Validating a Hypothetical Learning Progression (LP) to Support Upper Elementary School Students to Learn and Apply Artificial Intelligence Concepts

Srijita Chakraburty, Cindy E. Hmelo-Silver, Krista D. Glazewski, Anne Ottenbreit-Leftwich, Jiyoung Kim, Vanessa Johnson, Dubravka Svetina Valdivia
srichak@iu.edu, chmelosi@indiana.edu, glaze@indiana.edu, aleftwic@indiana.edu, jk142@iu.edu, vanjohn@iu.edu, dsvetina@indiana.edu
Indiana University

Bradford Mott, James Lester
bwmott@ncsu.edu, lester@ncsu.edu
North Carolina State University

Abstract: This paper proposes a novel learning progression contributing to K-12 AI education by providing a framework that outlines learning trajectories for younger students. It identifies potential entry points and offers a clear description of increasing levels of core AI concepts. To ensure reliable student placement during validation of this LP, we compared the actual difficulties of assessment items, used for collecting evidence of student learning, against hypothesized item difficulties by generating Wright maps. We further explored variations in students' progression within different classroom contexts. Findings from the validation process revealed a consistent increase in students' progression across LP levels from pre- to post-tests for most constructs. Classroom comparisons indicated significant differences in students' progression levels for certain constructs. Further exploration through analysis of classroom videos highlighted differences in instructional strategies, likely attributed to the teacher's content-matter expertise, impacting student understanding of the content and hence their progression on the LP.

Introduction

Over the past decade, technology has continued to permeate our lives and evolve at a rapid pace, touching almost every aspect of our existence, particularly with the ever-present influence of Artificial Intelligence (AI). To understand what AI is and how this innovative technology functions, AI education has started to make its way into the K-12 educational landscape (DeLyser & Vargas-Vite, 2021). Despite some K-12 curriculum development efforts that have begun to incorporate AI concepts (Judd, 2020; Payne 2019), there is limited research into how to effectively measure students' understanding of these concepts and their application.

As AI's prevalence continues to grow, Wong and colleagues (2020) assert that K-12 students must acquire proficiency in three key dimensions of AI literacy: concepts, applications, and ethics/safety. Within this context, researchers have identified competencies in AI ethics, decision-making, data analysis, sensors, and data-driven learning. These competencies serve as the foundation for developing frameworks in curriculum design and learning paths. Furthermore, AI4K12, designed as a joint initiative between the Association for the Advancement of Artificial Intelligence (AAAI) and the Computer Science Teachers Association (CSTA), has identified five "big ideas" pivotal to AI education: (a) perception, (b) representation and reasoning, (c) learning, (d) natural interaction, and (e) societal impact (Touretzky et al., 2019). These frameworks are helpful in designing K-12 AI curricula; however, the field needs a description of learning trajectories and guiding standards for younger students (Ottenbreit-Leftwich et al., 2023). Guiding frameworks can help provide insight regarding what topics to consider, however, researchers have recently started to map pathways that will inform ways to better engage young K-12 learners with ideas about AI. There needs to be more focus on choosing concepts that are developmentally appropriate for young learners, as well as defining learning trajectories. An essential element of these trajectories is using reliable assessment items which can help to measure student understanding at each level of the expected learning trajectory and must undergo psychometric and cognitive testing to ensure they are reliable and valid. This is pivotal in measuring a student's grasp of the construct in the progression (Wiliam, 2010). In our study, we designed a hypothetical AI learning progression (LP), drawing from domain specific research, validated assessments, prior implementations of our AI curriculum (Chakraburty et al., 2023), and expert input. Next, we validated this LP in two classrooms, while considering potential cultural influences. The general aim of this research is to design and validate a hypothetical LP to support upper elementary school students to learn and apply

AI concepts. We address this through three research questions: (1) How does the progression of upper elementary students' knowledge of AI concepts unfold? (2) Where do the students lie on the progression before and after the curriculum intervention?, and (3) To what extent does students' progression differ depending on different classroom contexts?

Literature review: Learning Progressions

Learning Progressions (LP) start with a few foundational ideas in a specific content area that are classified into progress variables (Wilson & Scalise, 2006). These foundational ideas are generative disciplinary ideas that are built and refined over time. They consist of different levels describing the development of students' understandings that are research-based (Duncan & Rivet, 2018). Learning progressions have been developed for many areas of science, such as complex systems (Cisterna et al., 2020), genetics (Shea & Duncan, 2013), and physical science (Kaldaras et al., 2021). Learning progressions have a lower anchor representing initial understanding and an upper anchor representing expert understanding with several points in between. Given the complexity of content in AI and the inadequacy of current efforts for delivering and measuring AI learning, especially among younger learners, a more coherent approach to understanding students' development of AI knowledge is needed. LPs contribute to educational coherence in three distinct ways. First, they map the development of students' understanding, guiding their progression from basic to advanced thinking, establishing developmental coherence. This foundational understanding then facilitates the alignment of educational content, fostering horizontal coherence, while also bridging the gap between classroom-level assessments and broader evaluations, promoting vertical coherence. To further advance this field, we need research to expand the use of LP-based interventions and to enhance teachers' understanding and implementation of LPs (Jin et al., 2019).

To prepare students to step into the world of AI, they need to be introduced to these complex concepts early. In their study, Ottenbreit-Leftwich et al. (2023) discuss the significant challenges associated with designing a curriculum that introduces AI content to fourth and fifth grade students. One challenge they identify is students' prior knowledge about AI, including naïve ideas, by investigating their everyday experiences and ideas about AI. This highlights possible entry points to designing a learning trajectory for AI learning for younger students. Wong et al. (2020) argue that although there are lessons to be learned from university-level AI education, they cannot be completely implemented at a K-12 level in the same way. We argue that an AI LP that defines the learning trajectory across multiple grade-bands from elementary to high school would ensure students' preparation for higher level AI courses including preparation as citizens with fundamental AI literacy. With younger learners' experiencing AI devices and media representations at an early stage of their lives, they bring various pre-existing perceptions and ideas. It is important to critically assess student understanding of AI concepts to better evaluate their current perceptions of AI and to design a developmentally appropriate curriculum (Long & Magerko, 2020). These LPs are initially hypothesized based on expert targets and empirical evidence. To understand these instructional targets and how students are making progress in achieving them, reliable assessments are needed. These hypothetical models require testing and validation through iterative revision and refinement (Duncan & Hmelo-Silver, 2009). Validation of hypothetical LPs can occur through cross-sectional studies documenting knowledge and reasoning development across multiple grades (Mohan, Chen, & Anderson, 2009) or through longitudinal teaching sequences (Songer et al., 2009). It's essential to empirically evaluate LPs as the development of students' AI understanding isn't inevitable (Duncan & Rivet, 2018). In the field of AI, we need frameworks to support students in progressively developing more advanced AI knowledge and reasoning. In this study, we designed a hypothetical LP by conducting cognitive analysis, using data such as students' scores on assessment items from previous PrimaryAI curriculum studies (Chakraburty et al., 2022), classroom observation videos, existing research, and expert consultations. The second part of the study validates this LP in two upper-elementary classrooms with different cultures, grounded in socio-constructivist theories. We examine how learners understand complex AI concepts in relation to the LP progression and explore potential differences based on teachers' strategies and their impact on students' learning.

Data sources and analysis

To understand the functionality of our data collection instrument, we created Wright-maps for two scales that underwent psychometric testing. These maps illustrate both person abilities and item difficulties on the same scale, enabling visual examination on a single graph. We employed the Rasch model to generate these maps and estimate item difficulties. The scales were developed using assessment items from two years of data collection (n=105) in six Midwest classrooms during the implementation of our AI curriculum for upper-elementary students. To validate our LP, we collected data from two semi-urban schools in the Midwest (n=35). One school was taught by a content matter expert (a researcher on the team), while the other had limited content matter knowledge. Pre-and post-test scores placed students on the progression, and we conducted a frequency count of how many students progressed across the levels as the initial step of LP validation. Next, we used chi-square analysis to assess differences between the two classrooms for each construct. Additionally, we analyzed video data using interaction analysis (Jordan & Henderson, 1995) to examine teachers' instructional strategies and alignment with the content of the AI learning progression.

Learning Progression design

We initiated the development of the Learning Progression (LP) by researching domain standards and aligning it with the PrimaryAI curriculum. The curriculum, developed in accordance with AI4K12 standards, included lessons for each objective derived from these standards, aiming to teach distinct AI concepts. To create a learning progression that effectively supports the teaching of this content, it is crucial to establish clear distinctions among the various ideas targeted by these objectives. To articulate these levels and refine the progress variables further, we conducted a thorough review, which included analyzing classroom observation videos from prior intervention studies of the AI curriculum. We also performed a cognitive analysis to ensure that the progress variables accurately reflected increasing levels of sophistication in understanding AI concepts. To assess student understanding, we assigned assessment items from the AI curriculum to different LP levels. A team of content matter experts then reviewed the hypothetical LP to verify the content matter validity of the main ideas defined through various progress variables. We adapted lower anchor concepts from a previous study, and we refined higher levels through cognitive analysis. For example, an increase in sophistication may involve progressing from a basic understanding that computers use pixels to a more advanced comprehension of how pixels relate to RGB color bands in image processing. The LP is organized around six main ideas (Table 1). We conducted our data analysis in two stages. First, we examined whether items met their intended difficulty levels. Second, we investigated evidence from two classrooms to confirm the hypothesized progression of student understanding and to identify contributing factors to any differences in student progressions.

Results

Instrument functioning and Learning Progression Validation

To examine instrument functioning for the validated items, we reviewed the Wright Maps (Figure 1) for two subscales, mapping assessment items to different constructs. We assessed item difficulties to ensure alignment with hypothesized values. The item prefixes (CV/ML) indicate the scale, and the number (1/2) denotes the item's position within that scale.

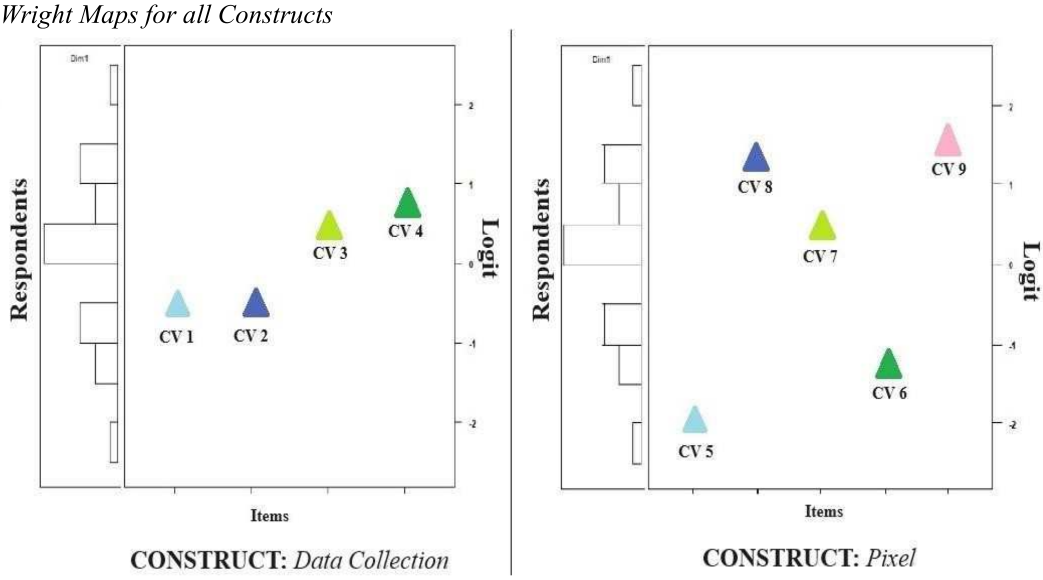
Table 1
Learning Progression in Artificial Intelligence

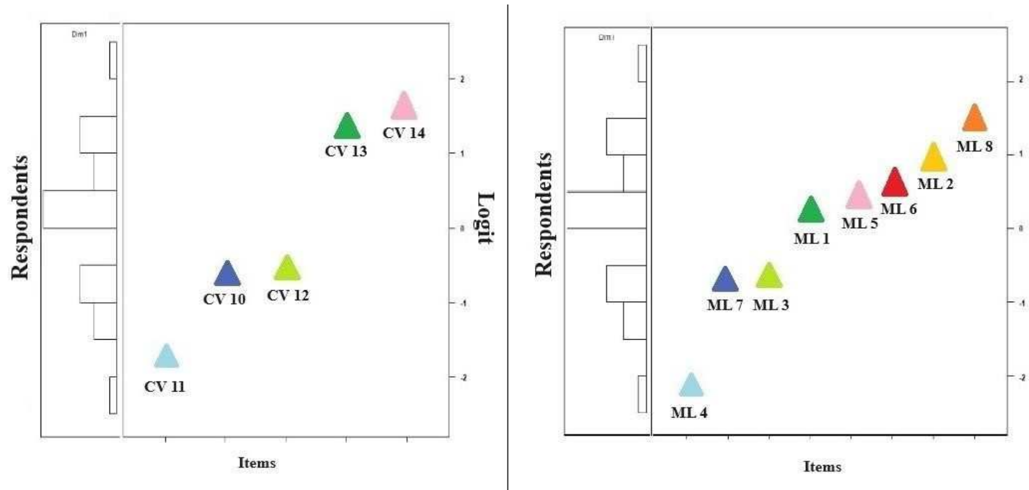
Constructs	1- Basic	2- Developing	3- Proficient
Data Collection	Recall that humans use their senses to collect data.	Identify that computers use different sensors (similar to humans) to collect data.	Explain the relationship between data, learning, and decision-making.
Pixels	Recognize that computers see images using pixels.	Explain that each pixel has a number associated with it.	Demonstrate the relationship between the numbers in the pixels and the color bands they represent (RGB).
Edges	Identify what an *edge-detected* image by a computer looks like.	Describe how computers find edges by comparing the RGB number values and finding patterns.	Explain how computers find and classify shapes using the detected edges.
Feature Extraction	Recognize that computers use rules called *algorithms* to classify the different shapes in an image.	Identify that the more detailed an algorithm is the more accurate the classification of the shape/image is.	Recognize that neural networks are used to combine different concepts like edge detection and feature extraction to see and classify pictures.

Learning	Identify that there are three different ways in which machines learn *supervised, unsupervised, and reinforcement.*	Differentiate between the different types of machine learning.	Compare and contrast the various applications that use different types of machine learning.
Data and Bias	Explain that the quality of the data is more important than the quantity of data in training a model.	Identify what bias in data looks like.	Explain the relationship between biased data and the decision-making it leads to.

We looked at item difficulties for the constructs: Construct 1: Data Collection; Construct 2: Pixels; Constructs 3 & 4 combined: Edge Detection and Feature Extraction; Constructs 5 & 6 combined: Learning, Data, and Bias. We combined the last two constructs due to limitations in the number of items in the individual constructs and the interconnected nature of the concepts, preventing us from generating separate Wright Maps for each construct. Our aim was to have at least 4-5 items per construct to generate these maps. We observed that for Constructs 1 (Data Collection) and 3 & 4 (Edge Detection and Feature Extraction), the actual order of item difficulty exactly matched the hypothesized order, confirming our conjecture. For Construct 2 (Pixels), two items did not match the hypothesized order. However, further investigation revealed that at least one item from each level followed the expected order, leading us to conclude that these items would ensure the reliable placement of students on the trajectory. For the last two constructs, (Learning and Data & Bias), four items were identified that did not match the expected order of difficulty. This was expected because as concepts become more complex, the nature of various topics within each idea may not strictly follow a linear progression. Understanding if the actual item difficulties matched the hypothesized item difficulties was an important step before we moved on to the validation of the progression since these items that were mapped to the different levels of the progression played an important role in the reliable placement of the students on the progression.

Figure 1
Wright Maps for all Constructs





CONSTRUCT: *Edges and Feature Extraction* **CONSTRUCT:** *Learning and Data & Bias*

We aimed to validate the LP by assessing students' understanding using pre-test and post-test scores gathered during the implementation of the AI curriculum intervention. A total of 35 students completed the Computer Vision (CV) subscale tests, while 28 students completed the Machine Learning (ML) subscale tests. Items from the CV subscale were used for the first four constructs, and items from the ML subscale were used for the latter two. Our hypothesis was that, for each construct, students would progress to higher levels in the post-test compared to the pre-test. Additionally, we expected students to generally perform at lower levels in the later constructs compared to the earlier ones.

Results indicated a consistent increase in levels for the first two constructs, Data Collection and Pixels, among all students. However, starting from the third construct, we observed a change in this trend, along with a decline in the number of students progressing across levels for certain constructs: Construct 3 (Edges) and Construct 6 (Data & Bias) from levels 0 to 3. The results for the last construct (Data & Bias) were not surprising, as the curriculum didn't emphasize the concept of bias extensively and only briefly touched upon the importance of data quality over quantity. We are actively working on incorporating more classroom activities related to this construct for future iterations. We also examined how many students progressed across levels from pre to post-tests (see Table 2). We observed a similar trend of initial progress across levels for the first few constructs, followed by a reduction in the number of individuals progressing across levels, or in certain cases, even dropping levels as the ideas behind the constructs became more complex (left to right in Table 2). This suggests that as ideas become more intricate, we need to provide additional curriculum resources and support to address the nuanced learning of these abstract concepts.

Table 2
Progression across levels before and after curriculum

Level Progression	Data Collection	Pixels	Edge	Feature Extraction	Learning	Data & Bias
Same Level	17	13	21	18	10	23
1 level up	14	8	7	0	7	2
2 levels up	4	13	3	15	7	0
3 levels up		1	1		2	
1 level down			3	2	2	3

The two fundamental questions proposed by Duschl et al. (2011) for evaluating LPs are as follows: "How well developed is the identification of foundational knowledge that facilitates and advances pathways of reasoning and understanding? How thorough is the description of the teacher-mediated learning pathways?" (Duschl et al., 2011, p. 173). While validating LPs, the first question helps us understand the coherence of the curriculum, and the second question addresses the alignment between the curriculum, instruction, and assessment. In the validation of our LP, we examined two classrooms where the same curriculum was taught, and we assessed student learning using the same assessment items. However, differences in classroom context and teachers' expertise levels prompted us to explore how teachers' instructional choices might have impacted student progression in the two classrooms.

Classroom comparison

Evidence from LP validation should consider teachers' pedagogical content knowledge, which can impact their instructional strategies and, consequently, students' learning outcomes. In our study, the two classrooms we used for validation had different levels of teacher expertise. Jennifer was new to AI teaching, while Alison was an AI expert. We used chi-square tests to determine if there was a significant association between student progression levels and the two teachers. Out of the six constructs studied, there were significant associations between the constructs and classrooms for Feature Extraction and Learning), with $\chi^2(2)=8.40$, $p=.003$; $\chi^2(2)=9.77$, $p=.007$, indicating a notable association between progression levels and classrooms, indicating that Alison's students moved further along the LP on these constructs. To delve deeper into the differences observed in the constructs of Feature Extraction and Learning between teachers' classes, we conducted interaction analysis on lesson videos. We reviewed 110 minutes of video footage from both classrooms, selecting segments where lessons on the constructs of Feature Extraction and Learning were being taught. Due to space constraints, we discuss the results for Feature Extraction here. We present our findings from two activities in the following section, emphasizing the nuances of student responses across the two classrooms to highlight differences in their understanding of the discussed concepts and the role teachers play in eliciting these responses. The three LP levels of this construct are mentioned in Table 1.

<u>Activity 1: Quick, Draw</u>

The first activity featured the game Quick, Draw! by Google, where players drew objects, and the AI either correctly identified the drawing by recognizing patterns from its database of previously drawn examples of the same object, or the player ran out of time. This activity addressed the concept that: *Computers use algorithms to classify shapes in an image.*

We observed students actively engaging in the activity right from the beginning in both classes. In Jennifer's class, when the game failed to identify the objects some students drew, Jennifer addressed certain student concerns, such as "my panda... said it looks like sunglasses or a donut," by probing them to understand why that might have happened. She asked, "What do sunglasses, pandas, and donuts have in common?" This led them to think about the common shape across the three objects, with one student remarking, "if that's a *circle*, it could be sunglasses or a doughnut." Similarly, in Alison's class, student responses indicated that they also reached an understanding of a common shape that was being used to identify similar objects. One student said, "Well, I see a lot of *rectangles*; some look like a wallet... a *rectangle* is definitely a common shape." However, we observed Alison further probe by saying, "think along the lines of *edge detection* and *feature extraction*." This led the students to expand their thinking and grasp the nuances of how these shapes are classified. When discussing why the game might have mistaken a drawing of stairs as that of a chair, one student responded, "Just a little bit more of an extra *edge* piece makes it (a chair), stairs." Like the panda, sunglasses, and donut discussion in Jennifer's class, students in Alison's class also delved into the identification of common shapes across different images, such as rectangles representing a toaster or a wallet. However, we observed more sophisticated thinking and the discussion reaching a higher level of complexity in Alison's class, where students connected this to prior concepts and identified how the addition of extra *edges* could transform an image from a chair into stairs.

<u>Activity 2: Dog vs Table</u>

In this second activity, the teachers engaged their classes in a discussion about establishing rules to distinguish between a picture of a dog and a table. This activity addressed the concept that: *The more detailed an algorithm is the more accurate the classification of the shape/image is.*

In Jennifer's class, student responses indicated that they were able to design different rules for the two objects. Some student responses were, "focus on the four legs for a dog," "a table would have a flat surface," "a dog would have two triangles for the ears." Although Jennifer acknowledged these responses, she didn't probe further to have a discussion on how these rules were similar to a certain extent but different because of the intricacies of the two images. There was no discussion in the class along the lines of comparing the rules of the two objects. In Alison's classroom, the students started by designing rules for the two objects separately, similar to Jennifer's class. Some of these rules were, "specifically talking about the head and its ears for the dog," "the table has four legs." Alison then asked them to compare these rules, which led to further responses like "The table has a flatter surface than the dog," "dignified markings on the table compared to the dog." The students then had a discussion on how the rules for the dog might not apply to a table but might apply to a cat when one student mentioned, "but the circle with two triangles could be a cat too." This was quickly refuted by another student who said, “we can be more *specific* about the length of the ears...for a dog.” The discussion in Alison’s class once again touched upon the nuances of how to detail algorithms or rules for more accurate classification (e.g., dog vs. table; dog vs. cat), whereas in Jennifer’s class, the discussion stopped at creating different rules for various shapes/images, without delving into how these rules may vary and require different levels of detail for accurate classification. For both activities, students in both classrooms demonstrated a good understanding of the fundamental concepts of algorithms. However, it was evident that students in Alison’s classroom engaged in more sophisticated discussions, placing greater emphasis on the intricacies of rules, particularly when dealing with complex images, and drawing connections to previous concepts like *edge-detection*. Although the initial responses in both classes were similar, Alison’s discussion prompts encouraged students to think more elaborately about these concepts. This observation can be attributed to the abstract nature of the content being taught. Alison's strong grasp of the subject matter as a content-matter expert provided her with an advantage in handling the nuances of the concept compared to Jennifer. While Jennifer delivered the content as expected and demonstrated understanding, she couldn't make these seamless connections with the same ease.

Conclusions and implications

Our study concentrated on creating an AI learning progression for upper-elementary students. We validated it by assessing two classrooms with reliable items for varying understanding levels. During validation, we observed a consistent increase in student comprehension for four out of six constructs from pre- to post-tests. However, our results also indicated that as concepts grew more complex, progress across levels decreased, suggesting a potential need for additional resources to support the understanding of increasingly abstract concepts. This observation aligns with findings in the learning sciences literature, which highlight the nuanced nature of knowledge structure and the processes of learning (Duschl, 2008). This specifically holds true for AI learning, considering the relative “newness” and abstract nature of the domain. Our proposed LP provides a framework for designing and refining AI curricula and ensuring alignment with assessments and instructional strategies. Additionally, we conducted a comparative study across two different classroom settings to examine the impact of various contexts and instructional strategies on students' comprehension of complex AI concepts. We argue that learning is not inevitable, and the context and quality of teaching play a crucial role in student learning. To investigate this, we compared student progression between two classrooms: one taught by an AI expert (Alison) and one with limited AI teaching experience (Jennifer). This comparison revealed significant differences in students' progression levels for two of the more complex constructs: Feature Extraction and Learning. Analysis of classroom videos showed notable differences in instructional strategies, likely attributed to the teacher's expertise and hence ability to improvise teaching the content depending on need. This finding was expected, given the sophistication and nuanced nature of these concepts, which an AI expert can more effectively address. This highlights the importance of pedagogy, particularly in teaching AI concepts. Our findings underscore the need to define and elaborate on instructional progressions that can support students' understanding of these core AI ideas within our LP. The challenge lies in deciding what and how much to emphasize and what to exclude.

Future research should focus on enhancing teachers' capacity and expertise in AI education by integrating pedagogical knowledge with AI understanding through instructional strategies guided by the LP. In our next phase of refining assessment items and curricula using this LP framework, we intend to design a broader range of assessment items and learning performances to gather additional evidence of student understanding before placing them on the progression. Our project has identified possible entry points (lower anchors) and defined the increasing levels of core AI ideas, making a novel contribution. This is particularly important as the AI education field for K-12 currently lacks a framework that describes learning trajectories guided by standards for younger students. Such a framework will also facilitate the development of coherent AI curricula, aiding students in building a sophisticated understanding of key AI concepts.

References

Ottenbreit-Leftwich, A., Glazewski, K., Jeon, M., Jantaraweragul, K., Hmelo-Silver, C. E., Scribner, A., ... & Lester, J. (2023). Lessons learned for AI education with elementary students and teachers. *International Journal of Artificial Intelligence in Education, 33(*2), 267-289.

Chakraburty, S., Jeon, M., Glazewski, K. D., Hmelo-Silver, C. E., Ottenbreit-Leftwich, A., Jantaraweragul, K., & Lester, J. (2023). An Analysis of Teacher Practices and Student Participation in Contrasting Activity Systems in an AI Educational Program. In Proceedings of the 17th *International Conference of the Learning Sciences-ICLS 2023, pp*. 1026-1029. International Society of the Learning Sciences.

Cisterna, D., Liu, L., Chen, S., van Rijn, P., Graf, A., Rehmat, A. P., Housh, K., & Hmelo-Silver, C. E. (2020). Students' Ideas about the NGSS Crosscutting Concept of System and System Models: Evidence From Parallel Assessments. In *The Interdisciplinarity of the Learning Sciences, 14th International Conference of the Learning Sciences (ICLS) 2020*, Volume 2 (pp. 829-830).

DeLyser, L. A., & Vargas-Vite, E. (2021, March). AlignCS: Evaluating the Landscape of K-12 CS Framework Aligned Curricula. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (pp. 1334-1334).

Duncan, R. G., & Hmelo-Silver, C. E. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal of Research in Science Teaching, 4*, 606-609.

Duncan, R. G., & Rivet, A. E. (2018). Learning progressions. In *International handbook of the learning sciences* (pp. 422-432). Routledge.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of research in education*, *32*(1), 268-291.

Duschl, R., Maeng, S., & Sezen, A. (2011). Learning progressions and teaching sequences: A review and analysis. *Studies in Science Education*, *47*(2), 123-182.

Jin, H., Mikeska, J. N., Hokayem, H., & Mavronikolas, E. (2019). Toward coherence in curriculum, instruction, and assessment: A review of learning progression literature. *Science Education, 103*(5), 1206-1234.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The journal of the learning sciences*, *4*(1), 39-103.

Judd, S. (2020, February). All Means All: Bringing Project-based, Approachable AI Curriculum to More High School Students through AI4ALL Open Learning. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (pp. 1409-1409).

Kaldaras, L., Akaeze, H., & Krajcik, J. (2021). Developing and validating Next Generation Science Standards-aligned learning progression to track three-dimensional learning of electrical interactions in high school physical science. *Journal of Research in Science Teaching, 58*(4), 589-618. Mohan, L., Chen, J., & Anderson, C. W. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *Journal of Research in Science Teaching*, 46, 675-698.

Payne, B. H. 2019. AI+Ethics Curriculum for Middle School. Cambridge, MA: MIT Media Lab.www.media.mit.edu/ projects/ai-ethics-for-middle-school/overview.

Reiser, B.J., Krajcik, J., Moje, E.B., & Marx, R.W. (2003). Design strategies for developing science instructional materials. Philadelphia, PA: Paper presented at the Annual Meeting of the *National Association for Research in Science Teaching*.

Shea, N. A., & Duncan, R. G. (2013). From theory to data: The process of refining learning progressions. *Journal of the Learning Sciences, 22*(1), 7-32.

Songer, N. B., Kelcey, B., & Gotwals, A. W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching, 46*, 610-631.

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., ... & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46*, 632-654.

Touretzky, D., Martin, F., Seehorn, D., Breazeal, C., & Posner, T. (2019, February). Special session: AI for K 12 guidelines initiative. *In Proceedings of the 50th ACM technical symposium on computer science education (*p. 492-493).

Wong, G. K., Ma, X., Dillenbourg, P., & Huan, J. (2020). Broadening artificial intelligence education in K-12: where to start? *ACM Inroads, 11*, 20–29.

Wilson, M., & Scalise, K. (2006). Assessment to improve learning in higher education: The BEAR Assessment System. *Higher education, 52*(4), 635-663.

Acknowledgements
This research was supported by NSF Grants DRL-1934128 and DRL-1934153.