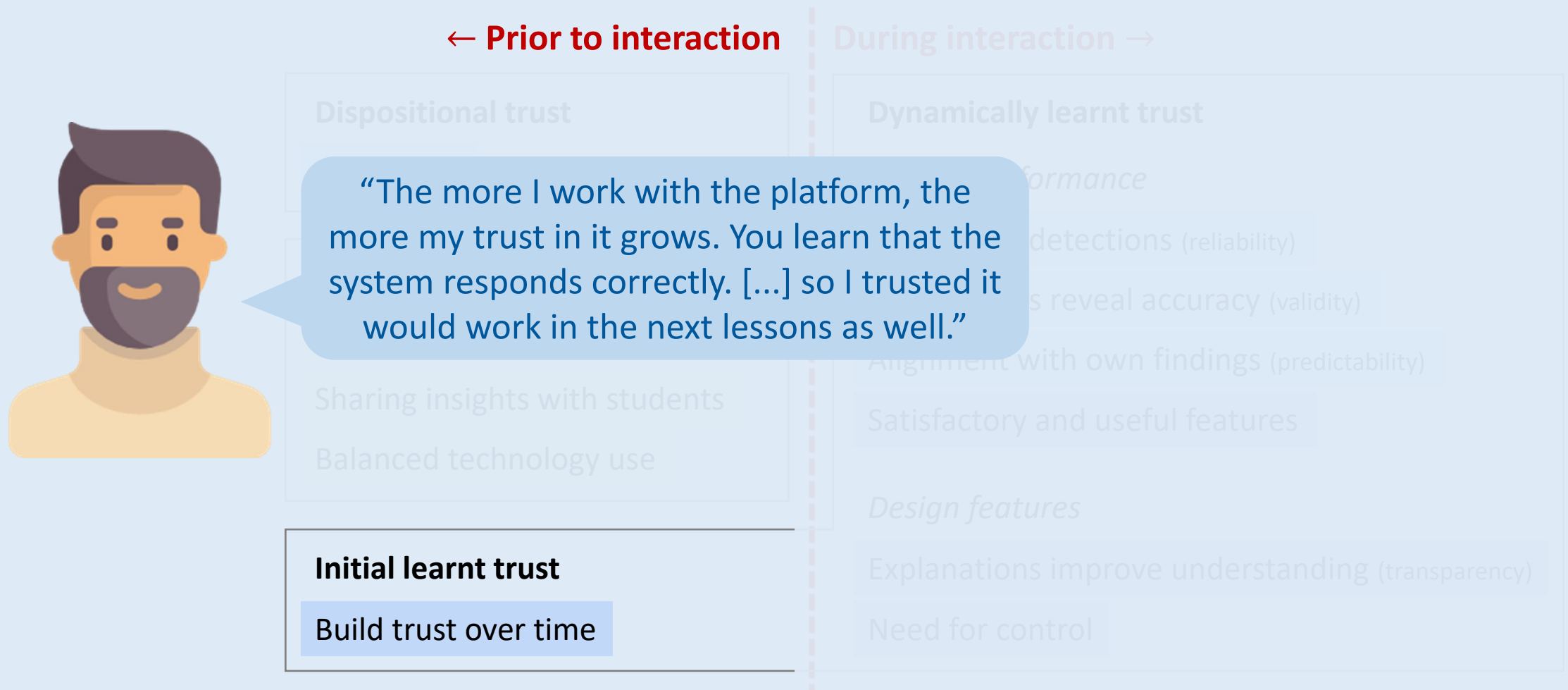
Satisfactory and useful features

*Design features*

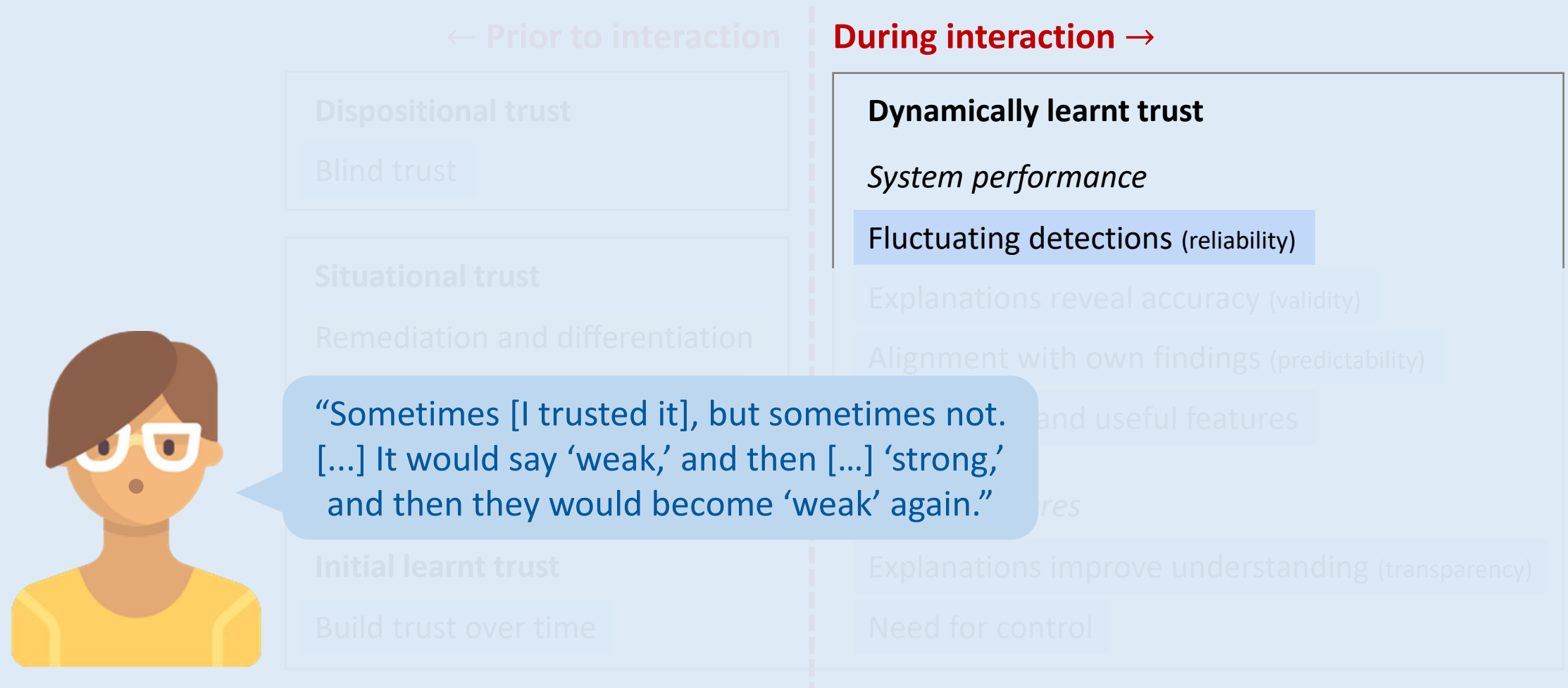
Explanations improve understanding (transparency)

Need for control

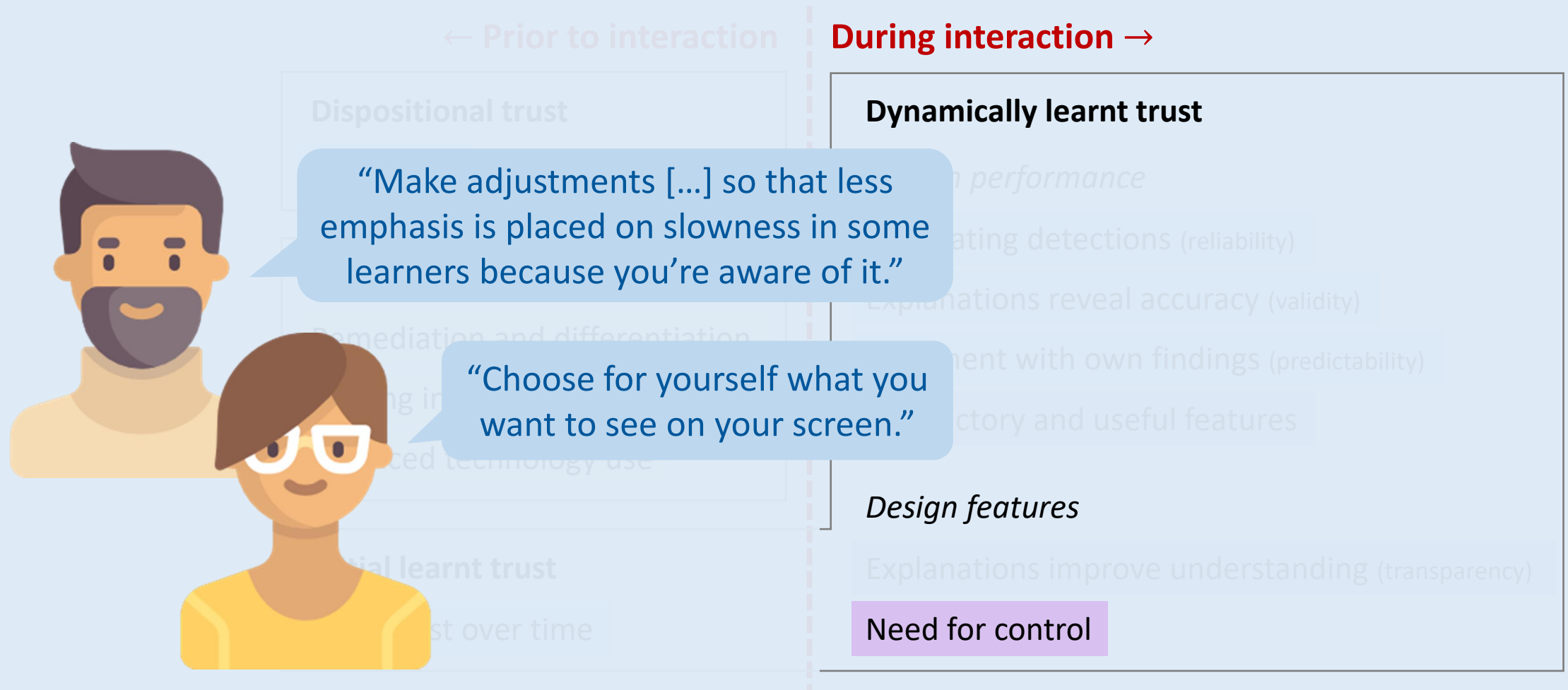
# RQ1: Which factors affect teachers' trust?



# RQ1: Which factors affect teachers' trust?

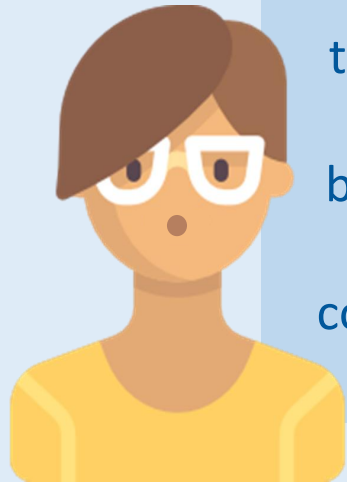


# RQ1: Which factors affect teachers' trust?



# RQ2: What is the impact of explanations?

**Model-centric** explanations can give a basic understanding of the algorithm



“I noticed that students whom I know are weaker were being highlighted as stronger. And I thought, ‘Really?’ But [...] it was true [...]. 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.”

## **Dynamically learnt trust**

*System performance*

*Fluctuating detections (reliability)*

**Explanations reveal accuracy (validity)**

*Alignment with own findings (predictability)*

*Satisfactory and useful features*

*Design features*

**Explanations improve understanding (transparency)**

*Need for control*

# RQ2: What is the impact of explanations?

Many teachers mainly paid attention to the **data-centric** explanations



“I could see on the timeline that they were guessing. [...] So, I trusted that the system was identifying the students correctly.”

Dispositional trust

Blind trust

Situational trust

Initial learnt trust

Build trust over time

**Dynamically learnt trust**

*System performance*

*Fluctuating detections (reliability)*

**Explanations reveal accuracy (validity)**

*Alignment with own findings (predictability)*

*Satisfactory and useful features*

*Design features*

**Explanations improve understanding (transparency)**

*Need for control*

# RQ3: How do teachers use the dashboard?

Dispositional trust

Blind trust

**Situational trust**

Remediation and differentiation

Sharing insights with students

Balanced technology use

Initial learnt trust

Build trust over time

Dynamically learnt trust

System performance

Alignment with own findings (predictability)

Satisfactory and useful features

Design

Explanation

Need for control

“I would love to use it to differentiate tasks.”

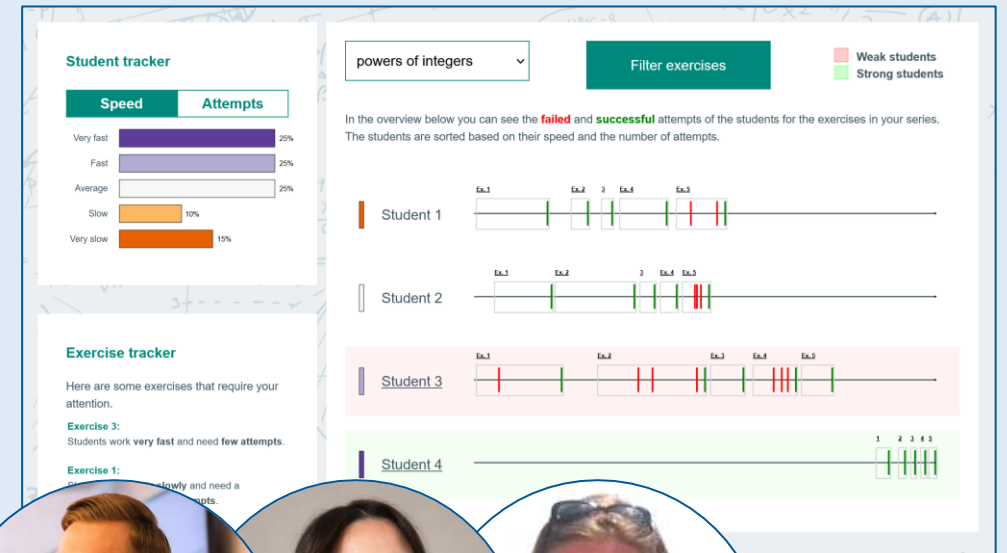
“I could say to them, ‘Hey, the system recognises you as a strong student’, and you could see them light up with pride.”

“It’s difficult to monitor the dashboard and walk around at the same time.”



# Design implications for outlier detection dashboards

1. Align with teachers' goals and educational context
2. Support portable devices
3. Include (data-centric) explanations
4. Provide control mechanisms
5. Avoid fluctuating detections



Find the paper, demo, and slides on [jeroenooge.be/lak26](https://jeroenooge.be/lak26)

