UNDERSTANDING UNDERSTANDING AI

Andreas Mühling[1] & Lukas Scheppach[2]
[1]Leibniz Institute for Science and Mathematics Education, Germany,
muehling@leibniz-ipn.de
[2] Leibniz Institute for Science and Mathematics Education, Germany

Focus Topics: Explanatory Models, AI and Data Science Competencies

Introduction

With an increasing demand to teach artificial intelligence and machine learning in K12 settings, there are several gaps that computer science education research needs to address. Determining suitable learning goals for the age group is one of them and has led to first curricula, such as AI4K12 (Touretzky et al., 2019). It is designed around "Big Ideas" and "Key Insights" and follows a very basic idea of "opening up black boxes" as far as the mathematical and CS background of the students will allow to make them understand the underlying principles behind the technology so dominant in our everyday lives (Essinger & Rosen, 2011; Mariescu-Istodor & Jormanainen, 2019; Touretzky et al., 2019). This approach is in line with typical science lessons that also aim to help students understand the natural world. It is, however, not in line with a CT or engineering based approach to computer science lessons in which construction and not understanding is the ultimate goal. In this line of thinking, Tedre et al. (2021) have proposed CT 2.0 as a new variant of computational thinking that is not based on sequential, procedural programs but instead on the notion of learnable machines (mostly neural networks).

Regardless of the chosen approach, developing good teaching materials typically involves understanding how students perceive a topic, what kind of prior knowledge they might bring into lessons and what kind of misconceptions might develop. This – together with instructional strategies – can form the basis for a body of pedagogical content knowledge (Shulman, 1986) that teachers should acquire. While there is work on conceptions of machine learning and artificial intelligence (Mühling & Große-Bölting, 2023; Vo et al., 2024; Whyte et al., 2024), we currently do not know much about students' progressions in understanding when learning about AI, neither is there much knowledge about misconceptions or the suitability of teaching approaches.

A Phenomenographic Model

In recent work (Mühling & Große-Bölting, 2023), we explored how beginners conceptualize machine learning based on students' responses and identified a phenomenographic outcome space that is structured along two dimensions: The *learning process* itself and the *internal model* of the learning agent. For the dimension of the learning process, four consecutive stages of understanding – None, Unclear, Repetition and Improvement – were identified, whereas for the internal model dimension there are three stages: None, Implicit and Explicit. A detailed description of the stages including anchoring examples are presented along with the model (Mühling & Große-Bölting, 2023).

This outcome space was also used to classify the responses of students from grades 12-13 prior and after a short 90 minute intervention based on an unplugged activity (Gardner & Michie, 1982) centered around reinforcement learning of a simple game (see Figure 1). Even this short workshop already had a medium effect on improving learners conceptualization regarding the learning process ($W = 999$, $r = 0.36$, $p = 0.0002$), however only a small and non-significant effect ($r = 0.16$) was observable regarding the model dimension.

Operationalizing the Outcome Space

Based on this initial work, we are currently investigating how to operationalize the outcome space into a diagnostic assessment that could be used to determine students' stages of understanding. In a first attempt, we used actual statements from students together with our understanding from coding to create items that students can agree or disagree with on a 5-point Likert scale.

We piloted a version of such an assessment with 12 items in a three-hour workshop on artificial intelligence with students from grades 9 and 10 in a pre-post setting and again could observe an improvement in the learning-process dimension ($W = 113.5$, $r = 0.36$, $p = 0.02$) and a small but non-

In S. Podworny & S. Schönbrodt (Eds.), *Towards Fostering AI and Data Science Literacy in Schools Across Disciplines. Proceedings of the 1st Symposium on Integrating AI and Data Science into School Education Across Disciplines (AIDEA1 2025)*, Salzburg, Austria.

significant effect on the dimension of the internal model (r = 0.10). However, the items also show only a weak internal consistency (Cronbach's alpha = 0.54). As they are combining two dimensions of the original model into one set of items. This is expected to a degree - however it also raises questions about the structure of the items and the overall design of the instrument. A rather large sample might be needed to investigate the internal consistency in such a setting.

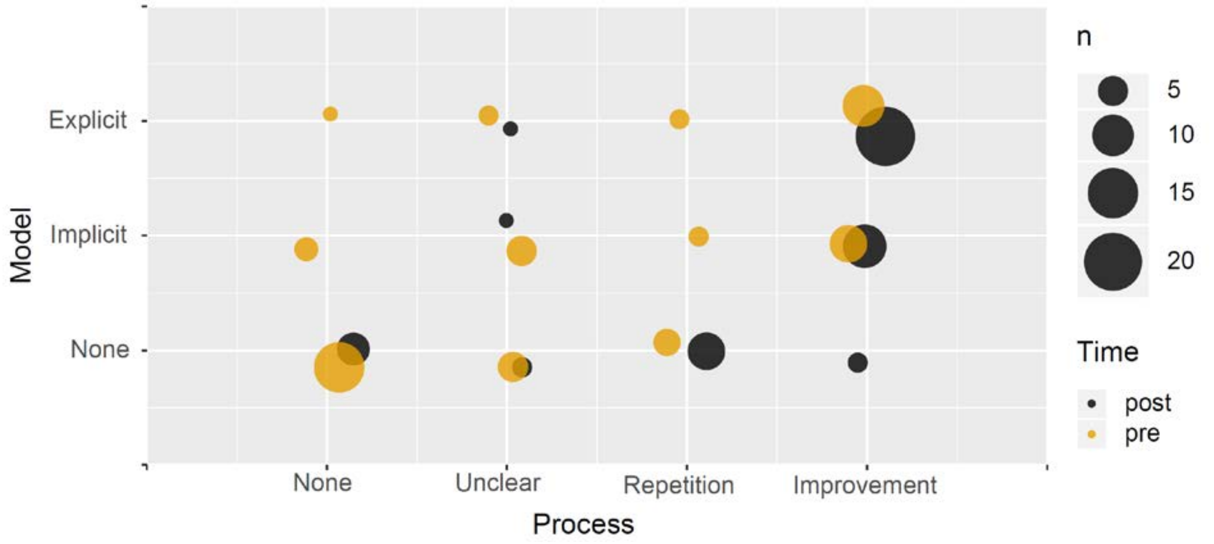

Figure 1: The phenomenographic outcome space and improvements between the classifications of pre- and post-test of a workshop (Mühling & Große-Bölting, 2023)

In an alternative approach we are currently developing concept cartoons (Keogh & Naylor, 1999) centered around each of the two dimensions in order to investigate whether this method might be a suitable way to operationalize a phenomenographic outcome space and diagnose students' understanding. Concept cartoons allow presenting phenomena and possible explanations from the perspective of peers of the learners and thus better align with the idea of phenomenography in which a normative, expert-like or "correct" understanding is not necessarily in the focus of the model but instead a description of the various ways of learners' sense-making (Odden & Russ, 2019).

Concept cartoons were already used by Babari et al. (2023) as a summative and formative assessment tool to assess children's conceptions about the internet. One issue in their work was that multiple inconsistent conceptions about the internet existed in parallel which made it difficult to reveal the conceptions of the test takers. This is problem should not occur in our instrument since the stages here represent different levels of understanding of the phenomenon machine learning and are thus more coherent.

Design of the Concept Cartoons

We designed an initial set of cartoons based on explanations of recommender systems and text-generative AI and piloted them in a workshop with two classes of grade 10 students (Scheppach, 2024). The results – in particular based on open answers that students could give - indicate that students tend to think that the questions are looked up on the internet rather than being generated by a pre-trained model. We used those initial results to refine the cartoons and create a more diverse set for the next round of piloting.

Those newly designed cartoons are based around different apps and websites most students know from their everyday life like Spotify, YouTube, Netflix or Amazon (Feierabend et al., 2024) all using recommender systems to provide their users suggestions. By choosing those apps and websites we try to ensure that most students experienced the workings of the machine learning algorithms behind those applications. Furthermore, we hope to generate curiosity with the cartoons about the question of how these systems provide their recommendation. Both factors are important when it comes to how engaged students feel when answering a questionnaire (Pekrun, 2006).

In each concept cartoon a specific context is given. For example, that they listened to a lot of music of a specific genre and now they get some recommendations from the website or application (see Figure 2). The students then are asked to answer how they think those suggestions are made, using the options presented in the cartoons following the description. For the *internal model* dimension, students have to decide between two different statements representing the None and Explicit stage of the model. They give their answer on a four-point Likert scale reaching from totally agree with person A to totally agree with person B.

For the *learning process* dimension, there are four statements in the cartoons each representing one stage of this dimension. Students then have to provide their agreement on a four-point Likert scale reaching from totally agree to totally disagree for each of these four statements.

To ensure that our results are not biased by the appearance of the persons displayed in the cartoons we designed different versions of the questionnaire were the assigned statements are swapped between the persons.

Validity and Reliability

For concurrent validity we will use the mathematics and computer science grade of the students, their self-assessed prior knowledge in this topic, and the existing questionnaire AILQ by Ng et al. (2024).We expect students reaching a higher stage to also score better in those categories. For prognostic validity we will ask the students to also explain why they decided to agree with a certain statement and also do a think aloud study to get deeper insights into students' thoughts and reasonings when confronted with the concept cartoons.

For reliability we will calculate Cronbach´s Alpha for the answers of each dimension. Since they all measure the same construct, students that disagree with the None statement in one cartoon should also disagree with the None statement in the other cartoons and vice versa. Likewise for the *learning process* dimension.

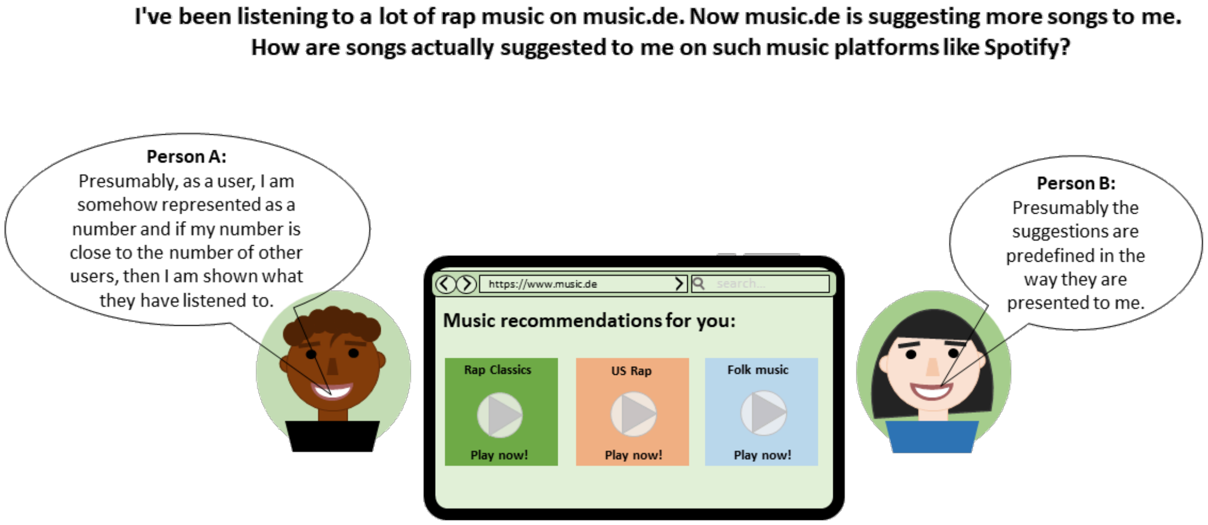
I've been listening to a lot of rap music on music.de. Now music.de is suggesting more songs to me. How are songs actually suggested to me on such music platforms like Spotify?
Person A: Presumably, as a user, I am somehow represented as a number and if my number is close to the number of other users, then I am shown what they have listened to.
https://www.music.de
search...
Music recommendations for you:
Rap Classics
Play now!
US Rap
Play now!
Folk music
Play now!
Person B: Presumably the suggestions are predefined in the way they are presented to me.

Figure 2: Example of a concept cartoon for the internal model dimension (translated from german)

Discussion and Future Work

The work has some limitations, most prominently the limited scope of the intervention – focusing solely on reinforcement learning – that was used to derive the outcome space. Nevertheless, the two dimensions map on the central aspects of the third ("Learning") and second ("Representation and Reasoning") "Big Idea" of the AI4K12 curriculum (Touretzky et al., 2019) and also align well with the modern notion of a learning agent (Russell & Norvig, 2016) that keeps a model of the world it acts in and uses data to improve this model. Both provide some external validity to the structure of the outcome space.

However, this only applied to the “final” stages of each dimension, i.e. the ones that would also be considered “correct” from a normative point of view. From a phenomenographic perspective, it is important to note, that the intermediate stages should not be considered incorrect. They all serve the purpose of explaining a phenomenon subjectively based on the experiences that one encountered so far (Marton & Booth, 2013).

This poses a rather fundamental question of how best to operationalize such a model. Concept cartoons, for example, usually work by combining correct answers with distractors that are derived from known misconceptions. In our case, the stages of understanding do not necessarily represent useful misconceptions however. If a student does understand that a model may be necessary within a learning agent, but does not yet have the capabilities of explicating parts of this model, the student does not hold a misconception. Therefore, designing a distractor that indicates an “implicit” understanding of the model in contrast to an “explicit” understanding – that would be considered correct from a normative perspective, becomes a difficult and eventually maybe even impossible task.

On the basis of these considerations and the results of our initial piloting, we will investigate whether the currently designed concept cartoons, which focus on only two stages of the *internal model* dimension, performs more adequately than our initial attempt.

The final instrument, regardless of its format, can then be used to investigate the effectiveness of teaching interventions and – in particular – how to address the model dimension that currently appears to be not as affected as the dimension of the learning process. Therefore, in another line of future work we aim to look at the validity of the construct, in particular regarding the dimension of *internal model* that may be aligned with a more general understanding of modelling and models, as described in literature (e.g. Upmeier zu Belzen & Krüger, 2010).

Using the Phenomenographic Model in Teaching

Finally, an interesting future aspect to consider is the suitability of the phenomengrpahic model as an explanatory model (Höper et al., 2024) in (K12) teaching. Teaching could then follow along the stages of the model to iteratively deepen students’ understanding. Since the finale stages align well with curricula and experts ideas, the model could present a suitable an empirically derived series of reductions that can be effective as a teaching device. For example, such a progression in teaching could mean that lessons first leave the internal model at an implicit level, while presenting repetition as the core idea of machine learning algorithms and then make the model explicit to also pinpoint what the purpose of repetition is: optimization of model parameters. Alternatively, a series of lessons could start with arriving at an explicit idea of the internal model, while leaving the aspect of optimizing the model based on data on a basic level of understanding and then deepening this understanding of how model parameters are optimized during training.

References

Babari, P., Hielscher, M., Edelsbrunner, P. A., Honegger, B. D., Waldvogel, B., & Marinus, E. (2023). Using Concept Cartoons for Assessing Children’s Conceptions about the Internet. In S. Sentance & M. Grillenberger (Eds.), *Proceedings of the 18th WiPSCE Conference on Primary and Secondary Computing Education Research* (pp. 1–4). ACM. https://doi.org/10.1145/3605468.3605496

Essinger, S. D., & Rosen, G. L. (2011). An introduction to machine learning for students in secondary education. In *2011 Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)* (pp. 243–248). IEEE. https://doi.org/10.1109/dsp-spe.2011.5739219

Feierabend, S., Rathgeb, T., Gerigk, Y., & Glöckler, S. (2024). *JIM-Studie 2024: Jugend, Information, Medien.*

Gardner, M., & Michie, D. (1982). *Logic Machines and Diagrams* (2nd ed.). University of Chicago Press.

Höper, L., Schulte, C., & Mühling, A. (2024). Learning an Explanatory Model of Data-Driven Technologies can Lead to Empowered Behavior: A Mixed-Methods Study in K-12 Computing Education. In P. Denny, L. Porter, M. Hamilton, & B. Morrison (Eds.), *Proceedings of the 2024 ACM Conference on International Computing Education Research - Volume 1* (pp. 326–342). ACM. https://doi.org/10.1145/3632620.3671118

Keogh, B., & Naylor, S. (1999). Concept cartoons, teaching and learning in science: an evaluation. *International Journal of Science Education*, *21*(4), 431–446. https://doi.org/10.1080/095006999290642

Mariescu-Istodor, R., & Jormanainen, I. (2019). Machine Learning for High School Students. In P. Ihantola & N. Falkner (Eds.), *Proceedings of the 19th Koli Calling International Conference on Computing Education Research* (pp. 1–9). ACM. https://doi.org/10.1145/3364510.3364520

Marton, F., & Booth, S. (2013). *Learning and Awareness*. Routledge. https://doi.org/10.4324/9780203053690

Mühling, A., & Große-Bölting, G. (2023). Novices' conceptions of machine learning. *Computers and Education: Artificial Intelligence*, *4*, 100142. https://doi.org/10.1016/j.caeai.2023.100142

Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology*, *55*(3), 1082–1104. https://doi.org/10.1111/bjet.13411

Odden, T. O. B., & Russ, R. S. (2019). Defining sensemaking: Bringing clarity to a fragmented theoretical construct. *Science Education*, *103*(1), 187–205. https://doi.org/10.1002/sce.21452

Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Russell, S. J., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach*. *Prentice Hall series in artificial intelligence*. Prentice Hall.

Scheppach, L. (2024). Towards an Instrument to Assess K-12 Students' Conceptions of Machine Learning Using Concept Cartoons. In J. Leinonen & A. Mühling (Eds.), *Proceedings of the 24th Koli Calling International Conference on Computing Education Research* (pp. 1–2). ACM. https://doi.org/10.1145/3699538.3699574

Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.3102/0013189x015002004

Tedre, M., Denning, P., & Toivonen, T. (2021). CT 2.0. In O. Seppälä & A. Petersen (Eds.), *Proceedings of the 21st Koli Calling International Conference on Computing Education Research* (pp. 1–8). ACM. https://doi.org/10.1145/3488042.3488053

Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K-12: What Should Every Child Know about AI? *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(01), 9795–9799. https://doi.org/10.1609/aaai.v33i01.33019795

Upmeier zu Belzen, A., & Krüger, D. (2010). Modellkompetenz im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften : ZfDN*, *16*. https://doi.org/10.25656/01:31676 (Zeitschrift für Didaktik der Naturwissenschaften : ZfDN 16 (2010), S. 41-57).

Vo, G. M., Kreinsen, M., Ferdinand, R., & Pancratz, N. (2024). Draw, Find, and Describe AI for Me: Investigating Learners' Conceptions of Artificial Intelligence. In *2024 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1–5). IEEE. https://doi.org/10.1109/EDUCON60312.2024.10578628

Whyte, R., Kirby, D., & Sentance, S. (2024). Secondary Students' Emerging Conceptions of AI: Understanding AI Applications, Models, Engines and Implications. In T. Astarte, D. Hull, F. McNeill, & F. Moller (Eds.), *Proceedings of the 2024 Conference on United Kingdom & Ireland Computing Education Research* (pp. 1–7). ACM. https://doi.org/10.1145/3689535.3689552