Explainability and Control for Adaptive E-Learning Systems

RECIFE 2024

Jeroen Ooge

jeroenooge.be

Universiteit Utrecht

Input

AI model

Output

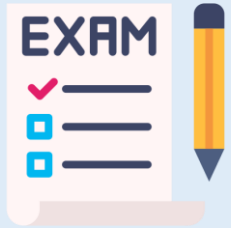
Available exercises

Recommended exercise

Learner profile

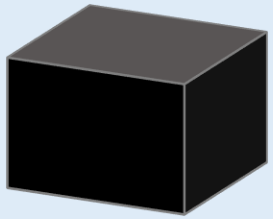
Predicted progress/risk

Textual prompt



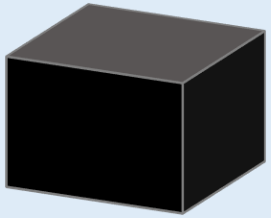
Exam question

Available exercises



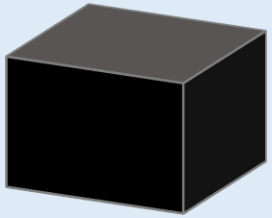
Recommended exercise

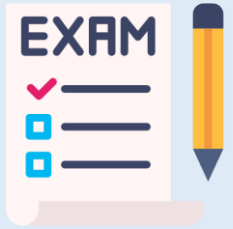
Learner profile



Predicted progress/risk

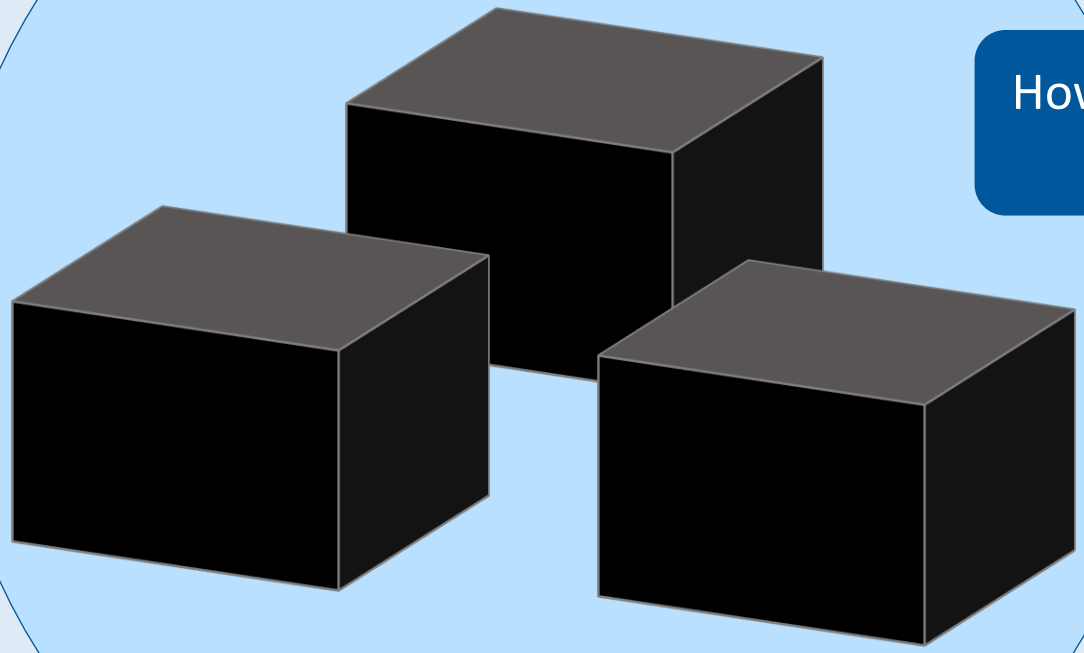
Textual prompt





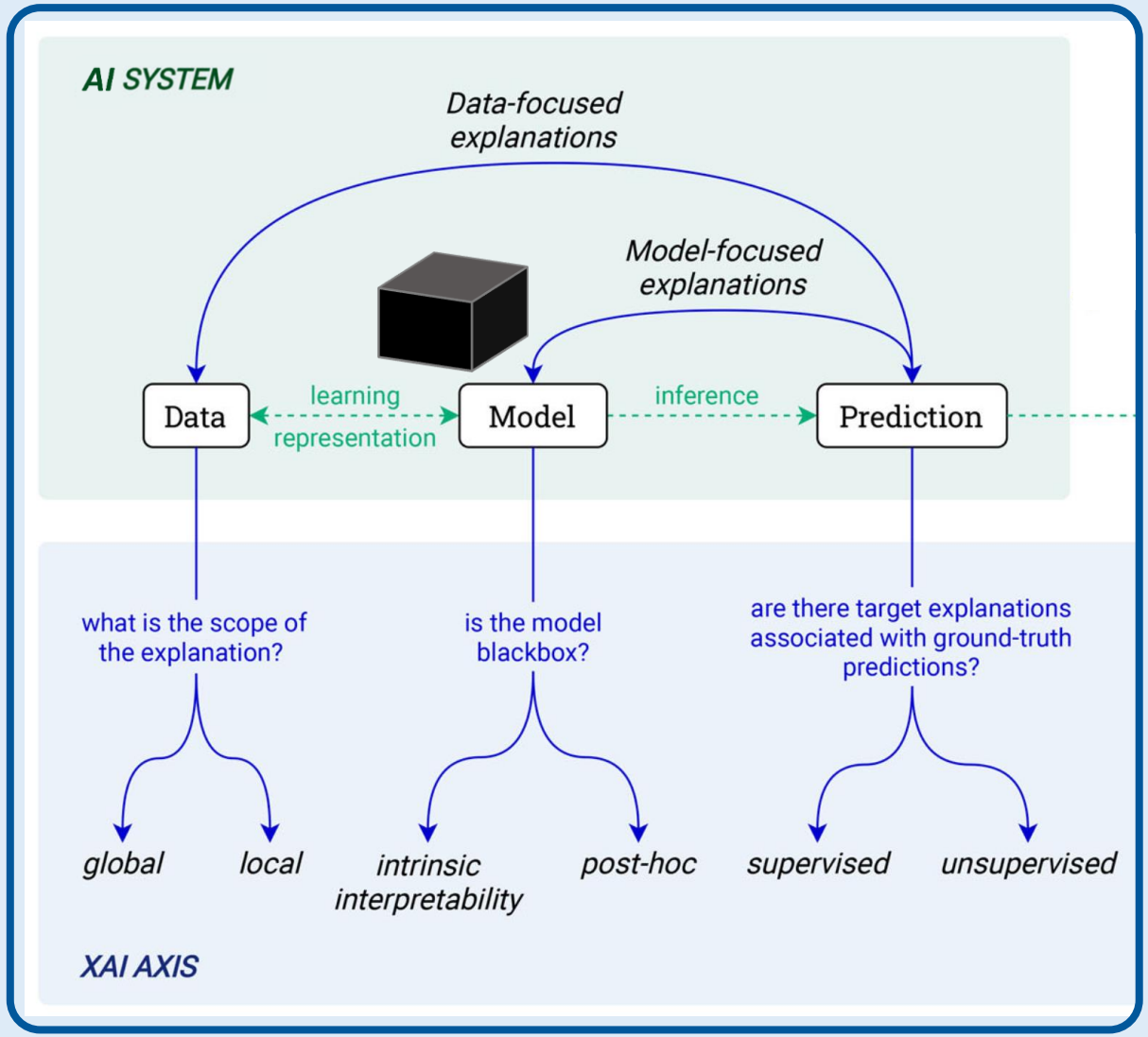
Exam question

Explanations



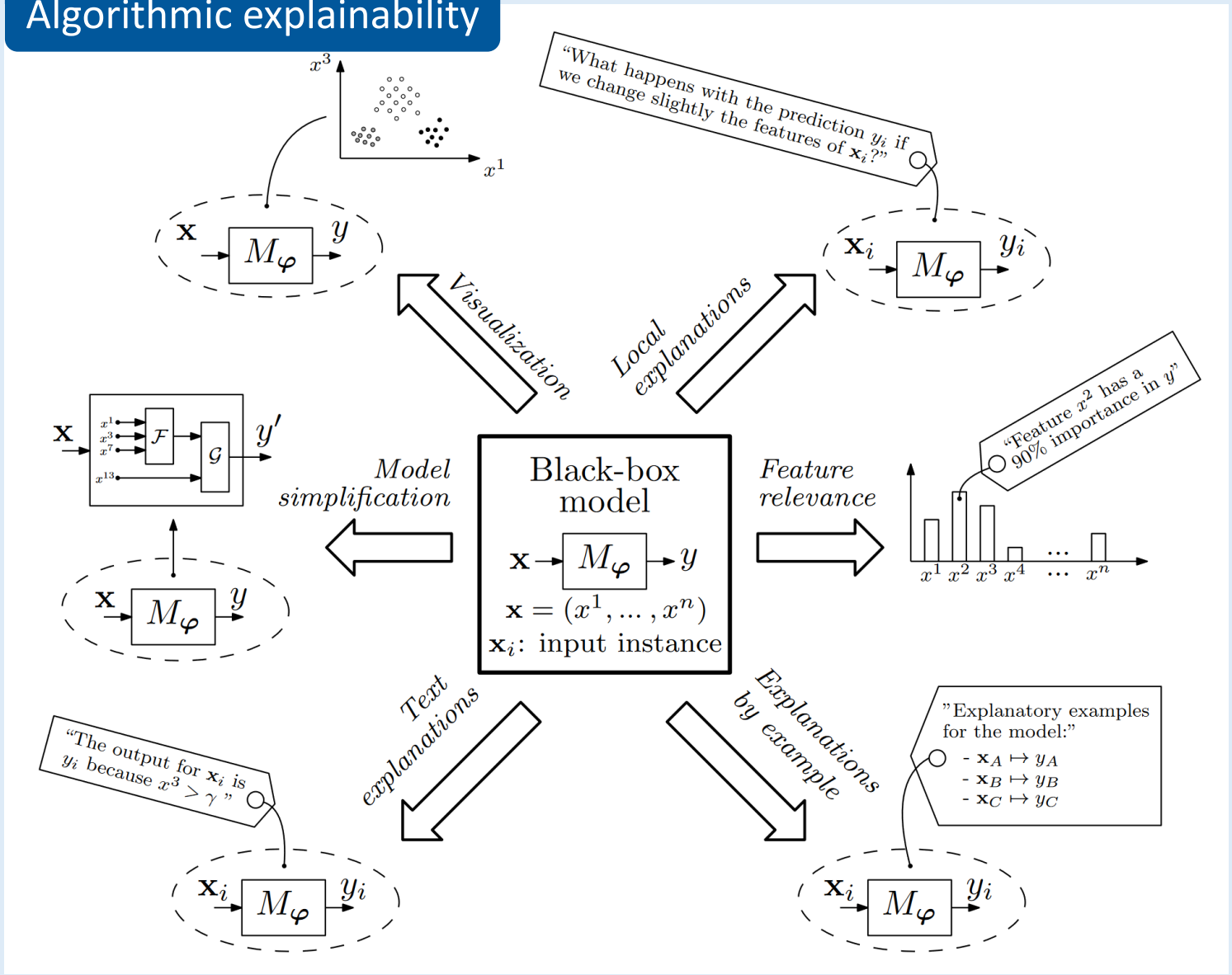
How can we generate explanations for these black boxes?

Algorithmic explainability



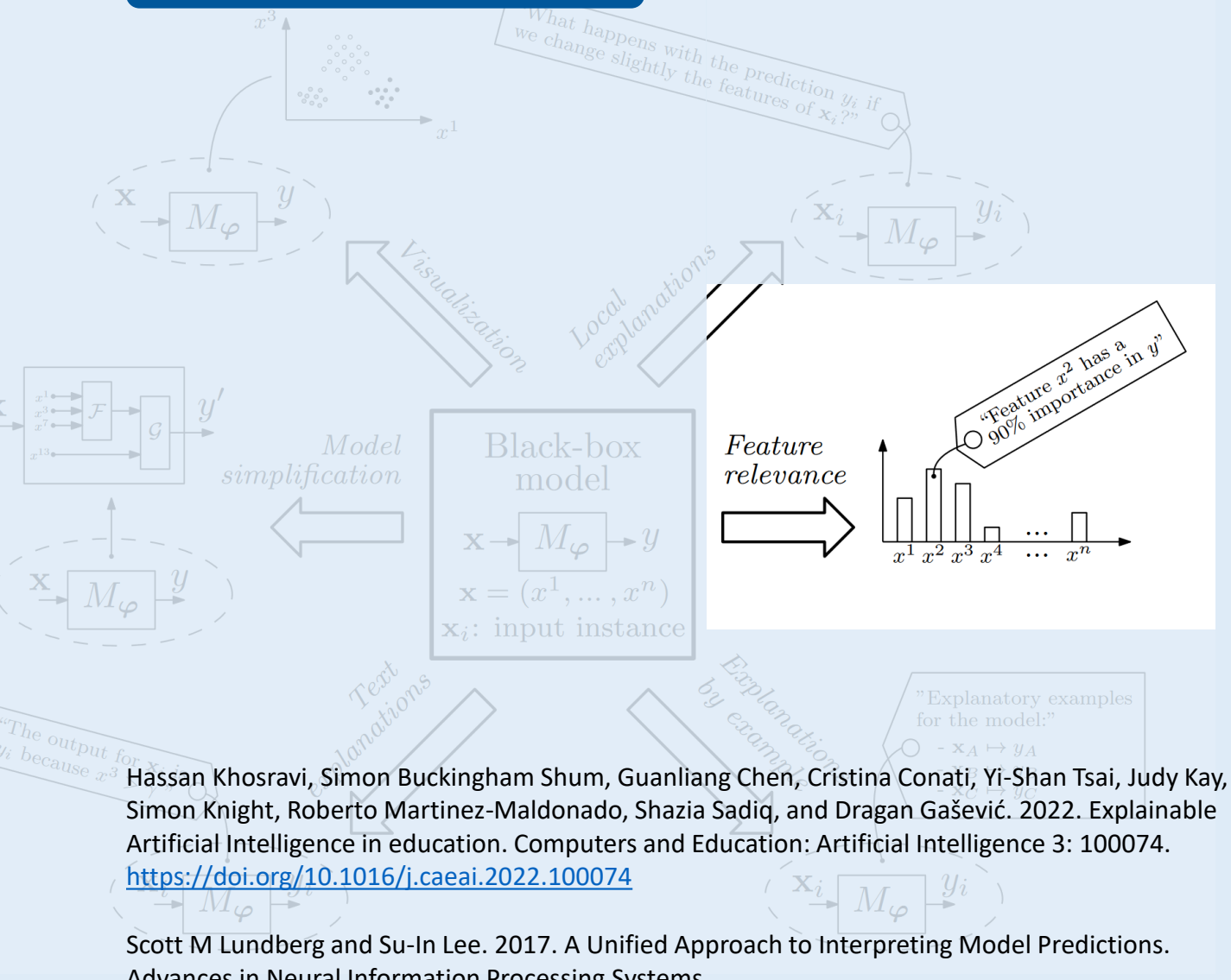
Darius Afchar, Alessandro Melchiorre, Markus Schedl, Romain Hennequin, Elena Epure, and Manuel Moussallam. 2022. Explainability in Music Recommender Systems. *AI Magazine* 43, 2: 190–208. https://doi.org/10.1002/aaai.12056

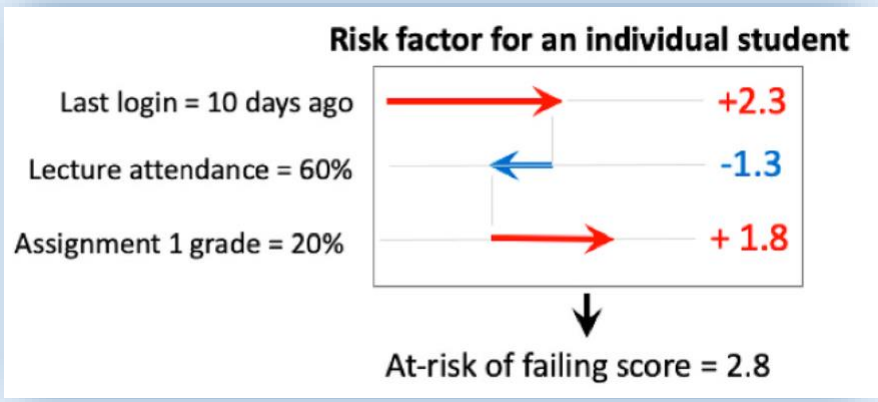
Algorithmic explainability

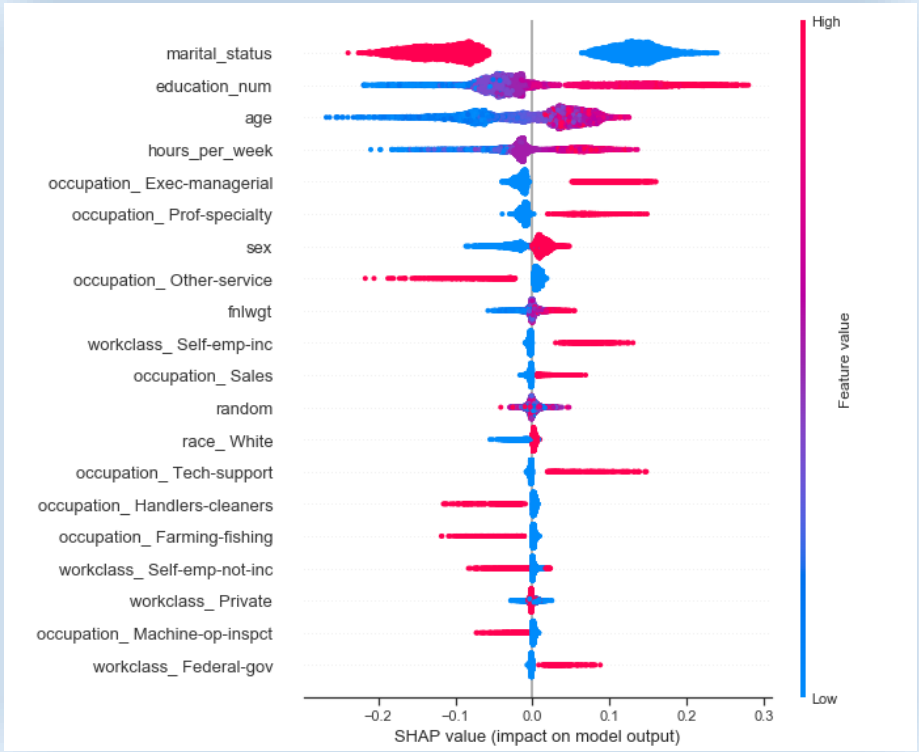


Alejandro Barredo Arrieta et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 58: 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Algorithmic explainability



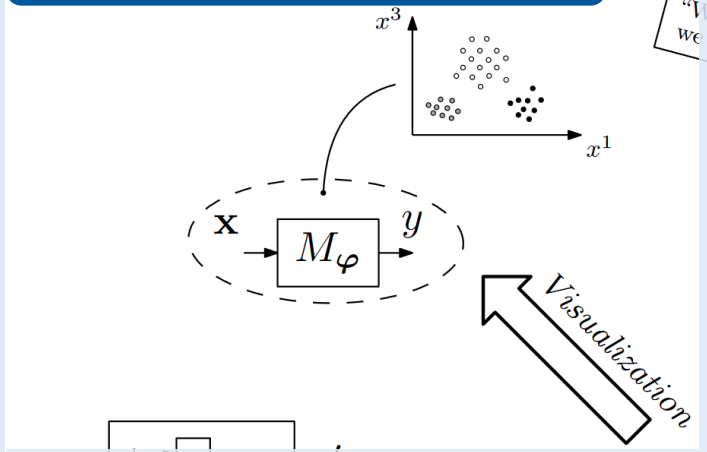


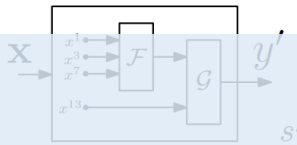


Hassan Khosravi, Simon Buckingham Shum, Guanliang Chen, Cristina Conati, Yi-Shan Tsai, Judy Kay, Simon Knight, Roberto Martinez-Maldonado, Shazia Sadiq, and Dragan Gašević. 2022. Explainable Artificial Intelligence in education. Computers and Education: Artificial Intelligence 3: 100074. https://doi.org/10.1016/j.caeai.2022.100074

Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems. https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

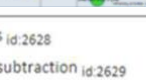
Algorithmic explainability





Model simplification

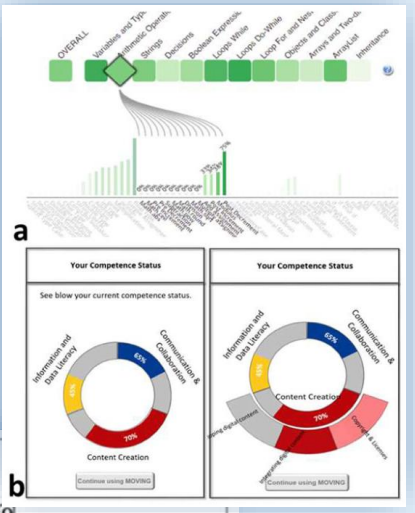
Black-box model



Local explanation

Explanation by example

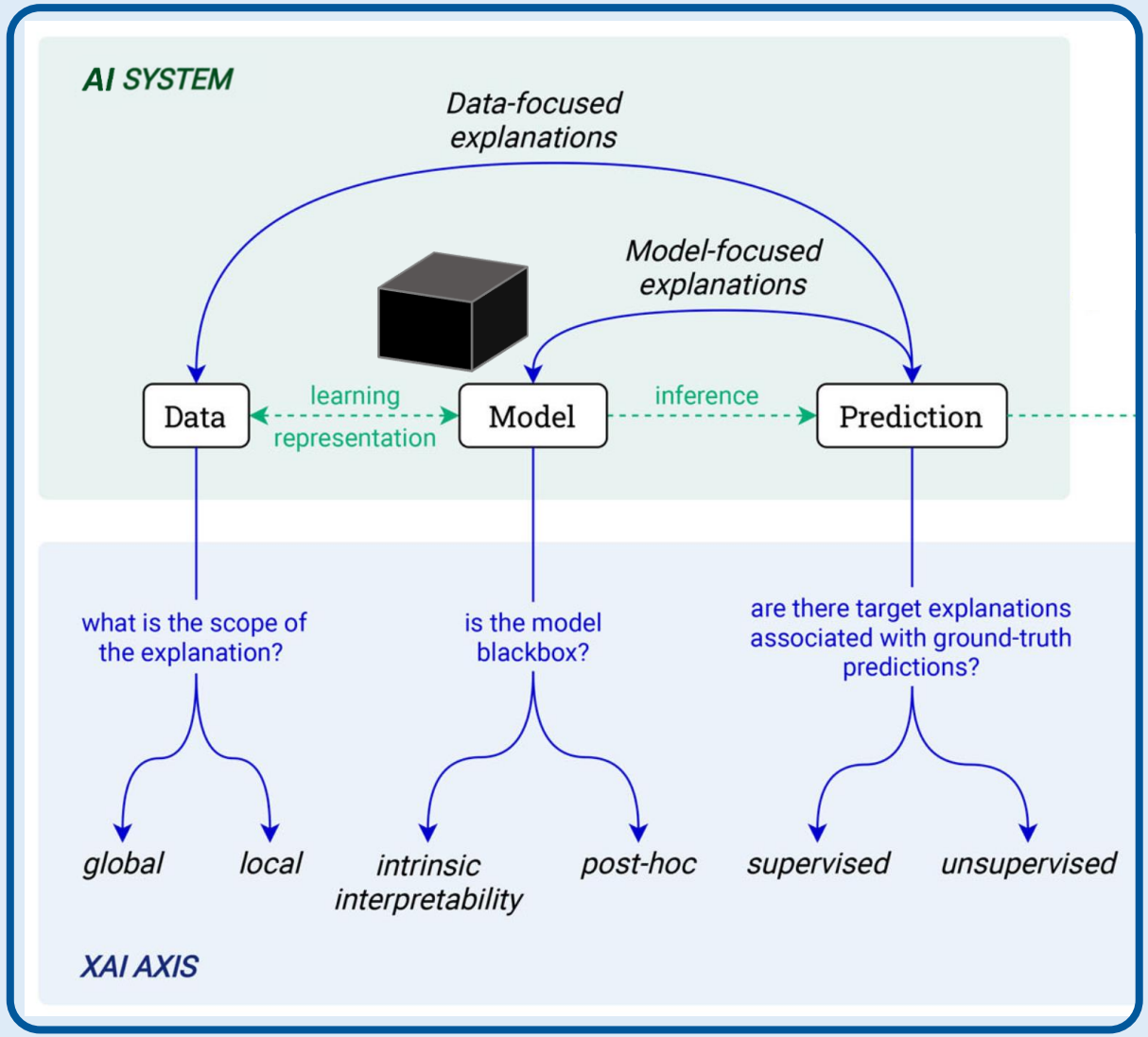
"The output for $\mathbf{x}_i$ is y_i because $x^3 > \gamma$"



Open learner models often require algorithmic skill estimation

Susan Bull. 2020. There are Open Learner Models About! *IEEE Transactions on Learning Technologies* 13, 2: 425–448. https://doi.org/10.1109/TLT.2020.2978473

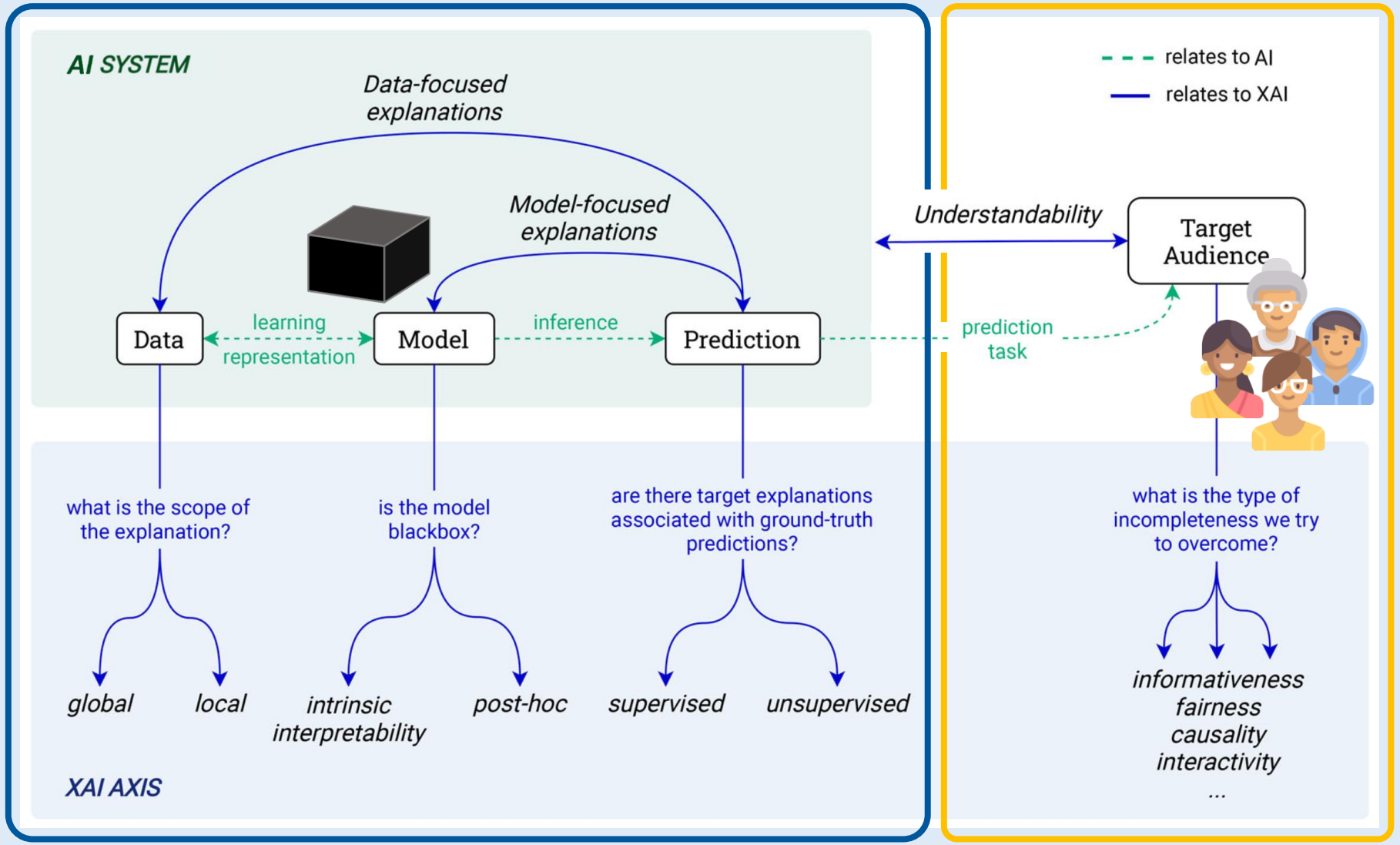
Algorithmic explainability



Darius Afchar, Alessandro Melchiorre, Markus Schedl, Romain Hennequin, Elena Epure, and Manuel Moussallam. 2022. Explainability in Music Recommender Systems. *AI Magazine* 43, 2: 190–208. https://doi.org/10.1002/aaai.12056

Algorithmic explainability

Human-centred explainability



Darius Afchar, Alessandro Melchiorre, Markus Schedl, Romain Hennequin, Elena Epure, and Manuel Moussallam. 2022. Explainability in Music Recommender Systems. *AI Magazine* 43, 2: 190–208. https://doi.org/10.1002/aaai.12056

Human-centred explainability

Different people and contexts
need different explainability solutions

Goal: Design tailored explanations and evaluate how they affect people's behaviour (e.g., trust, understanding, motivation)

The XAI-ED Framework

Who are the main *stakeholders*?
Leaners, parents, teachers, technologists, educational researchers, educational admins and policy makers

What potential *pitfalls* need to be considered?
Needless use of complex models, inaccurate or incomplete, explanations, misconceptions, promoting misbehavior.

What *approaches* are used for presenting explanations?
Globally self explaining like feature relevance or example based; locally self explaining like comparison based or counterfactual examples.

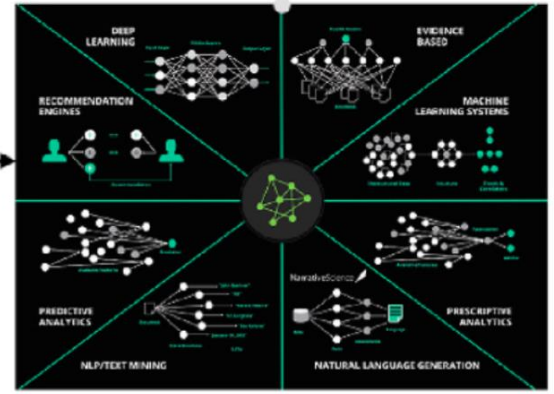
What AI *models* are commonly used?
General additive models, decision trees, rule-based models, clustering and natural language processing.

Users

What are the main *benefits*?
Agency, student-teacher interactions, AI literacy, accountability and trust.

How can educational AI tools be effective *designed*?
Using user experiences, theory driven design, centered design, participatory and co-design, HCI and interaction design.

AI interfaces

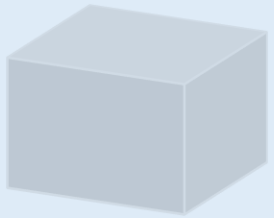


AI Models

Educational AI systems

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: 100074. https://doi.org/10.1016/j.caeai.2022.100074

Available exercises



Recommended exercise

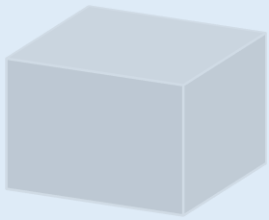
Learner pr

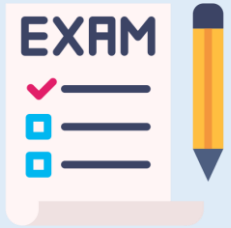
What if we want to steer the model outputs?

Predicted progress/risk

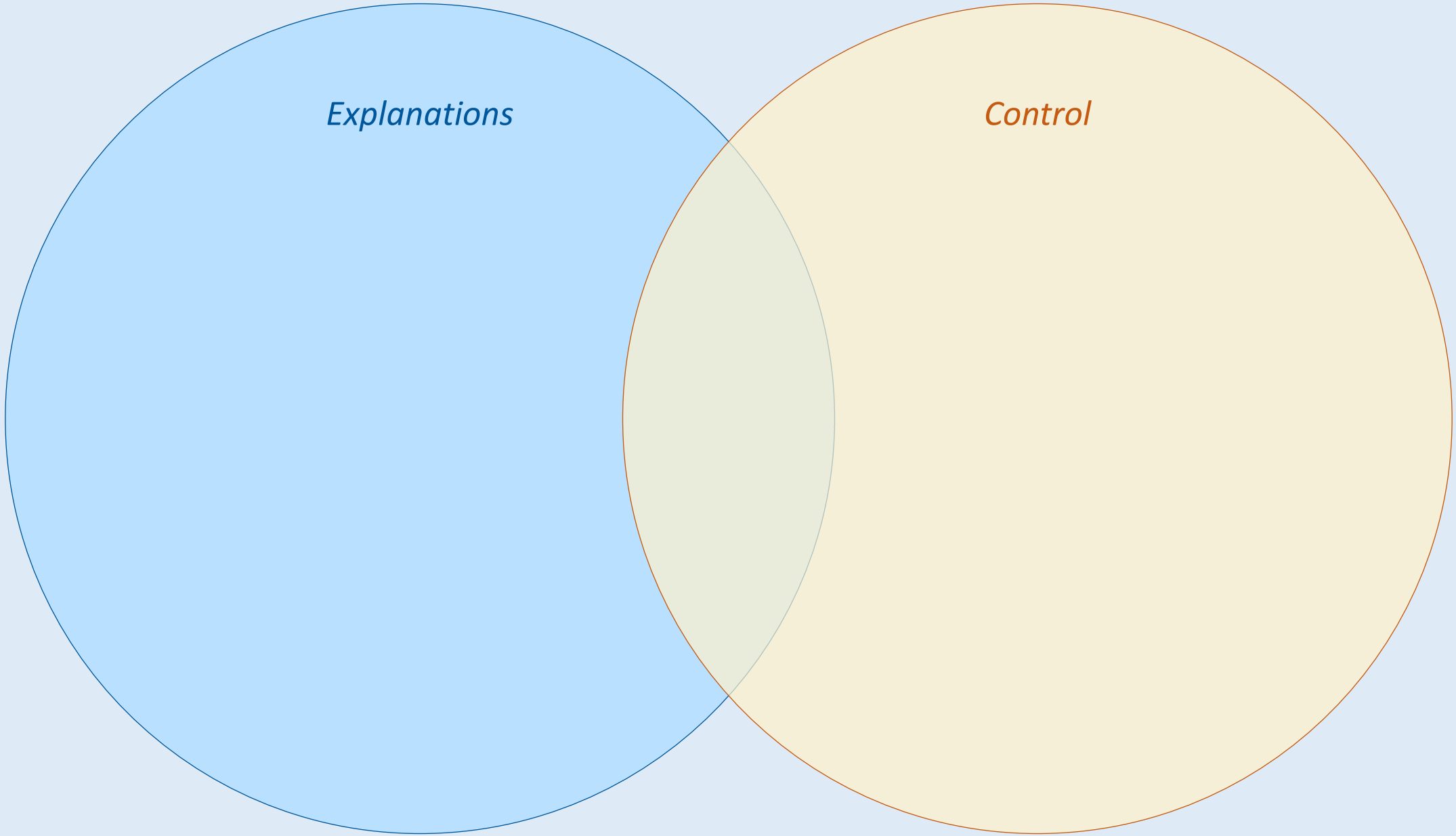
Aa

Textual prompt

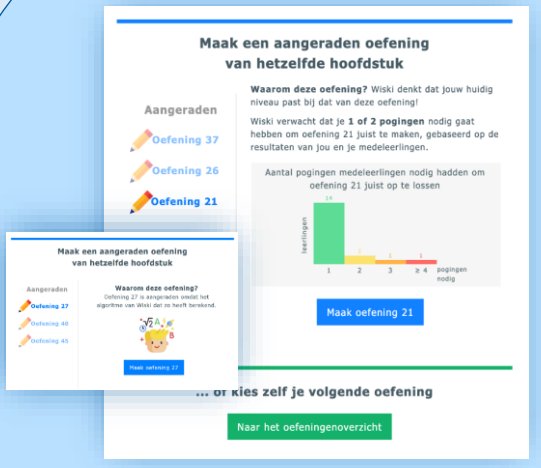


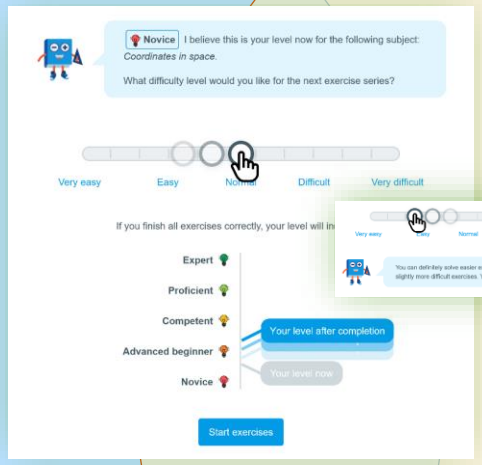


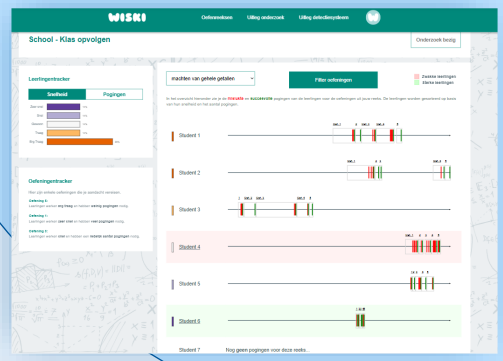
Exam question

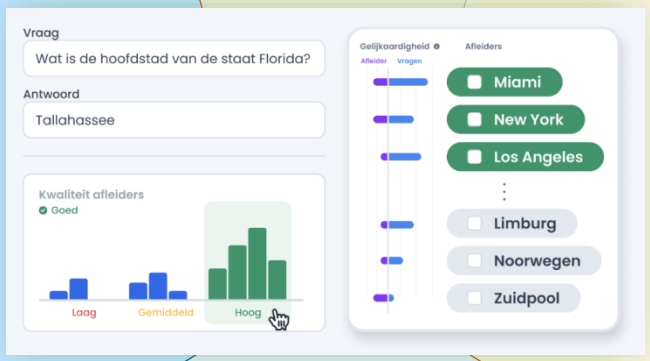

Explanations
Control

Explanations

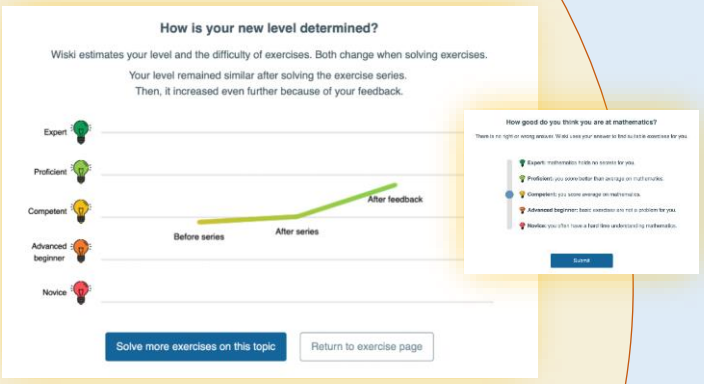


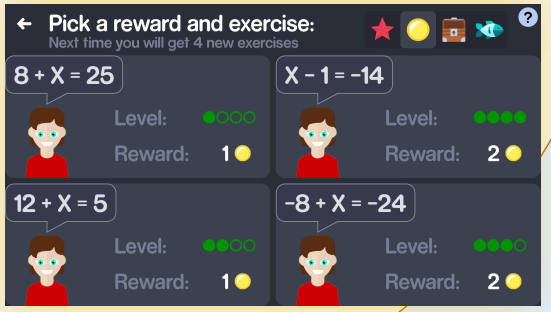






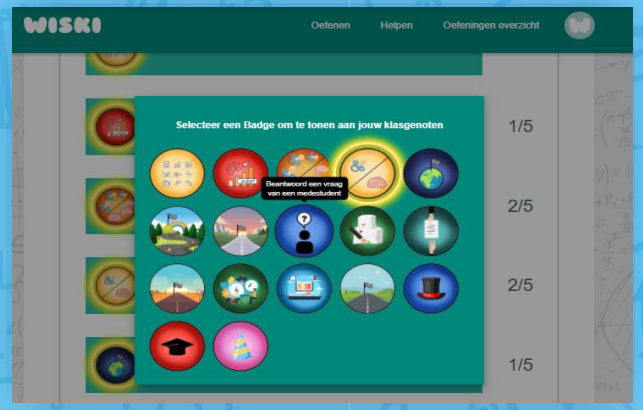
Control

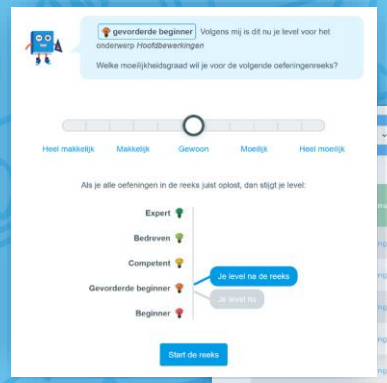


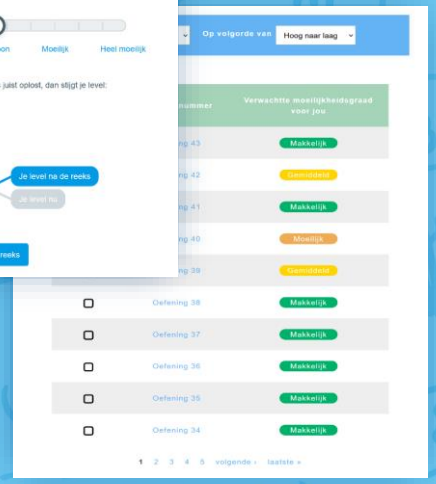


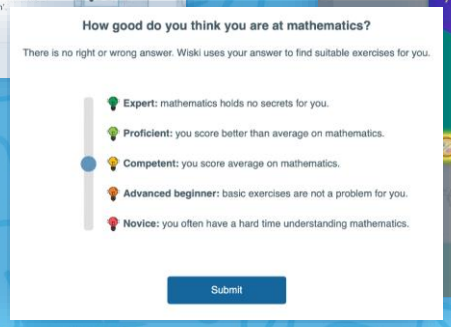
WISKI

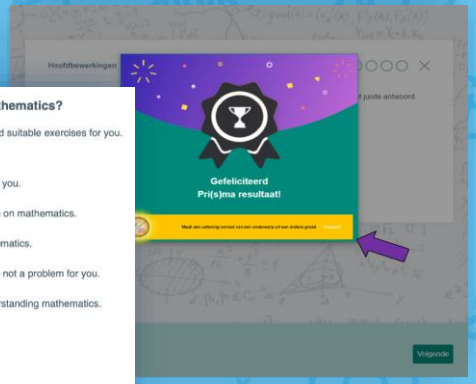
personalised maths practice with AI

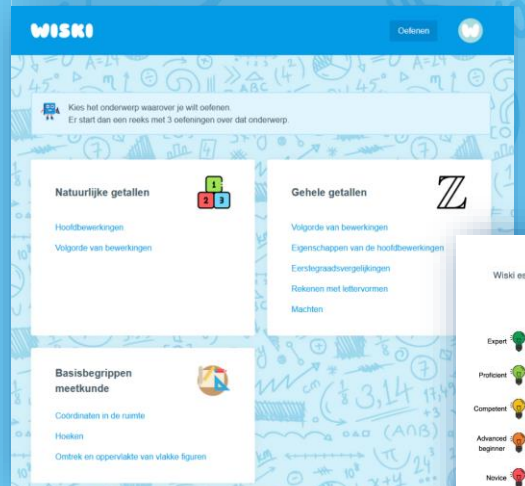


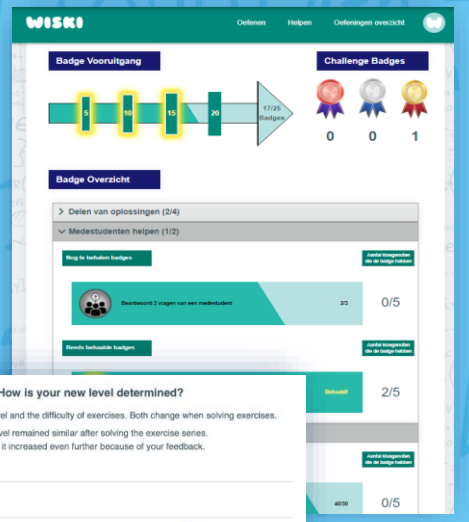












Volgorde van bewerkingen

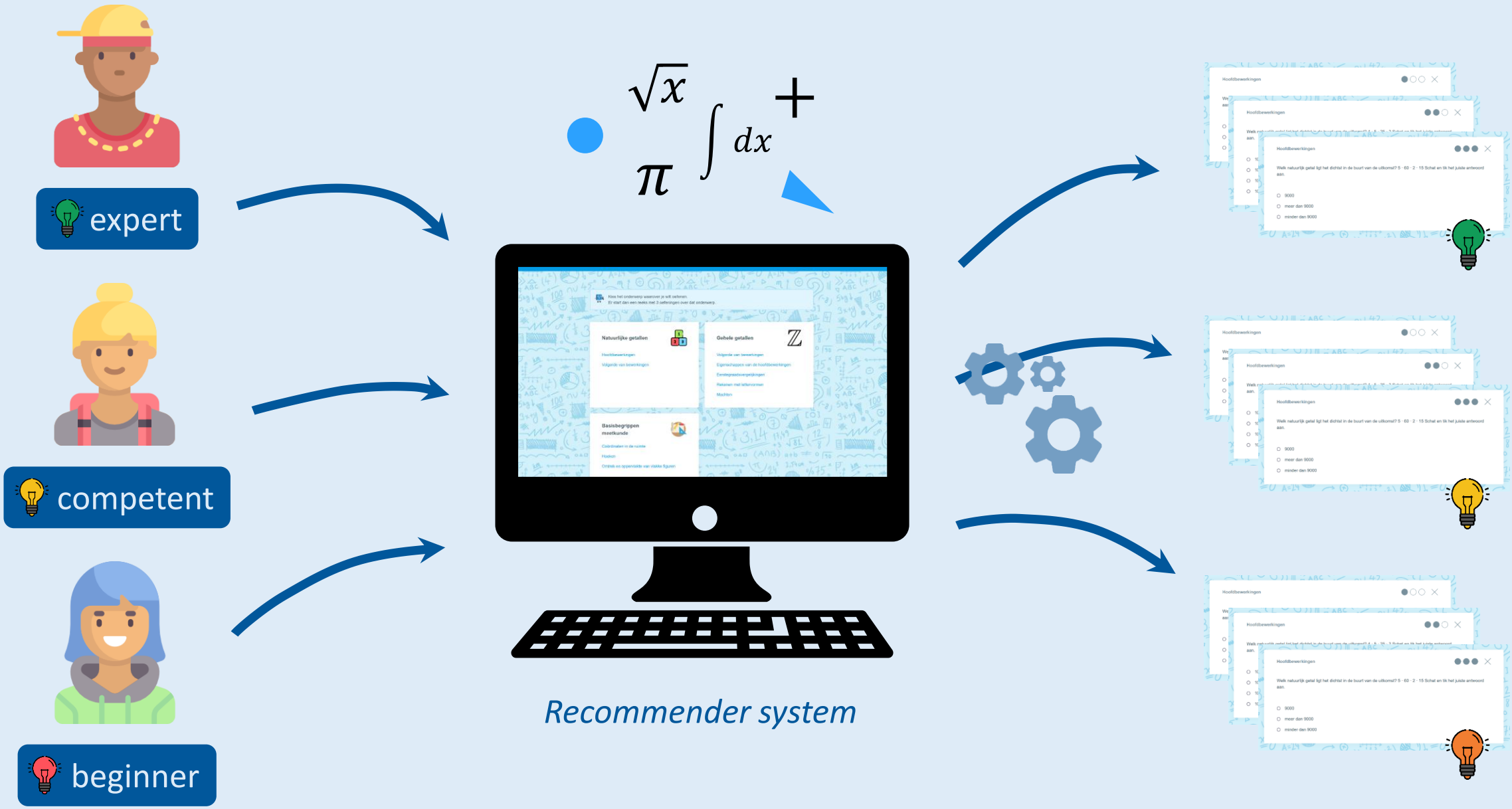
$6 \cdot (-12) + 36 \cdot 2 = \ldots$ Bereken en tik de juiste oplossing aan.

- (●) 0
- ( ) -144
- ( ) 72
- ( ) 144

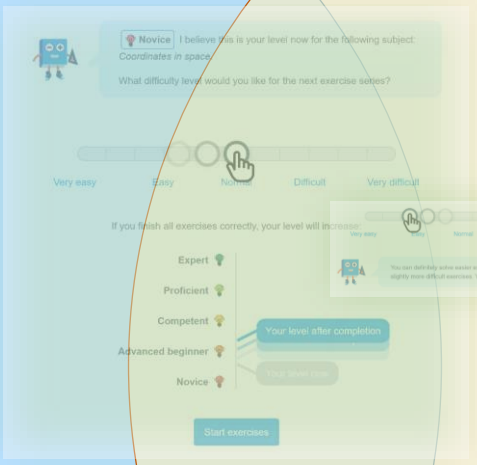
Correct

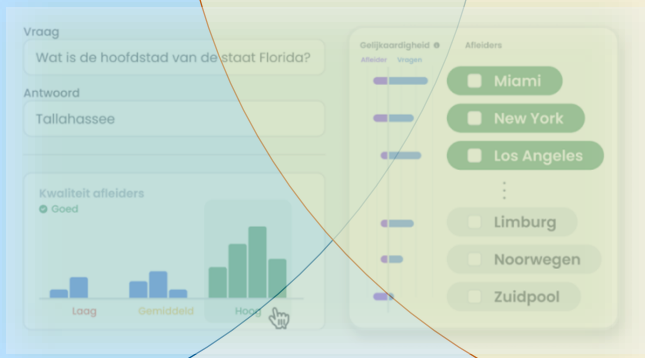
Top-π!

Overslaan Volgende

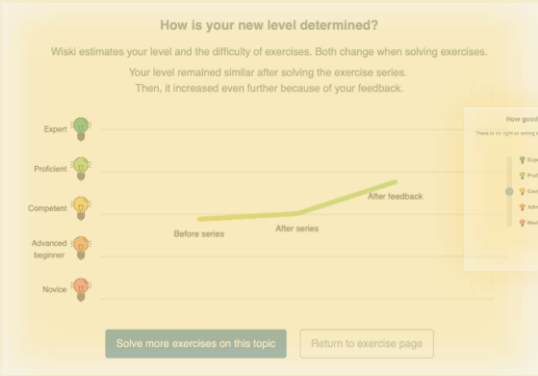

expert
competent
beginner
Recommender system

Explanations



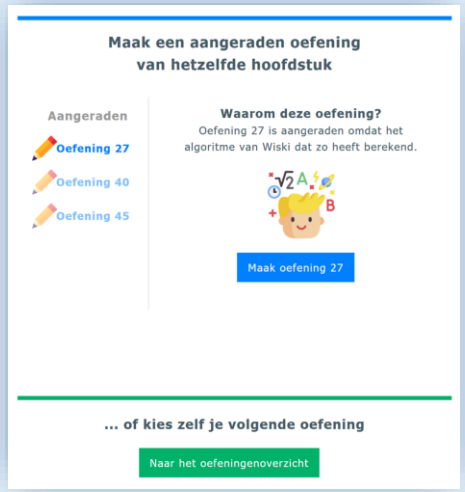


Control



Textual explanation

Visual explanation



Placebo explanation

No explanation

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), 93–105. https://doi.org/10.1145/3490099.3511140

Questionnaire on trust

Competence

Q1 Wiski is like an expert (for example, a teacher) for recommending exercises.
Q2 Wiski has the expertise (knowledge) to estimate my math level.
Q3 Wiski can estimate my math level.
Q4 Wiski understands the difficulty level of math exercises well.
Q5 Wiski takes my math level into account when recommending exercises.

Benevolence

Q6 Wiski prioritizes that I improve in math.
Q7 Wiski recommends exercises so that I improve in math.
Q8 Wiski wants to estimate my math level well.

Integrity

Q9 Wiski recommends exercises as correctly as possible.
Q10 Wiski is honest.
Q11 Wiski makes integrous recommendations.

Trust (one-dimensional)

Q12 I trust Wiski to recommend me math exercises.

Intention to return

Q13 If I want to solve math exercises again, I will choose Wiski.
Q14 If I want to be recommended math exercises again, I will choose Wiski.

Perceived transparency

Q15 I find that Wiski gives enough explanation as to why an exercise is recommended.

Trusting beliefs

Multidimensional trust

Effects on trust

Placebo explanations

1D trust

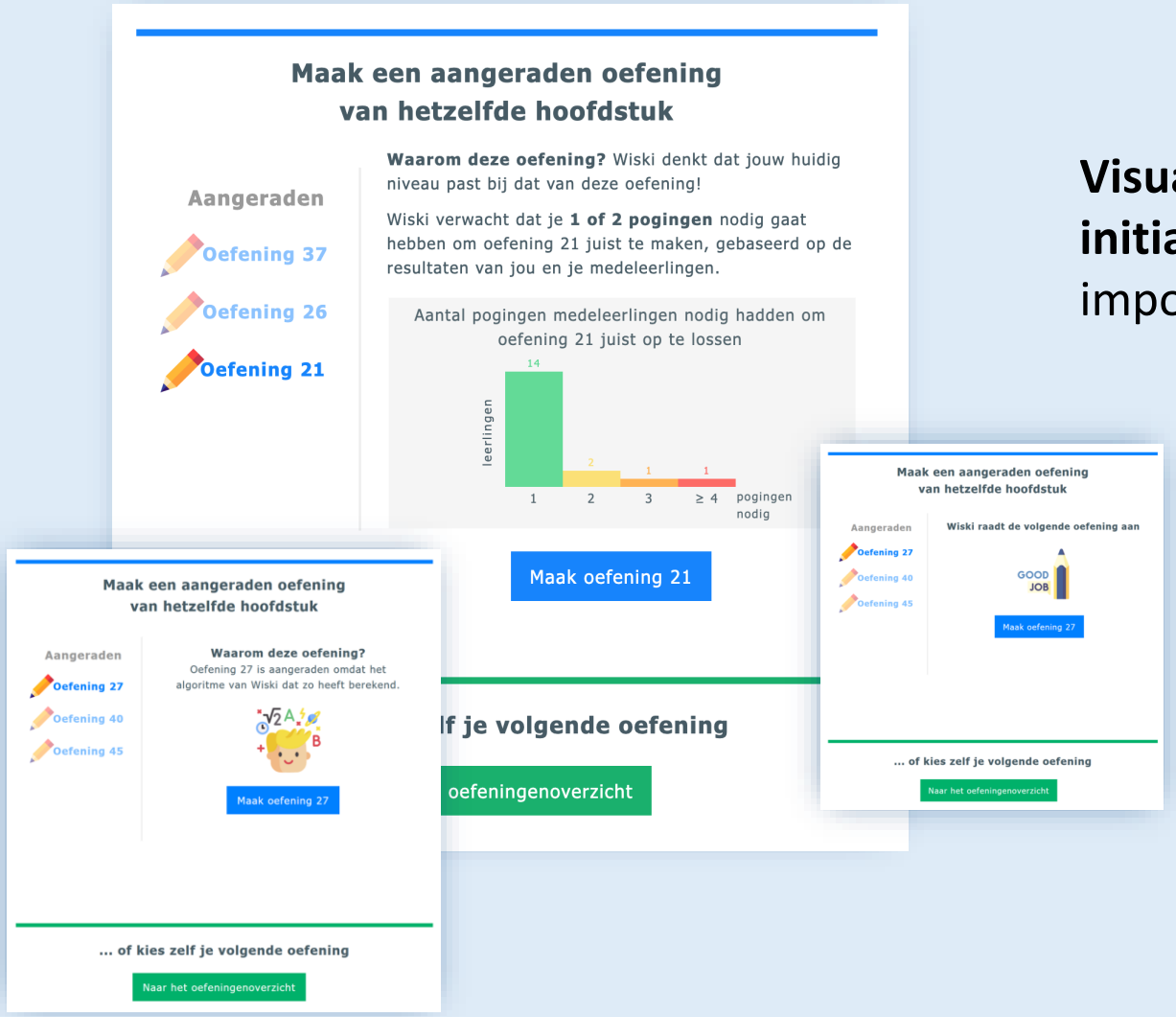
MD trust

Real explanations

1D trust

MD trust

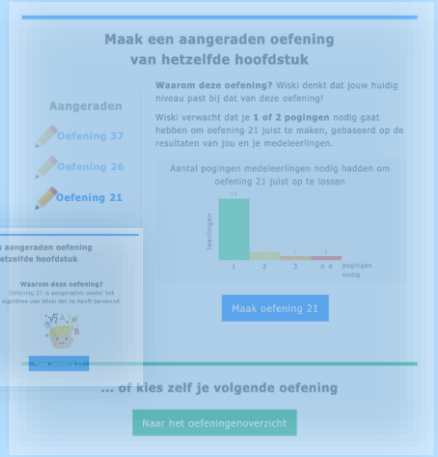
Jeroen Ooge, Shotallo Kato, and Katrien Verbert. 2022. Explaining Recommendations in E-Learning: Effects on Adolescents' Trust. In *27th International Conference on Intelligent User Interfaces* (IUI '22), 93–105. https://doi.org/10.1145/3490099.3511140

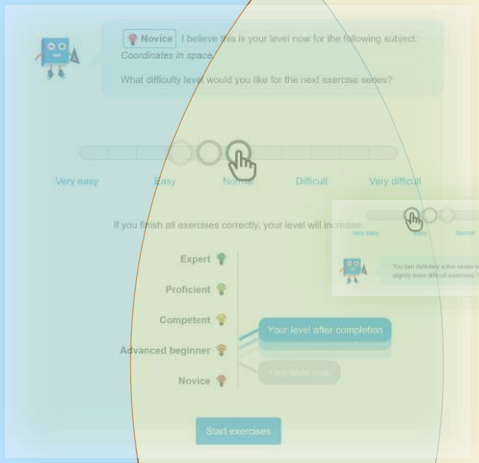


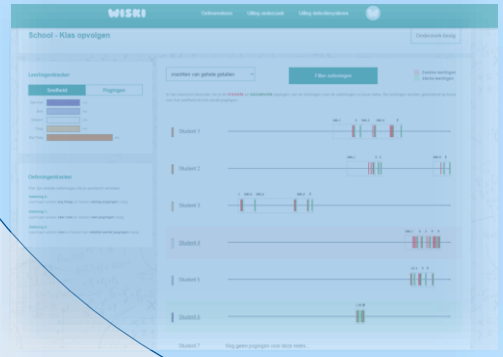
Visual explanations can increase initial trust but may not be the most important factor for building it

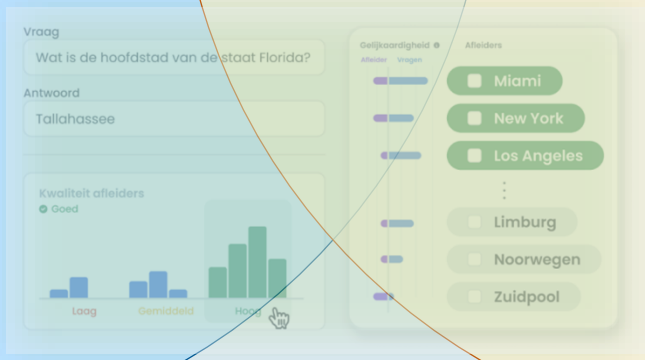
Jeroen Ooge, Shotallo Kato, and Katrien Verbert. 2022. Explaining Recommendations in E-Learning: Effects on Adolescents' Trust.
In *27th International Conference on Intelligent User Interfaces* (IUI '22), 93–105. https://doi.org/10.1145/3490099.3511140

Explanations

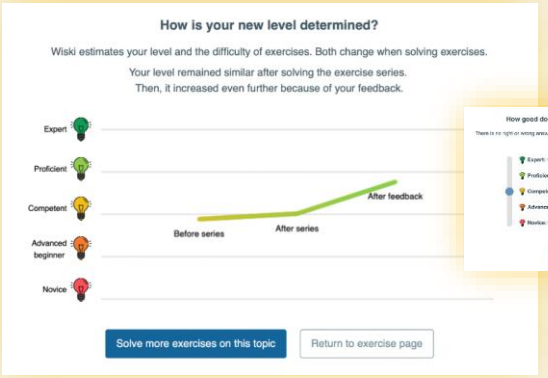


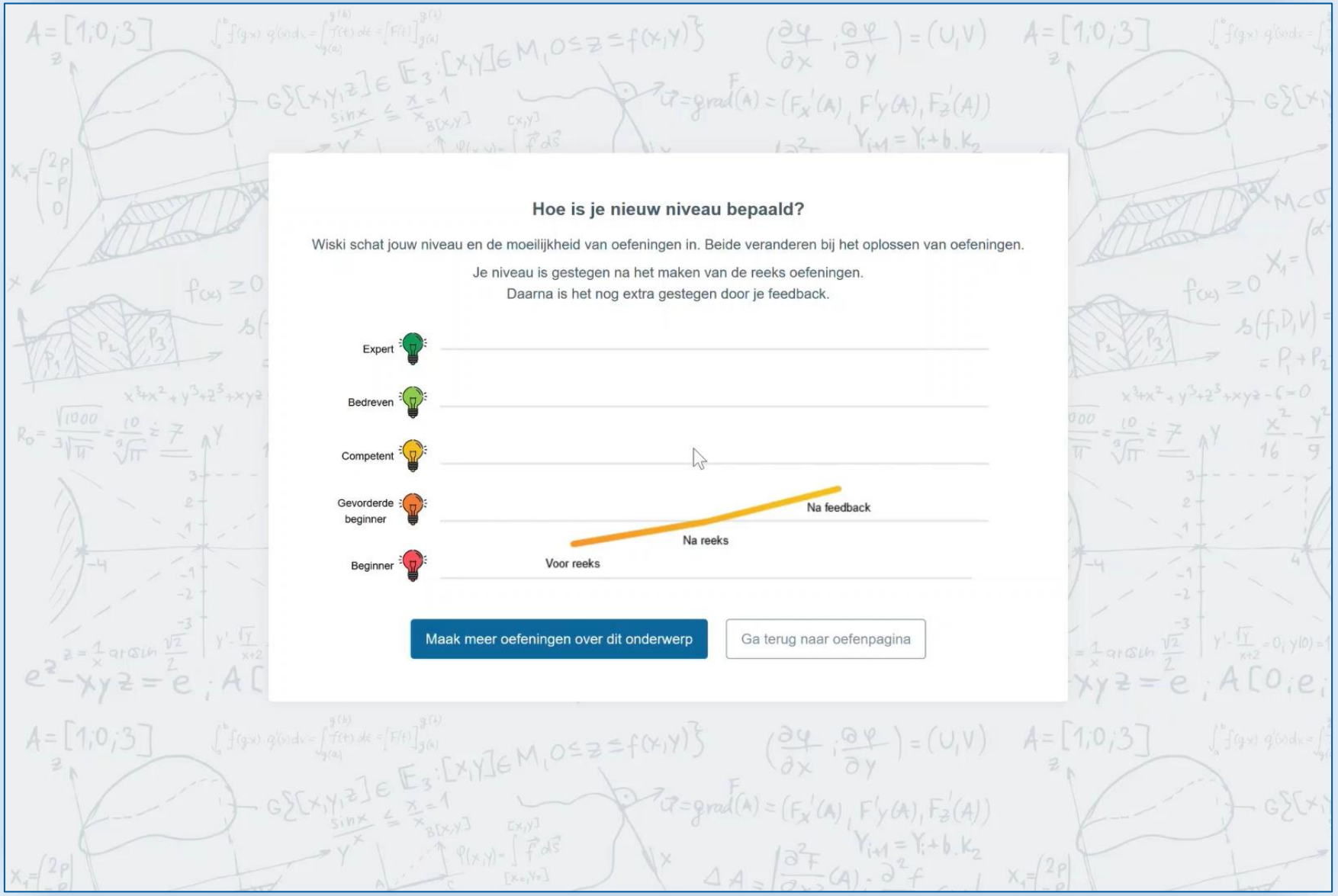




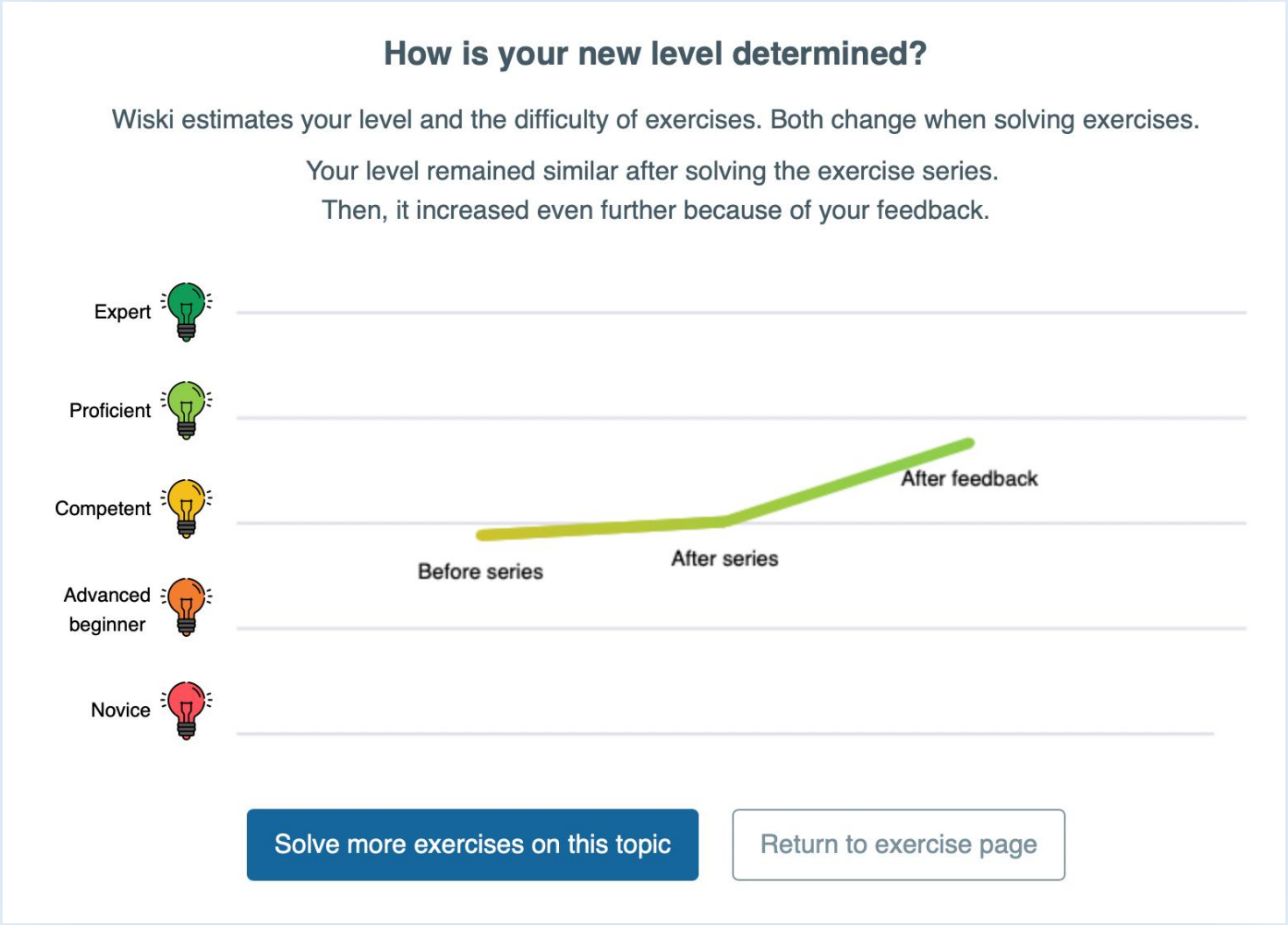


Control





Jeroen Ooge, Leen Dereu, and Katrien Verbert. 2023. Steering Recommendations and Visualising Its Impact: Effects on Adolescents' Trust in E-Learning Platforms. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (IUI '23), 156–170. https://doi.org/10.1145/3581641.3584046



Jeroen Ooge, Leen Dereu, and Katrien Verbert. 2023. Steering Recommendations and Visualising Its Impact: Effects on Adolescents' Trust in E-Learning Platforms. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (IUI '23), 156–170. https://doi.org/10.1145/3581641.3584046

Final questionnaire: **control**

Control	
Q18	I feel in control of telling Wiski what I want.
Q19	I don't feel in control of telling Wiski what I want.
Q20	I don't feel in control of specifying and changing my
Q21	Wiski seems to control my decision process rather t
Preference elicitation	
Q22	Wiski provides an adequate way for me to express m
Q23	I found it easy to tell Wiski about my preferences.
Q24	It is easy to learn to tell Wiski what I like.
Q25	It required too much effort to tell Wiski what I like.
Preference revision	
Q26	Wiski provides an adequate way for me to revise my
Q27	I found it easy to make Wiski recommend different t
Q28	It is easy to train Wiski to update my preferences.
Q29	I found it easy to alter the recommended exercises d
Q30	It is easy for me to inform Wiski if I dislike/like reco
Q31	It is easy for me to get a new set of recommended ex

7-point Likert-type questions + open comments

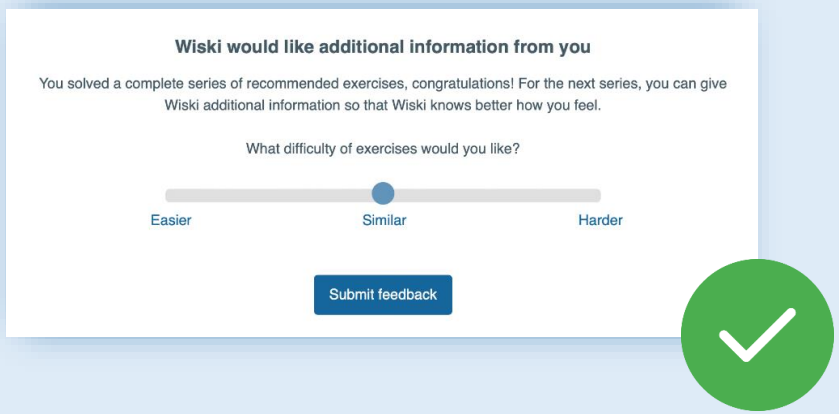
Final questionnaire: **trust**

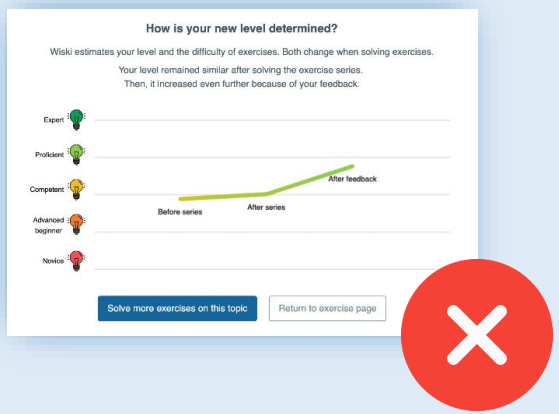
Competence	
Q1	Wiski is like an expert (for example, a teacher) for r
Q2	Wiski has the expertise (knowledge) to estimate my
Q3	Wiski can estimate my maths level.
Q4	Wiski understands the difficulty level of maths exer
Q5	Wiski takes my maths level into account when reco
Benevolence	
Q6	Wiski prioritises that I improve in maths.
Q7	Wiski recommends exercises so that I improve in m
Q8	Wiski wants to estimate my maths level well.
Integrity	
Q9	Wiski recommends exercises as correctly as possibl
Q10	Wiski is honest.
Q11	Wiski makes integrous recommendations.
Trust (one-dimensional)	
Q12	I trust Wiski to recommend me maths exercises.
Intention to return	
Q13	If I want to solve maths exercises again, I will choos
Q14	If I want to be recommended maths exercises again
Transparency	
Q15	I understood why the exercises were recommended
Q16	Wiski helps me understand why the exercises were
Q17	Wiski explains why the exercises are recommended

Trusting beliefs

Multidimensional trust

Jeroen Ooge, Leen Dereu, and Katrien Verbert. 2023. Steering Recommendations and Visualising Its Impact: Effects on Adolescents' Trust in E-Learning Platforms. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (IUI '23), 156–170. https://doi.org/10.1145/3581641.3584046



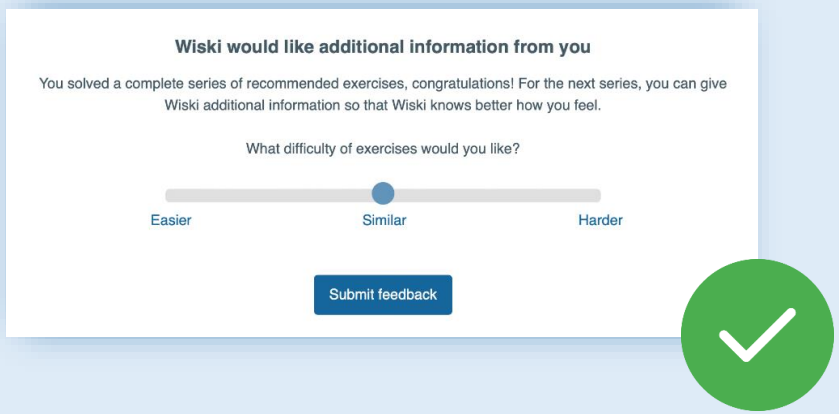


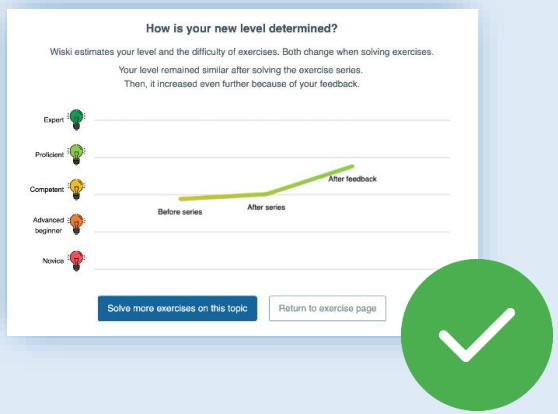
Trust in the platform

Reflection on own mastery

Reflection on recommendations

Jeroen Ooge, Leen Dereu, and Katrien Verbert. 2023. Steering Recommendations and Visualising Its Impact: Effects on Adolescents' Trust in E-Learning Platforms. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (IUI '23), 156–170. https://doi.org/10.1145/3581641.3584046



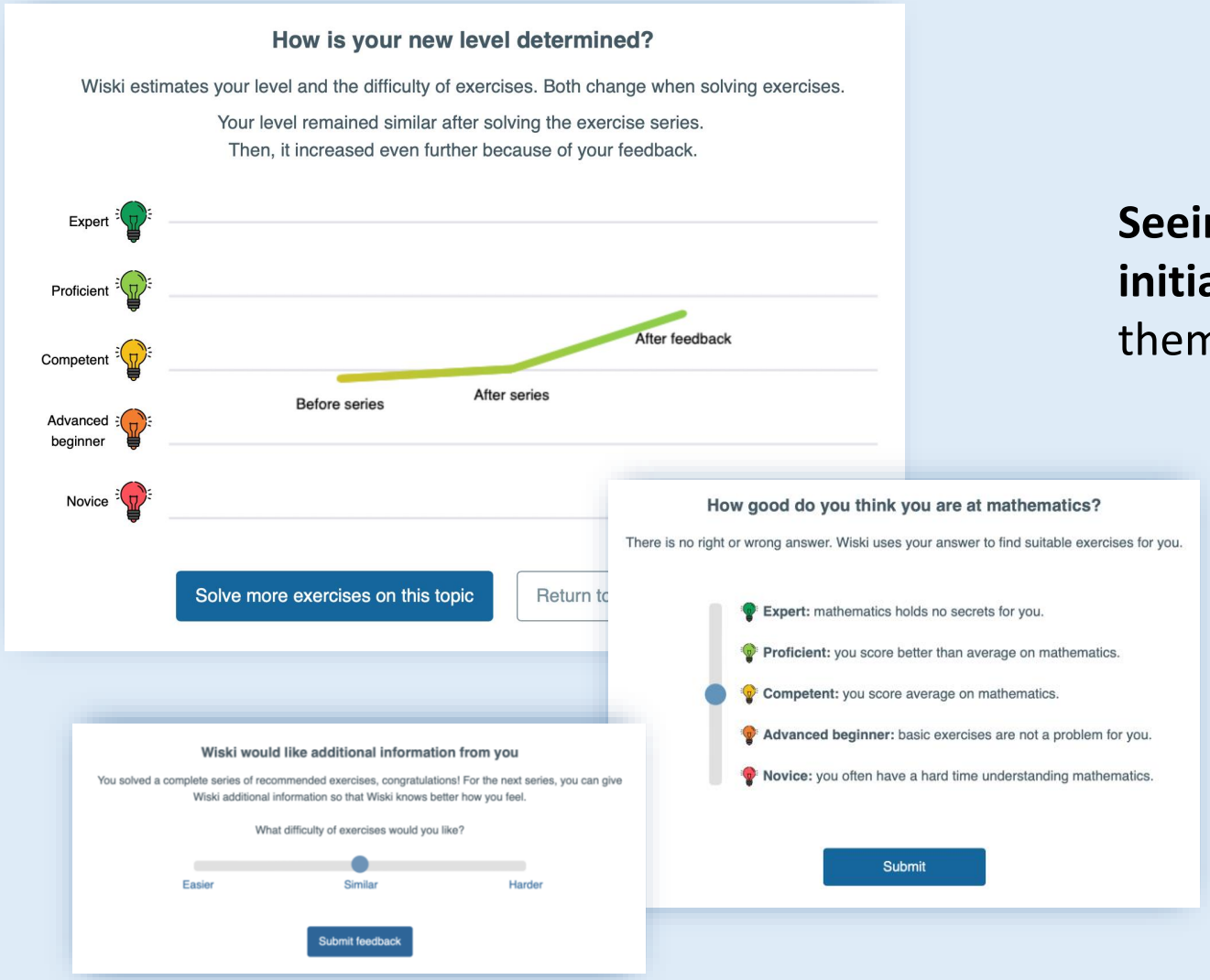


Trust in the platform

Perceived transparency

Reflection

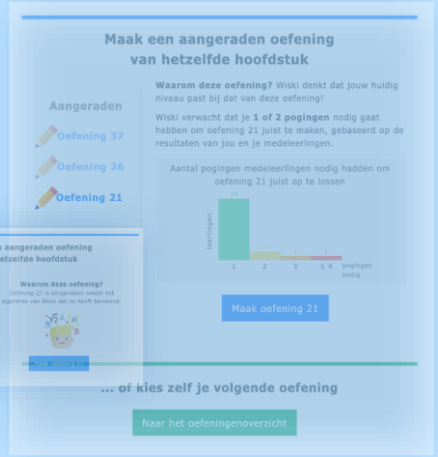
Jeroen Ooge, Leen Dereu, and Katrien Verbert. 2023. Steering Recommendations and Visualising Its Impact: Effects on Adolescents' Trust in E-Learning Platforms. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (IUI '23), 156–170. https://doi.org/10.1145/3581641.3584046

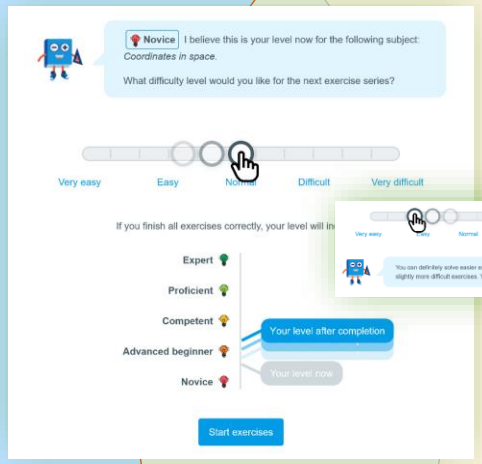


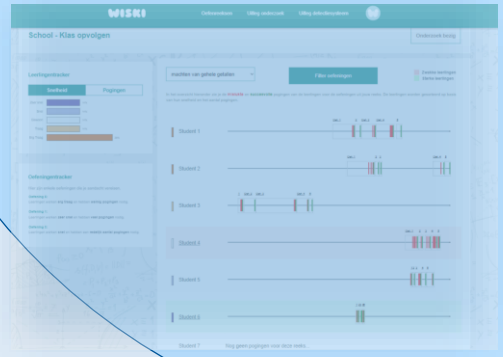
Seeing the impact of control can increase initial trust, but control mechanisms by themselves do not necessarily

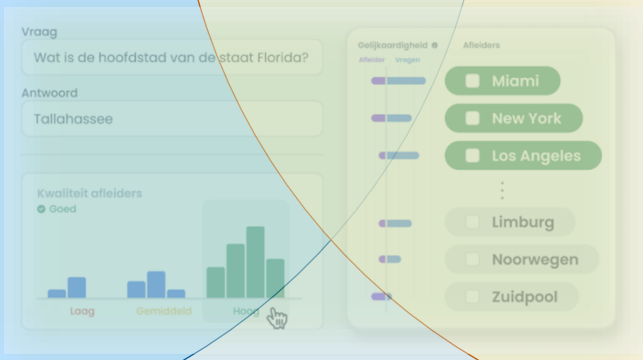
Jeroen Ooge, Leen Dereu, and Katrien Verbert. 2023. Steering Recommendations and Visualising Its Impact: Effects on Adolescents' Trust in E-Learning Platforms. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (IUI '23), 156–170. https://doi.org/10.1145/3581641.3584046

Explanations

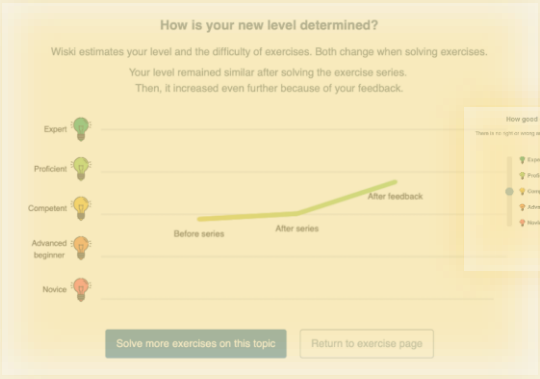


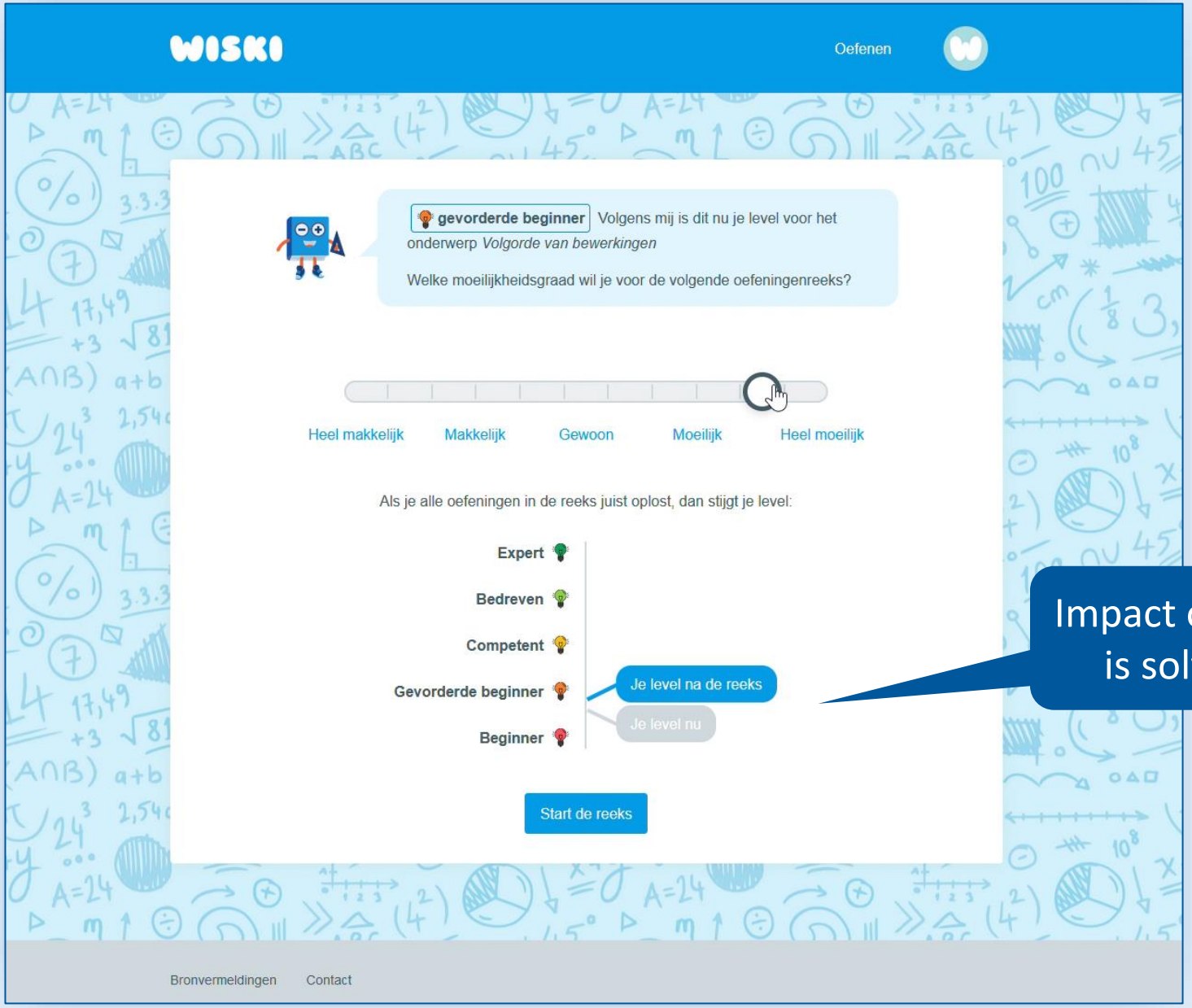




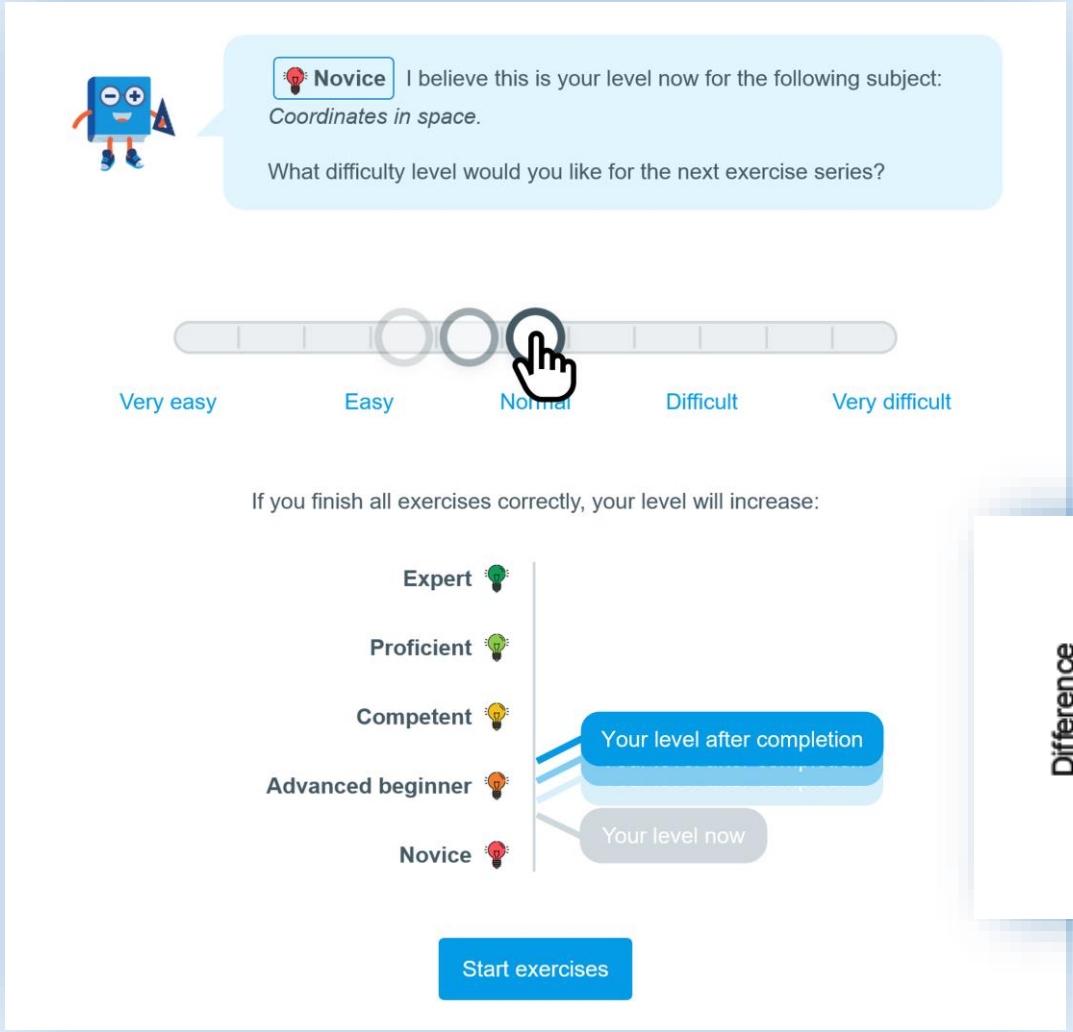


Control

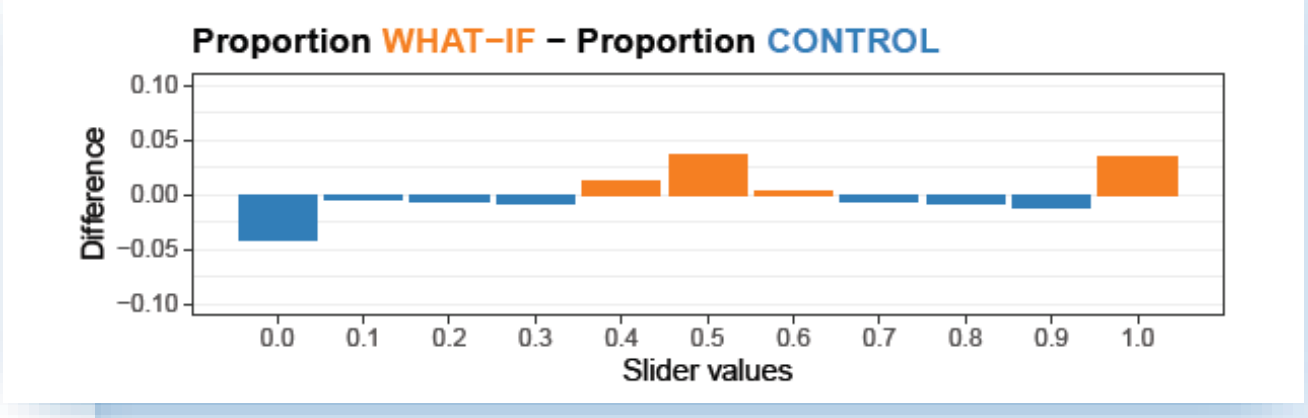




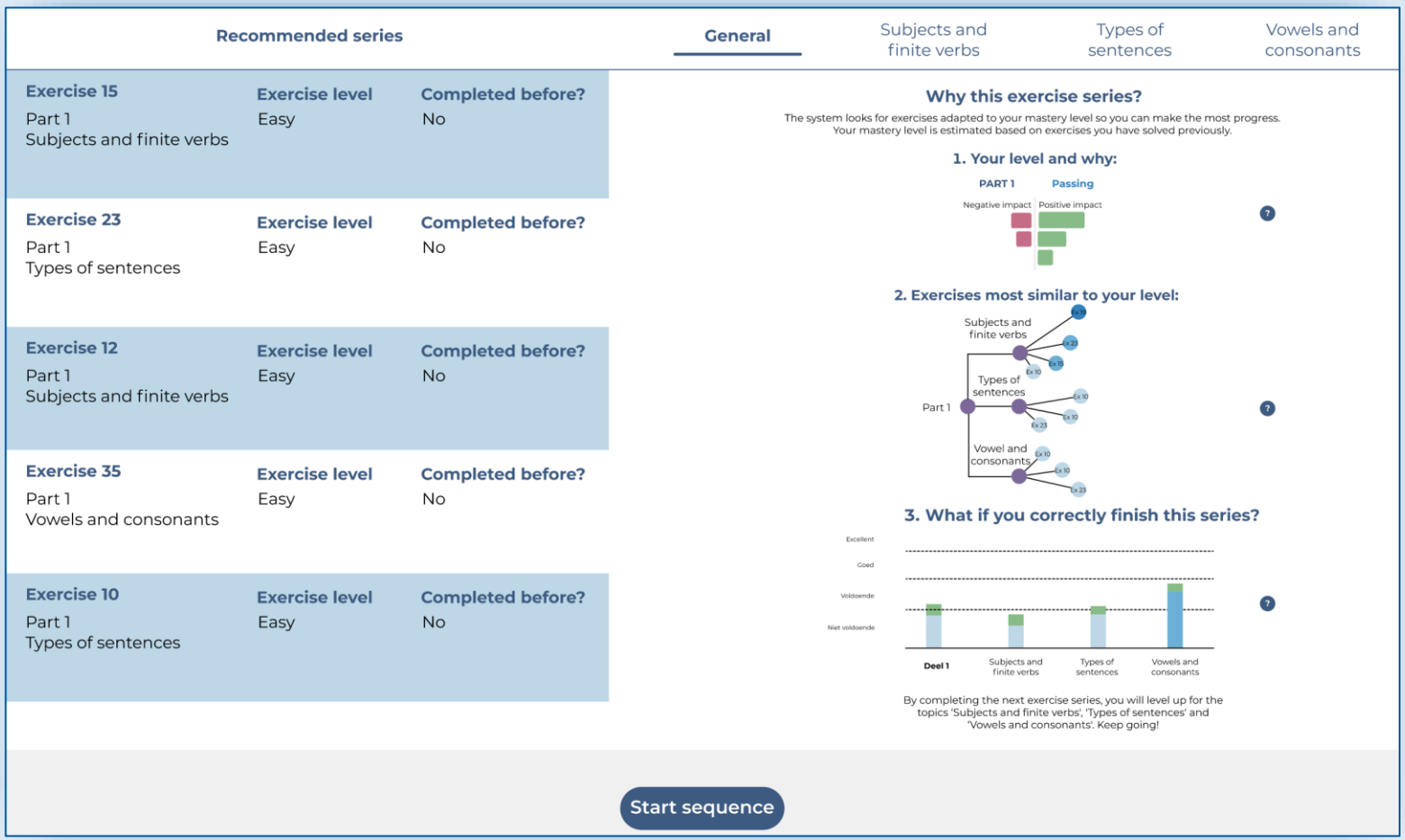
Impact on level if series is solved correctly



***What-if* explanations can encourage teenagers to try harder exercises**, but do not necessarily affect initial trust, metacognition, motivation, or performance

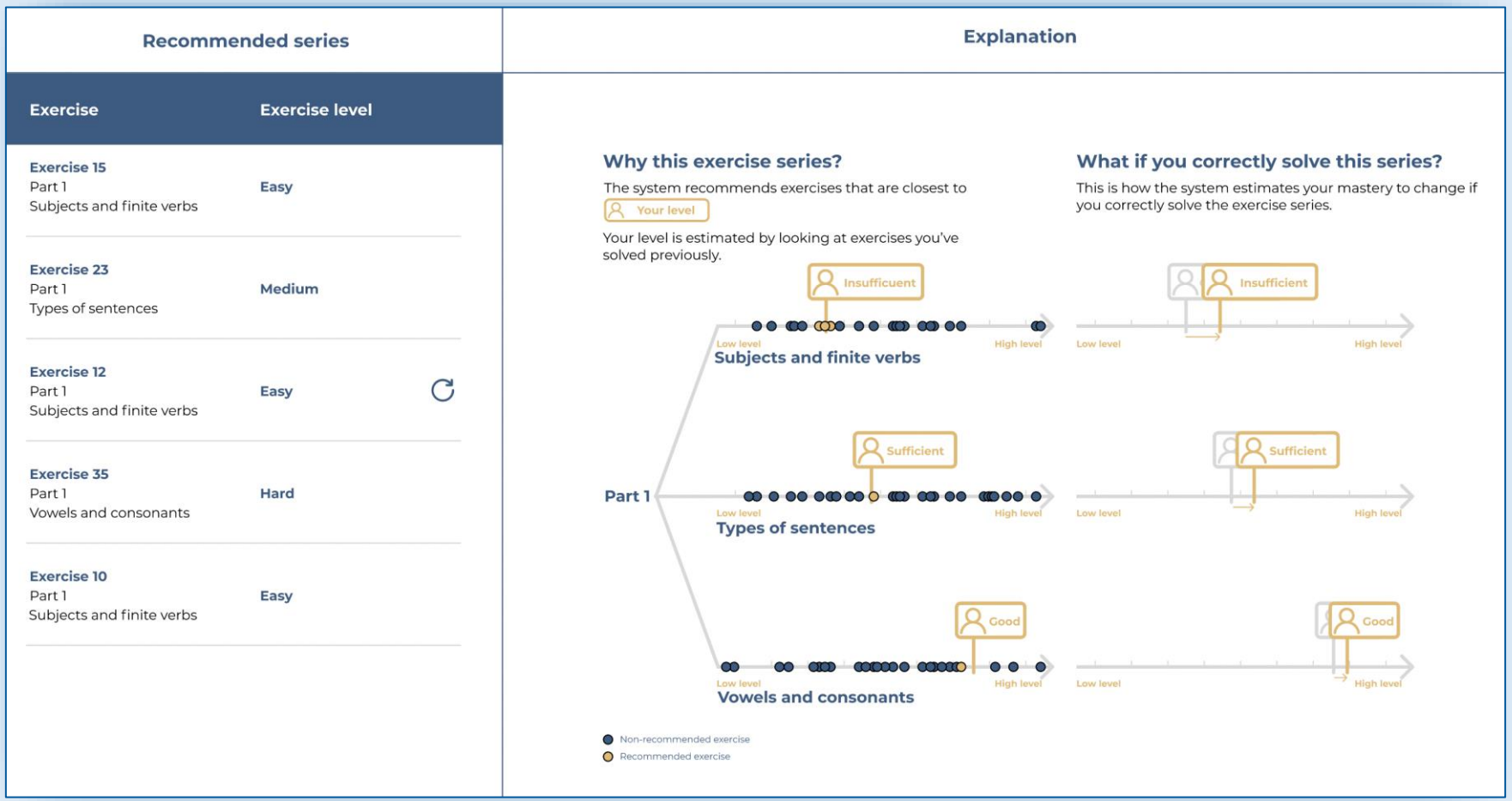


Jeroen Ooge, Maxwell Szymanski, Arno Vanneste, Robin De Croon, and Katrien Verbert. Steer, See Impact, Solve: How Learner Control and Visual Explanations Impact Learning, Motivation, and Trust. To be submitted.



Group discussions with teachers and people developing educational technologies

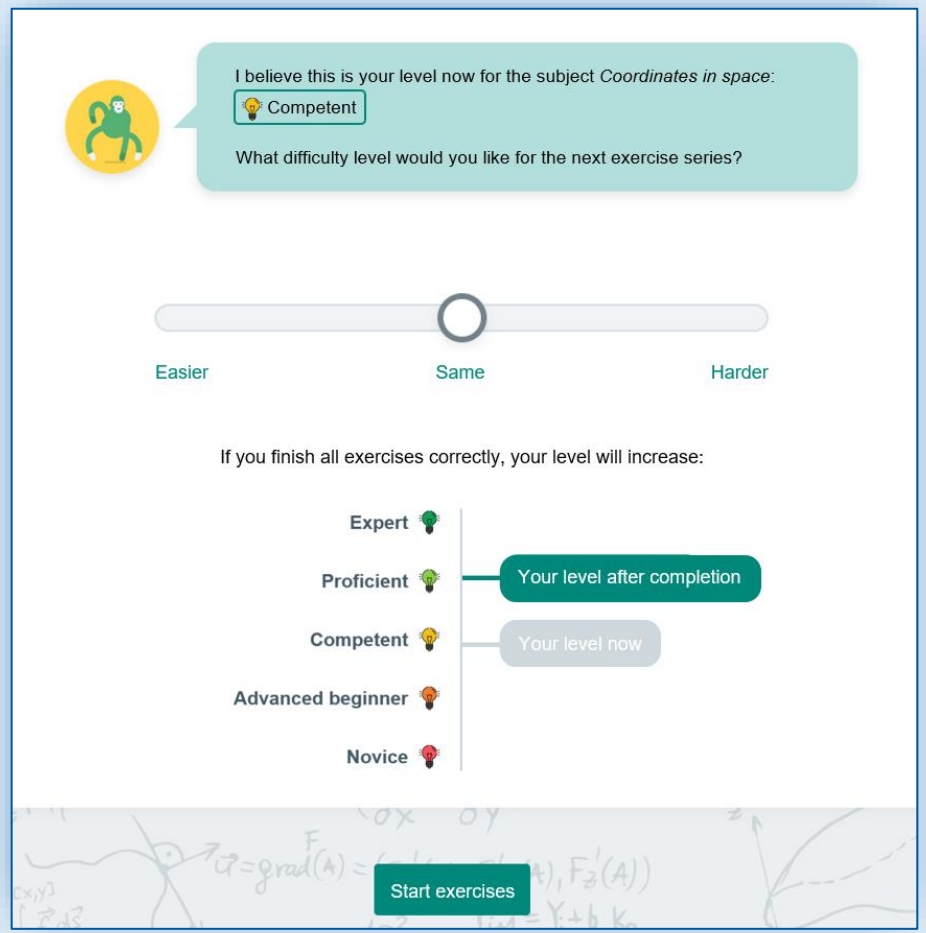
Jeroen Ooge, Maxwell Szymanski, Arno Vanneste, Robin De Croon, and Katrien Verbert. Steer, See Impact, Solve: How Learner Control and Visual Explanations Impact Learning, Motivation, and Trust. To be submitted.

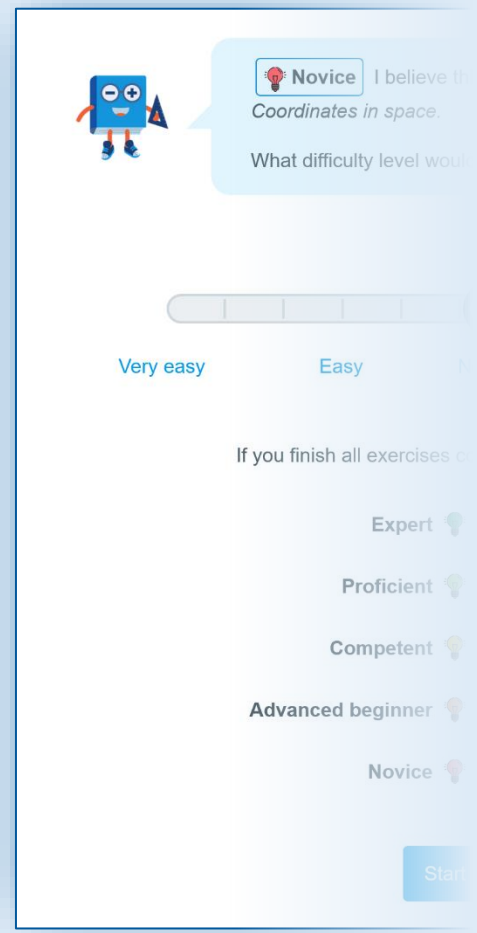


Focus groups with teachers and educational experts

Think-aloud sessions with adolescents

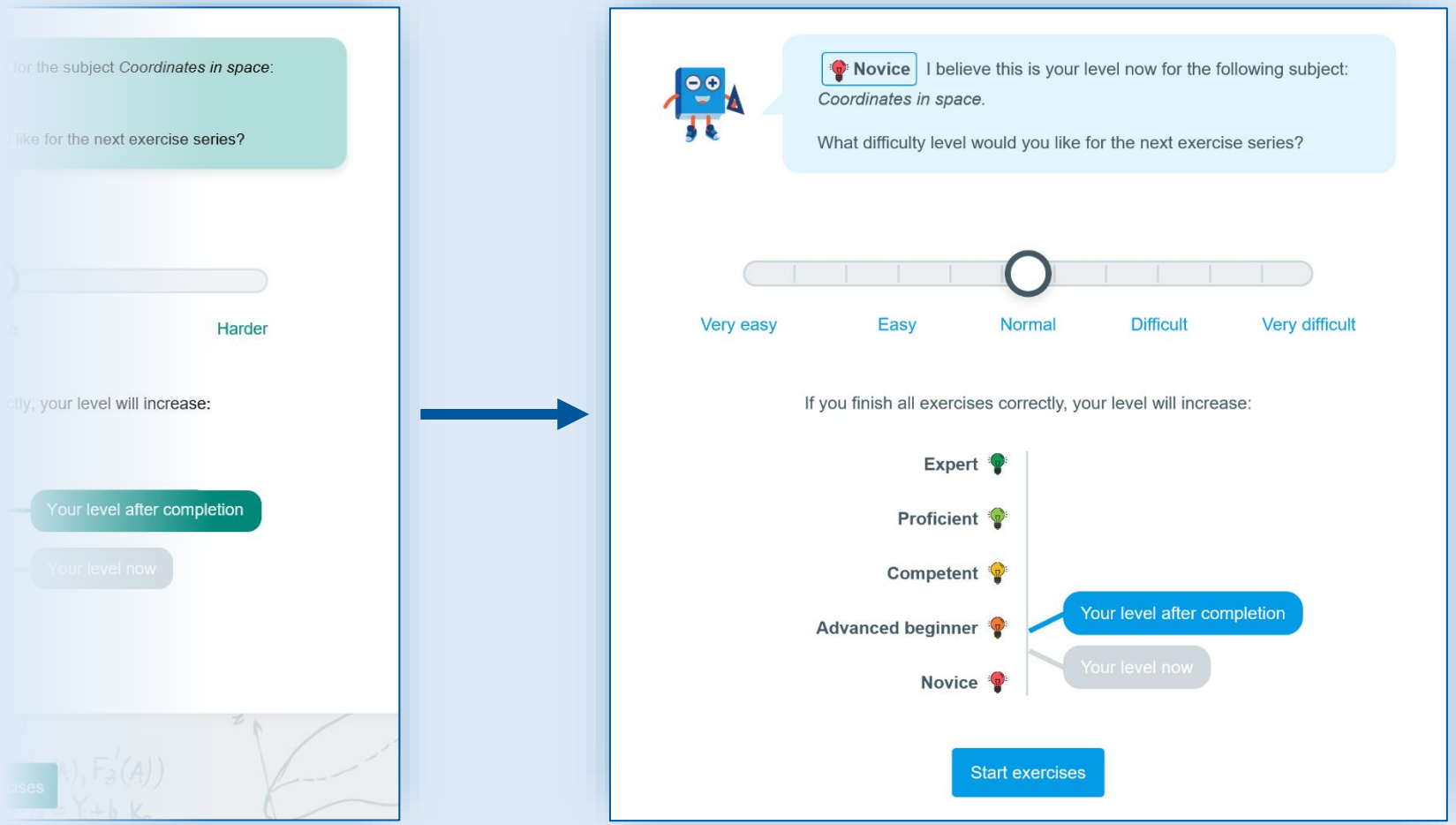
Jeroen Ooge, Maxwell Szymanski, Arno Vanneste, Robin De Croon, and Katrien Verbert. Steer, See Impact, Solve: How Learner Control and Visual Explanations Impact Learning, Motivation, and Trust. To be submitted.





More think-aloud sessions with adolescents

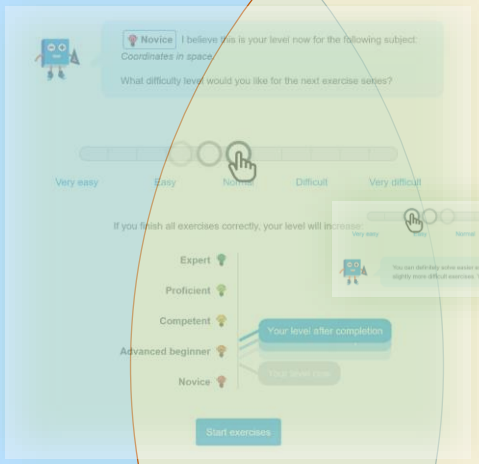
Jeroen Ooge, Maxwell Szymanski, Arno Vanneste, Robin De Croon, and Katrien Verbert. Steer, See Impact, Solve: How Learner Control and Visual Explanations Impact Learning, Motivation, and Trust. To be submitted.

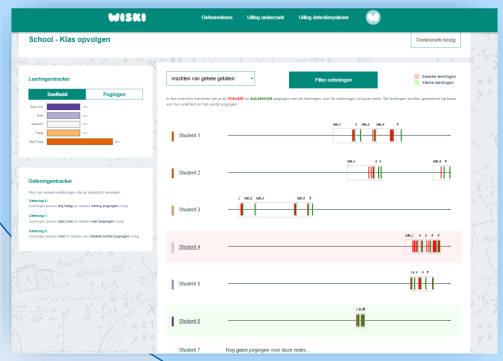


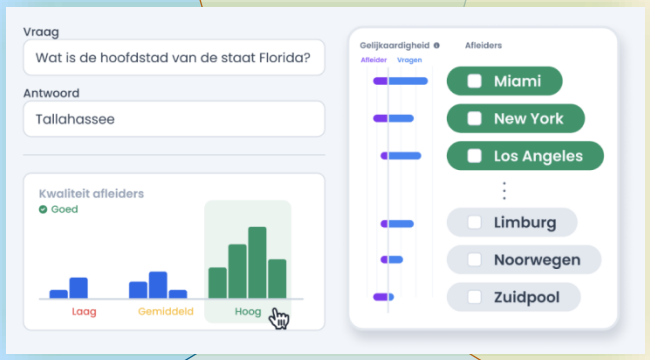
Final prototype for evaluation!

Jeroen Ooge, Maxwell Szymanski, Arno Vanneste, Robin De Croon, and Katrien Verbert. Steer, See Impact, Solve: How Learner Control and Visual Explanations Impact Learning, Motivation, and Trust. To be submitted.

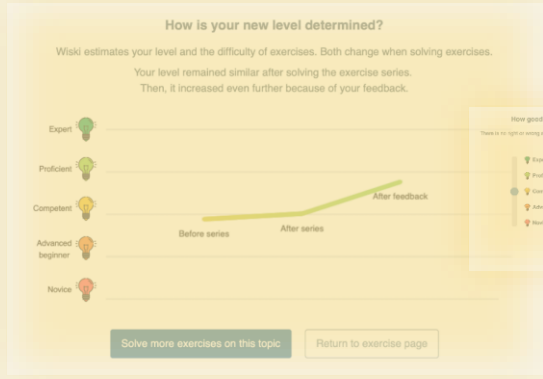
Explanations

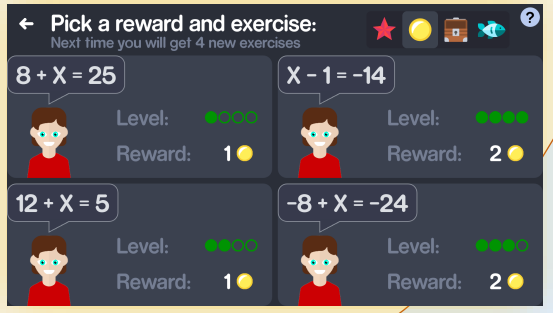


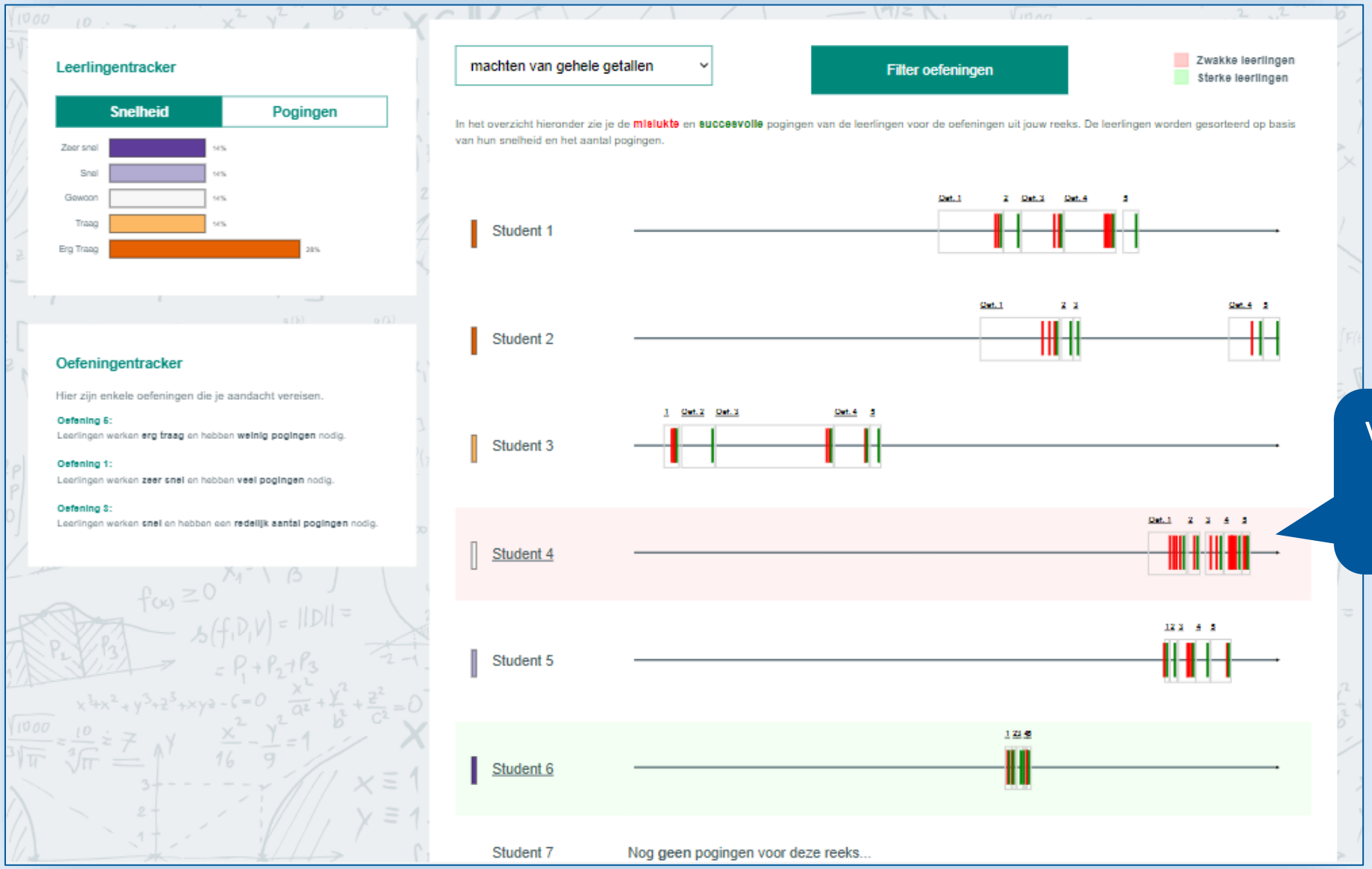




Control



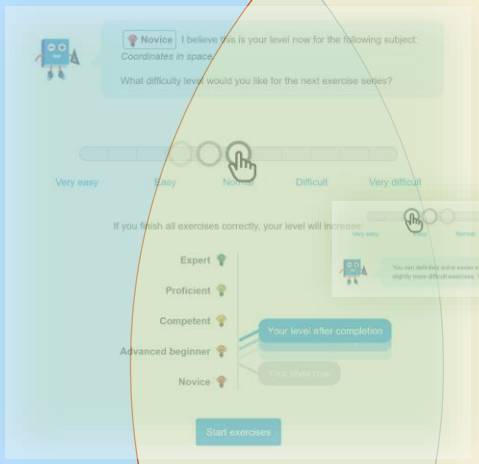


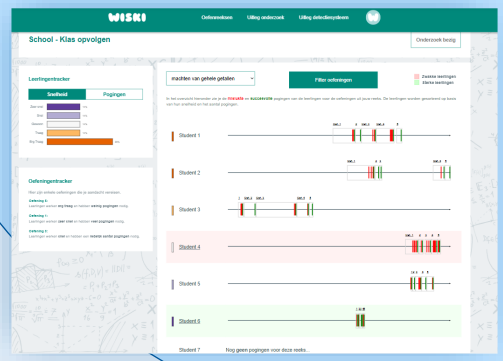


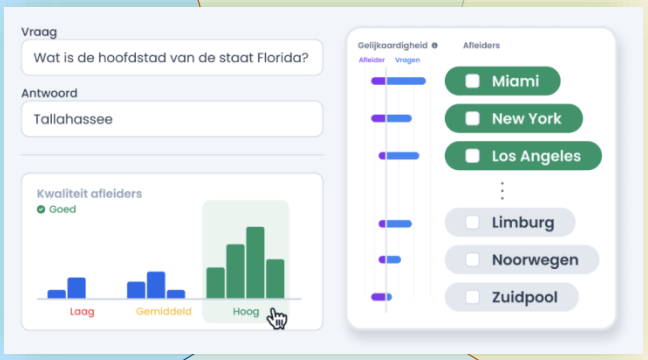
Visualise data used for automated outlier detection

Anissa Faik. Bringing a New Perspective to the Classroom: Detecting and Explaining Student Outliers. Master's thesis, 2023.

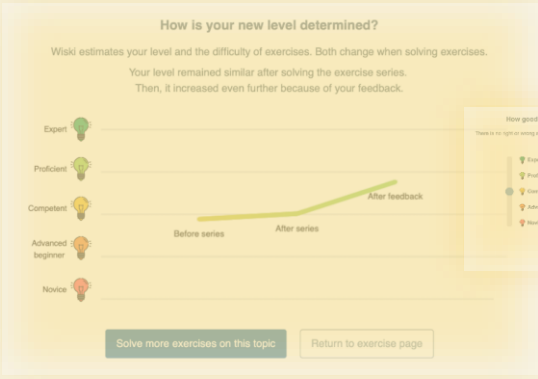
Explanations

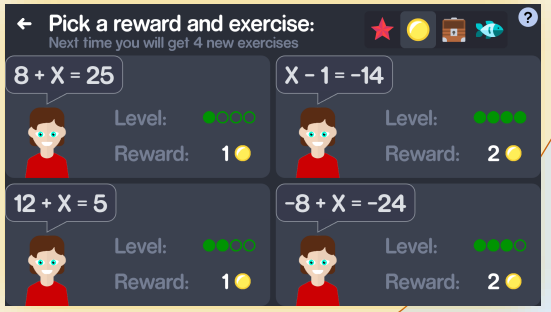


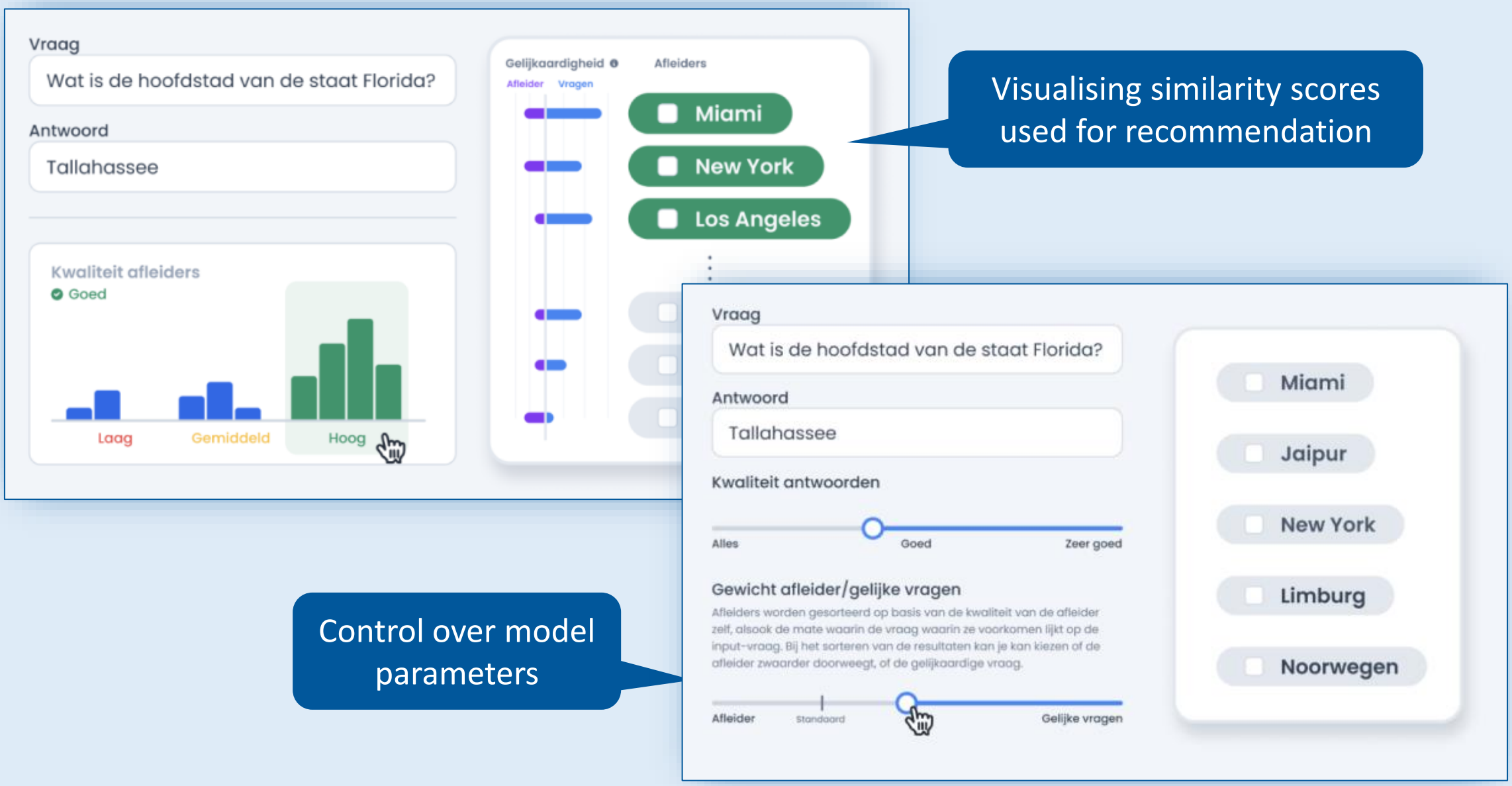




Control

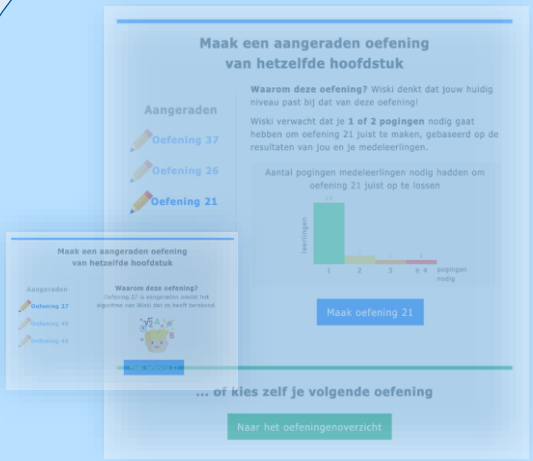


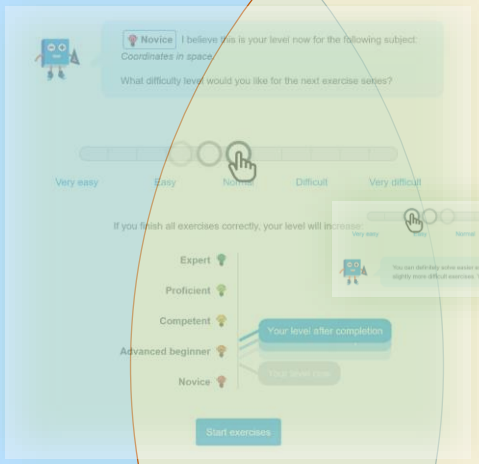


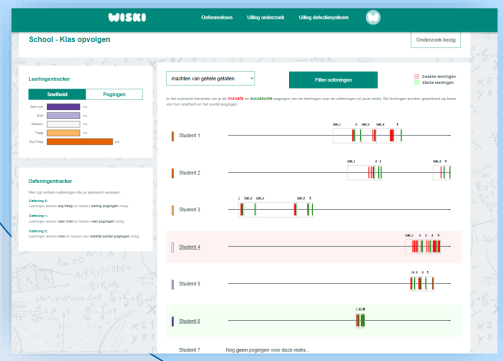


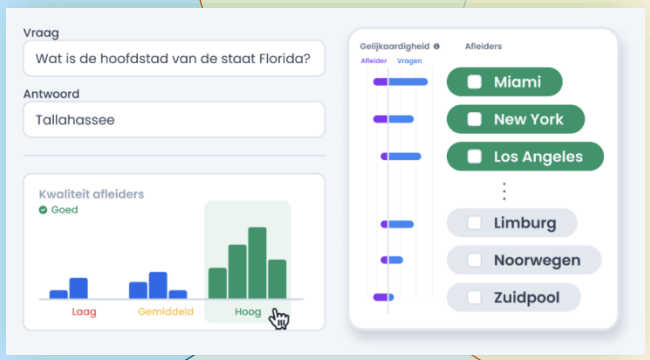
Maxwell Szymanski, Jeroen Ooge, Robin De Croon. Vero Vanden Abeele & Katrien Verbert. 2024. Feedback, Control, or Explanations? Supporting Teachers With Steerable Distractor-Generating AI. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*, March 2024, 690–700, https://doi.org/10.1145/3636555.3636933

Explanations

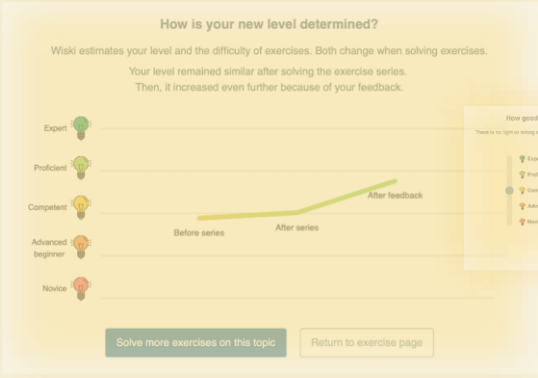


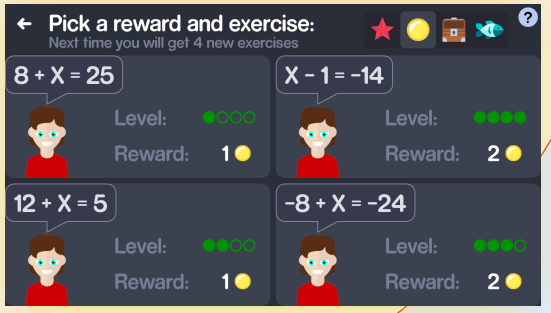


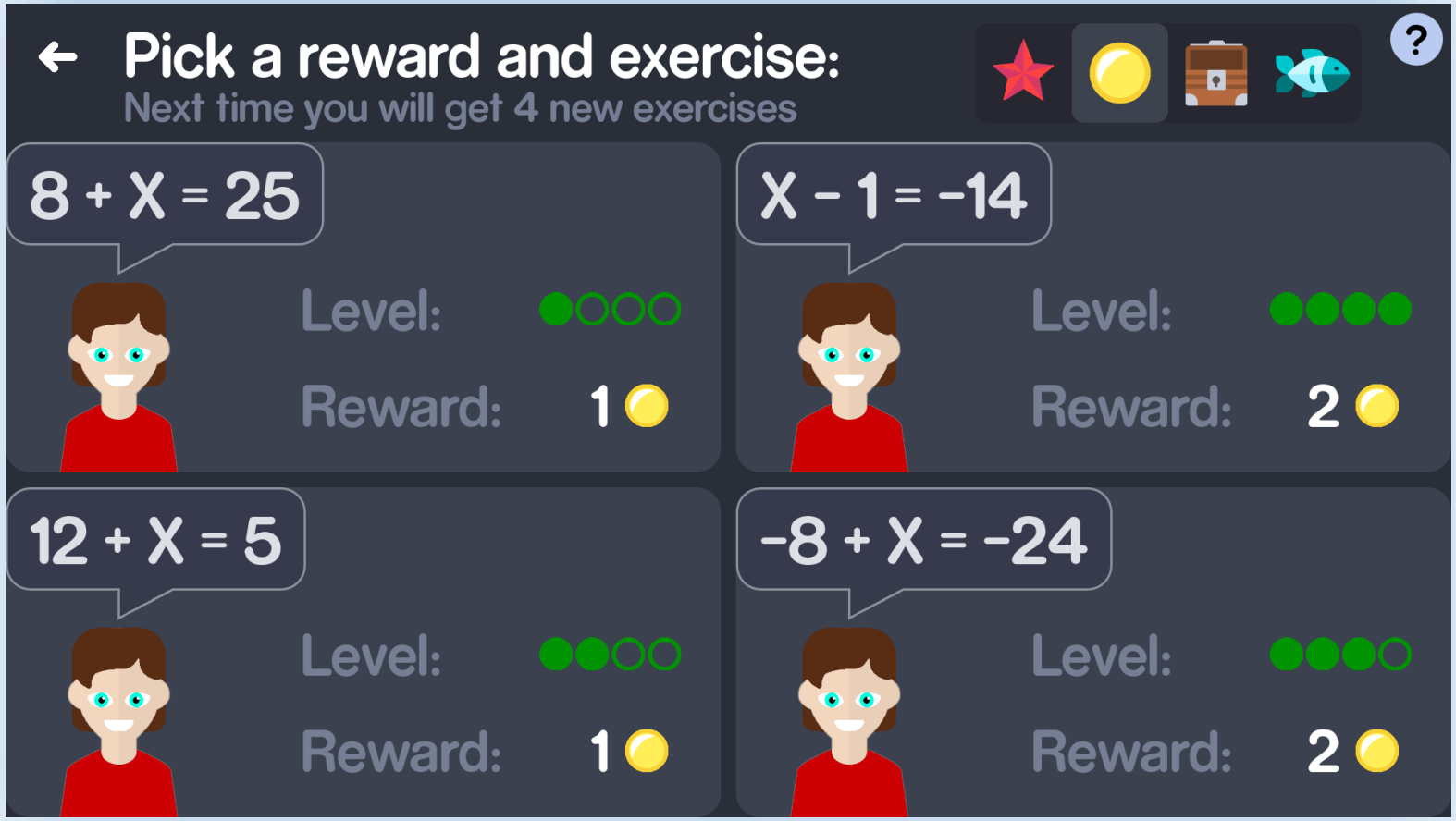




Control



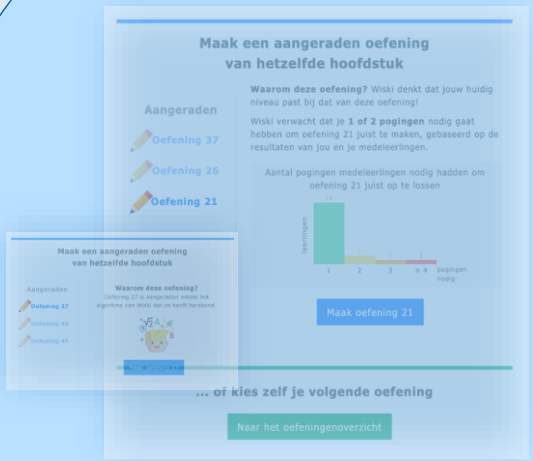


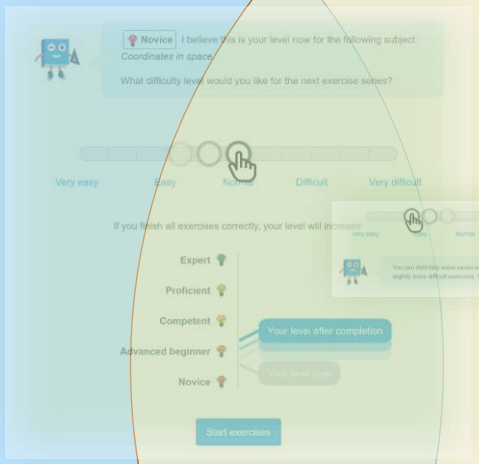


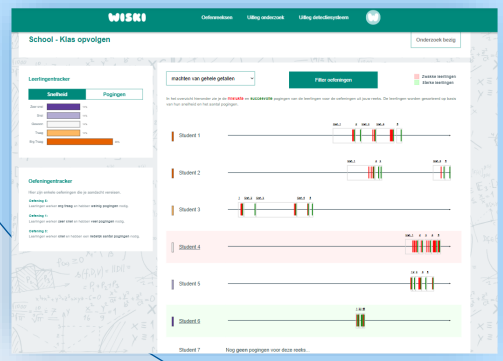
Simple control: choose among recommendations

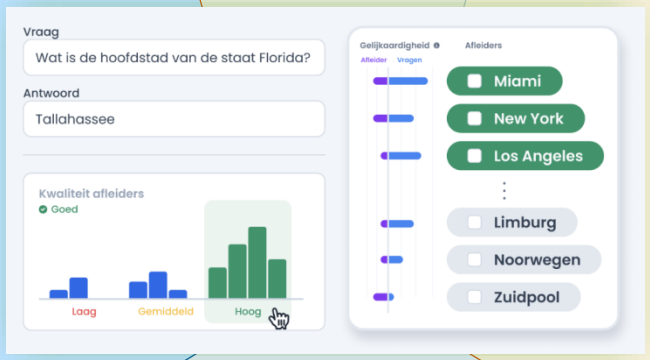
Jeroen Ooge, Joran De Braekeleer, and Katrien Verbert. 2024. Nudging Adolescents Towards Recommended Maths Exercises with Gameful Rewards. In: Artificial Intelligence in Education. AIED 2024. Lecture Notes in Computer Science, vol 14830. Springer, Cham. https://doi.org/10.1007/978-3-031-64299-9_28

Explanations

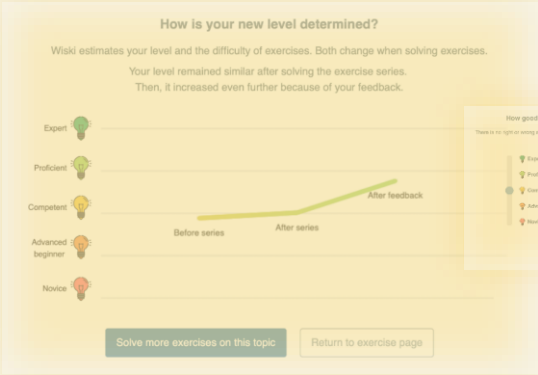


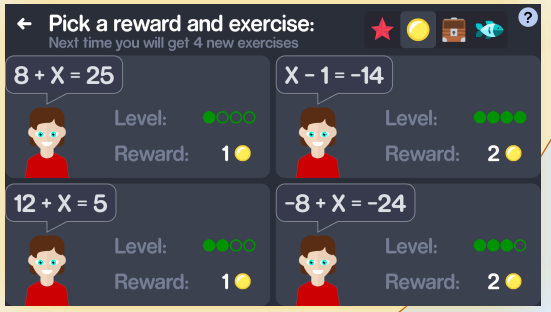




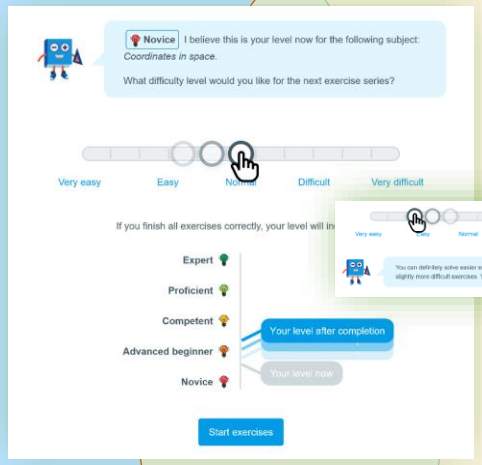


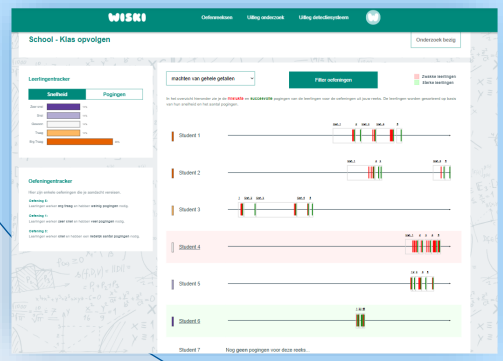
Control

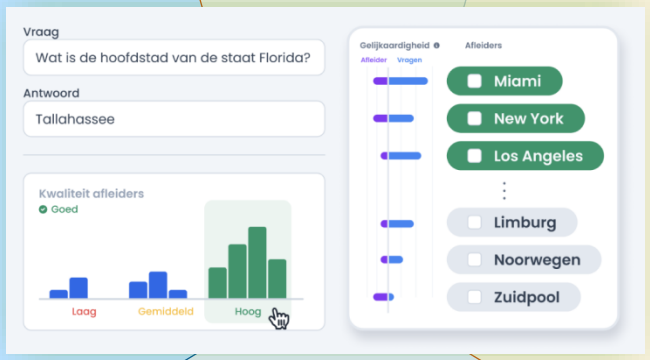




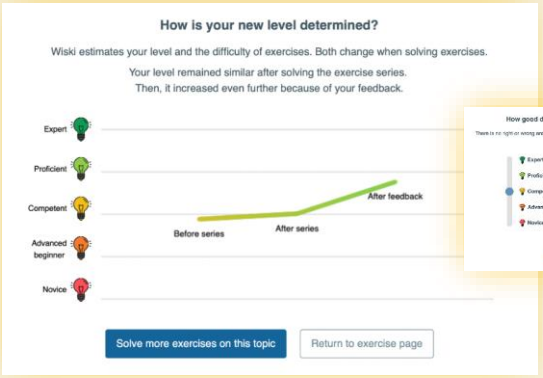
Explanations

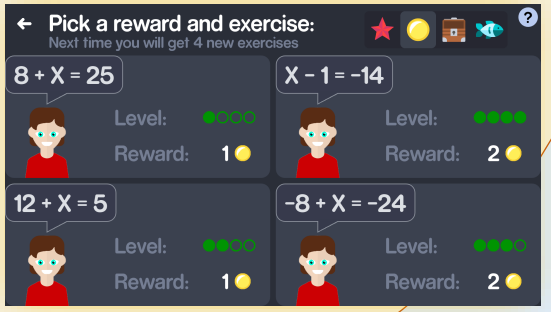






Control





Explainability and Control for Adaptive E-Learning Systems

RECIFE 2024

Jeroen Ooge

jeroenooge.be

Universiteit Utrecht