A Review of Assessments in K-12 AI Literacy Curricula

Randi Williams

MIT Media Lab
Cambridge, MA, USA
randiw12@mit.edu

Abstract

Today, Artificial Intelligence (AI) plays a role in many K-12 students' lives within the United States and around the world. Accordingly, there has been an increase in research about K-12 AI literacy curricula to prepare the next generation to participate in the future. However, there is little alignment from researchers on what to teach students or how to evaluate that learning. We examined 64 peer-reviewed articles on K-12 AI literacy curricula to catalog their approach to assessing students' AI knowledge and perspectives. We organized assessments by their type (formative or summative; open, fixed, activity-based) and the kind of content they assess (concept, practice, perspective). We found many examples of summative assessments of AI concepts, construction of AI, and psychological beliefs about AI. Few assessments were formative, activity-based, assessed students' analysis of AI, assessed students' AI communication skills, or assessed their critical understanding of AI. This work sheds light on which approaches to use in assessment and what assessment tools are missing.

1 Introduction

As artificial intelligence (AI) has become increasingly integrated into everyday life, researchers, policymakers, and educators have called for the general public to become more informed about AI. Various advocates of AI education have envisioned the kinds of knowledge and skills they believe people need to be well-informed citizens in an AI-powered world. As synthesized by Ng et al. (2021), different definitions of AI literacy found in previous works revolve around people understanding, knowledgeably using, creating, and evaluating AI systems by developing both technical and ethical knowledge. Along with these calls, there has been a surge in AI education research in formal and informal educational settings. Existing literature reviews on AI education, particularly for K-12 students and the general public, have examined or defined goals for AI education, explored design approaches, and proposed future directions for the field. However, existing reviews have yet to explore the effectiveness of different approaches.

This paper explores AI literacy curricula, specifically those designed for K-12 students, and how they evaluate their work. Specifically, we explore: *how do we best evaluate K-12 students' grasp of AI concepts, practices, and perspectives following an educational intervention?* Answering this research question is an important first step to future meta-reviews that could compare different pedagogical approaches by their chosen evaluation metrics. In the first part of this paper, we review published K-12 AI literacy curricula to propose a framework for taxonomizing AI literacy as AI concepts, practices, and perspectives. Then, we use this framework to summarize evaluation practices in the published articles. Finally, we propose approaches for designing K-12 constructionist AI literacy assessments.

2 Related Literature Reviews

As more research on K-12 AI literacy has been produced over the past few years, many researchers have published review articles to summarize work in the field. The goals of these articles are often quite different. Long and Magerko (2020), Ng et al. (2021), Ojeda-Bazaran et al. (2021), and Zhou et al. (2020) are broadest in scope, summarizing every K-12 AI article they found. Sanusi et al. (2021), Von Wangenheim et al. (2020), and Marques et al. (2020) more narrowly focus on K-12 machine learning curricula while Olari and Romeike (2021) focus on data science. Su and Yang (2022) looks at the use of AI in early education, including AI literacy curricula for children aged three to eight years old. Of these, Ng et al. (2021) contribute a particularly comprehensive review of K-12 AI literacy. First, they conceptualize a definition of AI literacy based on previous work. They describe AI literacy as understanding and using AI applications, applying AI knowledge to different scenarios, evaluating and creating with AI applications, and considering societal and ethical issues related to AI. Furthermore, they unpack pedagogical approaches, assessment approaches, and ethical issues related to the design of K-12 AI curricula.

Besides Ng et al.'s review, Marques et al. (2020) are the only others who review evaluation and assessment approaches in K-12 literacy. Both reviews describe the variety of quantitative and qualitative measures that researchers have used to gather evidence of their work's impact. Although both of these reviews contain useful information about how K-12 AI literacy is assessed, assessment is not the focus. The contribution of this work is that it centers on assessment. As many other reviews have pointed out, K-12 AI literacy work is very cross-disciplinary and there are divergent methods and terms used (Long and Magerko, 2020; Ng et al., 2021; Tedre et al., 2021; Su and Yang, 2022). By centering assessment, we hope to create a common language around evaluation that will make it easier to compare K-12 AI literacy research in the future.

3 Methods

3.1 Search and selection process

We conducted a literature review to summarize assessment approaches in K-12 AI education. We collected published literature in AI education using keyword searches followed by the "snowballing" method from Wohlin (2014). The snowballing method starts by identifying a start set of papers from relevant research communities and then discovering other relevant papers by looking at their references and citations. Our starting set included the nine literature review articles mentioned in the *Related Literature Reviews* section plus Touretzky et al.'s milestone 2019 paper on K-12 AI literacy. We also included papers on K-12 AI literacy from the 2019-2021 proceedings

Table 1: Inclusion and exclusion criteria for study selection

Inclusion criteria	Exclusion criteria
Is a peer-reviewed conference or journal paper	Is a thesis, book, website, or other non-peer-reviewed work
Specifies primary, middle, and/or high school students in the target population	Does not specify the age of the target population, combines results of participants of different ages, or only includes older learners
Describes an evaluative user study with K-12 participants	Is a proposal, theory, system description, or other non-experimental work
Provides the results of the evaluation	

of the Educational Advances in AI (EAAI) conference and the ACM's Special Interest Group on Computer Science Education (SIGCSE). We excluded EAAI and SIGCSE papers that did not focus on AI literacy or did not include students in the K-12 age range; this added 12 more papers to our starting set.

As we collected papers that cited and were cited by our starting set, we narrowed our scope to papers published during or before 2021. One reviewer downloaded and briefly reviewed the abstracts and results of these papers to determine the age range of the participants, whether a user study was conducted, and whether the results of the user study were published in the article. We adopted a set of inclusion and exclusion criteria, shown in Table 1, to standardize our study selection process.

3.2 Data coding and analysis process

First, we reviewed the selected studies and recorded descriptive information about the study: the country the authors came from, the year of publication, the age group studied, and the setting the study was conducted in. Next, we analyzed the AI content and evaluation methods employed in the studies. In this initial review, only one researcher conducted the analysis.

The researcher used the grounded theory method of analyzing the AI content delivered in the studies. For AI content, the researcher's goal was to align AI content with the concepts, practices, and perspectives (CPPs) framework proposed by Brennan and Resnick (2012). The researcher summarized the learning objectives found in the selected papers, then categorized them by whether they conveyed a concept, practice, or perspective. For example, "recognizing a system as using AI" was categorized as a practice while "awareness of the role of AI in personal life" was categorized as a perspective. For the assessments, the researcher used a similar process, clustering evaluation methods into broad categories such as "knowledge examination"

Table 2: Reviewed articles by year of publication

Year	N	Studies
2008	1	Bigham et al.
2012	1	Rosen et al.
2014	1	Benotti et al.
2016	3	Kandlhofer et al.; Burgsteiner et al.; Vachovsky et al.
2017	1	Srikant and Aggarwal
2018	4	Hitron et al.; Sakulkueakulsuk et al.; Kahn et al.; Ureta and Rivera
2019	11	Druga et al.; Hitron et al.; Williams et al.; Williams et al. Estevez et al.; Mariescu-Istodor and Jormanainen; Mobasher et al.; Tang et al.; Zhang et al.; Zimmermann-Niefield et al.
2020	13	Alturayeif et al.; Bilstrup et al.; DiPaola et al.; Lin et al. Norouzi et al.; Sabuncuoglu; Schaper et al.; Shamir and Levin Skinner et al.; Van Brummelen et al.; Vartiainen et al. Vartiainen et al.; Wan et al.
2021	29	Ali et al.; Ali et al.; Ali et al.; Choi and Park; Druga and Ko Forsyth et al.; Henry et al.; Jordan et al.; Kaspersen et al.; Lin et al.; Kaspersen et al.; Kim et al.; Lee et al.; Long et al.; Lyu et al. Melsión et al.; Olari et al.; Park et al.; Rodríguez-García et al. Shamir and Levin; Aki Tamashiro et al.; Tseng et al. Van Brummelen et al.; Vartiainen et al.; Voulgari et al. Williams; Yoder et al.; Zhang et al.; Zhu and Van Brummelen
Total	64	

or "discussion and interviews." These clusters were used to create a common set of codes that we could use to group studies. From there, the researcher identified common themes within the codes and used these to create subcategories within the CPP framework.

We acknowledge that this is a preliminary study where the coding results are highly likely to be unreliable or incomplete since only one researcher did the review work. In the future, we will validate this review procedure by having multiple researchers independently use the procedure and then calculate the inter-rater reliability of their codes. The results of the search process are publicly available for future researchers to build upon: `https://github.com/randi-c-dubs/k12-ai-ed`.

4 Results

4.1 Reviewed articles

Before excluding studies that did not include user studies, we collected 103 papers on AI literacy curricula, proposals, and educational tools for K-12 students. We note here that our search process yielded many more papers than previous literature

Table 3: Attributes of reviewed articles

Factor	Category	N	Percent
Age group	Primary	11	17%
	Middle	11	17%
	Secondary	22	34%
	Pre-K and Primary	2	3%
	Primary and Middle	6	9%
	Middle and Secondary	4	6%
	Primary to Secondary	6	9%
	Primary to Adult	2	3%
Setting	Informal	40	63%
	Formal	13	20%
	Laboratory	11	17%
Country of first author	USA	34	53%
	Denmark	4	6%
	Finland	4	6%
	Israel	4	6%
	Spain	3	5%
	Austria	2	3%
	Korea	2	3%
	Other	11	17%

reviews and certainly more papers than we expected. (Ng et al., 2021) had the largest review, with 67 K-12 AI literacy research articles. We believe that our collection of more than 100 papers is significant because research about K-12 AI literacy often claims that there is not much work to be found in the field. However, the results of our search suggest that there is more work out there than researchers are cognizant of. Here, we see an opportunity for further research about which papers receive more papers and citations while others are left out.

Of the papers that we identified for potential review, 78 conducted a user study with K-12 students. Of these, 64 presented the results of the study in their article. Table 2 contains all of the studies we included in our review categorized by their year of publication. This table shows that there has been a sharp increase in published AI literacy studies over the past five years.

4.2 AI concepts, practices, and perspectives

We also categorized previous work by the kinds of AI knowledge converted in the intervention. Brennan and Resnick (2012) published a taxonomy of computational thinking knowledge called the concepts, practices, and perspectives (CPPs) framework. We adapted this framework to outline concepts, practices, and perspectives for

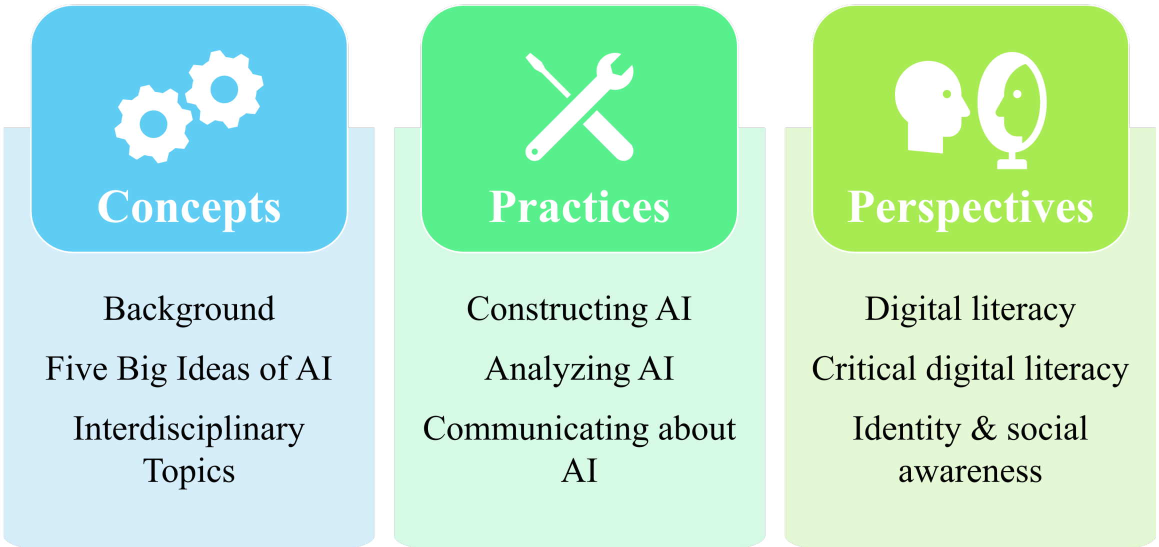

Figure 1: A summary of subcategories in the concepts, practices, and perspectives of our AI literacy CPP framework.

AI education. Rather than trying to capture all AI knowledge in a comprehensive list, our goal was to create a broad set of categories, shown in Figure 1, for sorting interventions.

Concepts. AI concepts are the set of knowledge that AI researchers and practitioners engage with as they do work. Since there is no agreed-upon taxonomy of AI concepts, we created subcategories for concepts along the lines of AI background knowledge, interdisciplinary knowledge, and the Five Big Ideas of AI. Background knowledge includes definitions and examples of AI, the historical development of AI, and comparisons between machine and biological intelligence. The Five Big Ideas of AI, developed by Touretzky et al. (2019) describe five sectors of AI knowledge: machine perception, representation and reasoning, machine learning, Human-AI interaction, and societal impact. AI concepts related to the Five Big Ideas were most commonly taught across all AI concepts, with machine learning as the best-represented set of concepts we found in previous work. Finally, interdisciplinary topics cover all other knowledge that researchers taught alongside AI (e.g., science) to contextualize learning and draw connections between AI and students' core curricula.

Practices. AI practices include the skills and methods that AI researchers employ in their work. Based on Brennan and Resnick's framework for computational thinking practices, we derived three subcategories for AI practices: construction, analysis, and communication about AI. Practices related to constructing AI were more common than any other kind of practice. This subcategory included using the design thinking process, using the scientific method, defining AI problems, prototyping, data set curation, creating models, evaluating models, adapting to feedback, programming, and user interface (UI/UX) design. The AI analysis subcategory included data analysis and interpretation, recognizing and deciphering AI systems, evaluating the presence of bias in AI systems, predicting the impact of systems, and identifying stakeholders and stakeholder values. The final and least commonly identified AI practices involved communicating about AI. This included scientific communication skills and activism skills, like advocacy, related to AI.

Perspectives. Similar to Brennan and Resnick, we define AI perspectives as the beliefs and evolving understandings that students develop as they learn about AI. The first "Perspectives" subcategory is digital literacy, or students' developing awareness of AI and its impact on society. This includes students recognizing how AI is relevant to their personal lives, future careers, culture, and society. The next subcategory is critical digital literacy, or students' developing consciousness of AI's strengths and limitations. This includes students recognizing humans' role in designing AI systems, seeing that AI systems can have both positive and negative impacts on society, and seeing that people can redesign AI systems with new goals. Finally, identity and social awareness cover students' recognizing their individual potential to influence AI, feeling that they are capable, feeling that they are part of an AI community, and being aware of their personal strengths and weaknesses.

4.3 Existing AI literacy assessments

Kong (2019) used the computational thinking CPP framework to explore researchers' approaches to assessment. In this section, we use a similar approach to review how researchers have assessed K-12 students' understanding of AI content. First, we identify the summative and formative approaches researchers have used for AI concepts and practices. Then we look at the myriad of approaches used to evaluate students' AI perspectives.

4.3.1 Summative Assessments of AI Concepts and Practices

Lancaster et al. (2019) described summative assessments as measures used to evaluate students' achievement at the end of an educational intervention. In our review, we found that K-12 AI education researchers used assessments including multiple-choice questions, open response questions, interviews, discussion, and project-based assessments to evaluate students' understanding of AI concepts and practices.

Multiple-choice questions. Multiple-choice questions are the least labor and time-intensive assessments for researchers and educators to implement. However, they are also the least flexible way to assess understanding and provide little opportunity for students to continue learning as a part of the assessment. AI education researchers have begun developing concept inventories for the topics they cover in their interventions. For example, Rodríguez-García et al. (2021) developed an assessment that covered general knowledge of AI and specific knowledge about supervised machine learning systems. Lee et al. (2021) used a concept inventory that covered a broad range of AI topics including basic AI knowledge, machine learning, decision trees, and generative AI concepts. Both examples used concept inventories as diagnostic tools to measure students' knowledge before any lessons began.

Open responses. Compared to multiple-choice questions, assessments with open response questions take longer for students to complete and are more labor-intensive

Table 4: AI Concept and Practice Assessments in Reviewed Articles

Type of Assessment	Relevant Articles
Multiple-choice: close-ended questions with a finite number of correct answers, evaluated quantitatively	(Kahn et al., 2018; Kandlhofer et al., 2016; Kim et al., 2021; Lee et al., 2021; Lyu et al., 2021; Melsión et al., 2021; Rodríguez-García et al., 2021; Shamir and Levin, 2020, 2021; Von Wangenheim et al., 2020; Williams et al., 2019b; Yoder et al., 2021)
Open response: open-ended short or extended response questions, evaluated qualitatively	(DiPaola et al., 2020; Estevez et al., 2019; Henry et al., 2021; Hitron et al., 2018, 2019; Kahn et al., 2018; Kandlhofer et al., 2016; Lin et al., 2020; Rodríguez-García et al., 2021; Rosen et al., 2011; Shamir and Levin, 2020, 2021; Tang et al., 2019; Tseng et al., 2021; Van Brummelen et al., 2020; Vartiainen et al., 2020a,b; Zhang et al., 2019, 2021; Zhu and Van Brummelen, 2021)
Discussion and interview: open-ended, conversations about AI concepts or students' projects, evaluated qualitatively	(Kandlhofer et al., 2016; Kaspersen et al., 2021a; Park et al., 2021; Shamir and Levin, 2021; Aki Tamashiro et al., 2021; Tseng et al., 2021; Vartiainen et al., 2020b, 2021; Von Wangenheim et al., 2020; Zimmermann-Niefield et al., 2019)
Project-based assessment: analysis of student-produced artifacts, evaluated quantitatively or qualitatively	(Alturayeif et al., 2020; Bigham et al., 2008; DiPaola et al., 2020; Jordan et al., 2021; Sakulkueakulsuk et al., 2018; Srikant and Aggarwal, 2017; Tseng et al., 2021; Ureta and Rivera, 2018; Van Brummelen et al., 2020; Vartiainen et al., 2020b, 2021; Von Wangenheim et al., 2020; Williams, 2021; Zhu and Van Brummelen, 2021)
Activity-based assessment: analysis of student behavior or task completion, evaluated quantitatively or qualitatively	(Ali et al., 2021c,b; Benotti et al., 2014; Druga et al., 2019; Druga and Ko, 2021; Forsyth et al., 2021; Kandlhofer et al., 2016; Lyu et al., 2021; Park et al., 2021; Aki Tamashiro et al., 2021; Tseng et al., 2021; Wan et al., 2020; Yoder et al., 2021)

to grade. They are also less objective since graders must make judgments about what knowledge a student is demonstrating in their answer. Even so, open response questions give researchers and educators a richer understanding of how much of the content students understand.

The most common kind of open response questions were definition questions where students gave a general definition of AI or explained a particular mechanism. Rodríguez-García et al. (2021) and Zhu and Van Brummelen (2021) both used students' open-ended responses to AI definition questions to create word clouds that showed how students' mental models of AI developed by the end of their educational modules. Kandlhofer et al. (2016) and Vartiainen et al. (2020b) gave open-ended assessments to younger (5-10-year-old) students, making them more accessible by having students draw to express their understanding.

Word problems are another kind of open response question, they allow students to demonstrate their understanding of concepts in a concrete context. For example, Hitron et al. (2019) and Tseng et al. (2021) asked students to explain AI mechanisms using scenarios they constructed. The scenarios provided some information students could use to make an educated guess about how a mechanism might work, even before they had gone through the AI lessons. Programming exercises, such as the ones used by Zhang et al. (2019) to have students demonstrate their mastery of logic programming, are another example of word problems.

Finally, a few research studies used extended response questions to evaluate AI understanding. For example, DiPaola et al. (2020) included assessment questions with case studies to have students demonstrate their grasp of stakeholder analysis. Hitron et al. (2019) and Vartiainen et al. (2020a) used essay questions, asking students how they might apply what they had learned to a problem in their own lives.

Discussion and interviews. Assessments using discussion and interviews are very labor-intensive, but create more opportunities to give students feedback during an assessment. Kandlhofer et al. (2016) started lessons with a discussion on students' prior knowledge, using their past experiences as a foundation for the AI lessons. At the end of the lessons, researchers interviewed students about their experiences and which topics stood out. In Vartiainen et al. (2021) and Vartiainen et al. (2020b), researchers facilitated small group discussions amongst students before and after learning about machine learning to assess their understanding. The discussions allowed students to articulate what they had learned and expand their thinking as they conversed with their peers. Finally, Von Wangenheim et al. (2020) used debates to get students to share varying perspectives about ethical issues in AI.

Project-based assessments. Researchers used cumulative final projects to assess the extent to which students could practically apply their knowledge. To evaluate students' projects, researchers used both qualitative and quantitative measures. Qualitative measures were more exploratory, analyzing the kinds of projects students created and what concepts they engaged with (Bigham et al., 2008; DiPaola et al., 2020; Jordan et al., 2021; Van Brummelen et al., 2020; Vartiainen et al., 2020b, 2021). Quantitive measures used rubrics or statistical measures to grade and compare students' projects (Alturayeif et al., 2020; Srikant and Aggarwal, 2017; Tseng et al., 2021; Sakulkueakulsuk et al., 2018; Williams, 2021; Ureta and Rivera, 2018; Von Wangenheim et al., 2020). For example, Alturayeif et al. (2020) gave students a close-ended project to complete and evaluated students' learning based on how many project tasks they completed and how long they took to complete them. Sakulkueakulsuk et al. (2018) used competition between students' projects based on performance metrics to rate students' work. Finally, Shamir and Levin (2021), Vartiainen et al. (2020b), and Aki Tamashiro et al. (2021) did artifact interviews where students' work was a starting point for discussing AI concepts in more depth with students.

4.3.2 Formative Assessments of AI Concepts and Practices

Lancaster et al. (2019) described formative assessments as gathering information about students' understanding as they progress through learning activities. In K-12 AI education, researchers tracked students' progress as they worked on projects and activities to understand the progression of their learning.

Formative project-based assessments. Formative project-based assessments considered students' processes as they worked on their projects. Vartiainen et al. (2021) documented a broad scope of students' creation processes, from their brainstormed ideas through implementation. They used this information to gain insight into how students reflected on and refined their projects as their understanding of machine learning grew. Tseng et al. (2021) evaluated students' conceptual understanding of machine learning by having them think out loud as they worked on their projects. In their method, researchers combined observation with interviews, leveraging the context of the projects students were working on to ask probing conceptual questions. Researchers' questions helped students both reinforce their understanding of machine learning concepts and clarify any misunderstandings.

Activity-based assessments. Researchers used formative activity-based assessments to better understand students' developing mental models of AI. Learning activities, unlike student projects, are smaller bit of work where an educator has more control over which learning objectives students engage with.

Although many papers described analytical methods they used to measure students' learning while they did activities, only Ali et al. (2021c) explicitly identified embedded assessments as part of their approach. Ali et al. described gathering evidence of students' learning using online forms and system logs as students played games and explored AI demos. They later analyzed this data to demonstrate what students had learned about GANs with their activities.

Similar to embedded assessment, other researchers recorded videos, transcribed, and gathered field notes as students completed the lessons to gather evidence of students' learning as they completed activities Druga et al. (2019); Druga and Ko (2021); Kandlhofer et al. (2016); Park et al. (2021); Tseng et al. (2021); Ali et al. (2021b); Forsyth et al. (2021); Aki Tamashiro et al. (2021); Benotti et al. (2014); Lyu et al. (2021); Wan et al. (2020); Yoder et al. (2021). Park et al. (2021) analyzed video recordings of students learning about AI agents to identify moments of revelation and obstacles to learning that they could use to improve their platform. Wan et al. (2020) identified four inquiry behaviors (question asking, creating arguments from evidence, suggesting, and sharing) and reviewed recordings of students participating in their activities to identify when students engaged in these behaviors.

4.3.3 Assessing Perspectives

Finally, researchers used surveys, open response questions, interviews, and assessments embedded in activities to understand students' changing values, beliefs about AI, and AI identities.

Surveys. Six (6) papers evaluated students' perceptions of artificially intelligent devices Williams et al. (2019a); Druga et al. (2019); Druga and Ko (2021); Lin et al. (2020); Van Brummelen et al. (2021); Zhu and Van Brummelen (2021). In these examples, researchers used examples of AI agents, such as toy robots and chatbots, as the basis for their investigations of perceptions. Researchers used multiple-choice Likert questions to explore students' mental models of the devices. A notable example

Table 5: AI Perspective Assessments in Reviewed Articles

Type of Assessment	Relevant Articles
Perception of AI surveys: subjective questionnaires about students' beliefs about AI artifacts, evaluated quantitatively	(Druga et al., 2019; Druga and Ko, 2021) (Lin et al., 2020; Van Brummelen et al., 2021) (Williams et al., 2019a) (Zhu and Van Brummelen, 2021)
Attitudes toward AI surveys: subjective questionnaires about students' motivation to learn AI, evaluated quantitatively	(Estevez et al., 2019; Kandlhofer et al., 2016) (Kim et al., 2021; Lee et al., 2021; Lin et al., 2020) (Mariescu-Istodor and Jormanainen, 2019) (Mobasher et al., 2019; Olari and Romeike, 2021; Rosen et al., 2011; Sakulkueakulsuk et al., 2018; Shamir and Levin, 2020; Vachovsky et al., 2016; Yoder et al., 2021; Zhang et al., 2019) Zhu and Van Brummelen (2021)
Discussion and interview: open-ended conversations about students' beliefs, evaluated qualitatively	(Druga et al., 2019; Lee et al., 2021) (Yoder et al., 2021)
Activity-based assessment: analysis of task completion, evaluated quantitatively and qualitatively	(Ali et al., 2021c; Lee et al., 2021)

is Williams et al. (2019a) who worked with very young students and connected their perceptions of AI with their developmental level, examining how age and students' theory of mind led to differences in students' attributions of intelligence to the devices.

Changes in students' attitudes toward AI and beliefs about their individual ability to participate in AI were of high interest to researchers. Zhang et al. (2019) and Sakulkueakulsuk et al. (2018) asked students about the extent to which they felt that learning AI in a relevant interdisciplinary context was important and beneficial to students. Estevez et al. (2019) and Yoder et al. (2021) used pre and post-Likert questions to explore how important students thought it was for people to learn AI given its role in society and important ethical issues. Other researchers used multiple-choice self-evaluation questions to gauge students' interest, feelings of competence, and self-efficacy with regard to AI. Notably, Mariescu-Istodor and Jormanainen (2019) and Kim et al. (2021) used models of student motivation and interest to assess multiple motivation factors such as relevance and perceived ability. Two other works specifically asked questions about participation in the AI community - Estevez et al. (2019) asked if students' assumptions about who could be successful in AI had changed and Vachovsky et al. (2016) asked if students felt they had role models and community in the field.

Discussion and interviews. Researchers' survey questions often went along with discussions and interview questions or where students could share more of their thoughts. Druga et al. (2019) gave students a questionnaire about their perceptions of different AI agents. After they completed the questionnaire, students would discuss their answers with their peers, lending further insight into how students thought about the technology. Lee et al. (2021) and Yoder et al. (2021) observed that during the post-activity discussions and interviews, students connected what they had done in the lessons to their developing beliefs about AI's impact on society.

Activity-based assessments. Finally, researchers used activities that gave students opportunities to explore and describe their beliefs about AI. Ali et al. (2021a) asked students to turn their understanding of AI into action in an activity where students proposed policies to regulate the use of AI. In their analysis, they explored how students' perspectives about AI's impact on society drove which policies they advocated for. Lee et al. (2021) included an activity where students connected potential future careers to AI. This activity showed how students thought about future AI applications and their role in them.

5 Discussion: Future opportunities for constructionist AI curricula

At the 2011 International Confederation of Principals, Andy Hargreaves remarked that you should "measure what you value instead of valuing only what you can measure." Along these lines, in constructionist AI literacy education we should value and measure what students learn as they engage in the process of making artifacts. In this section we build off of prior published work in AI literacy and computational thinking to synthesize a comprehensive approach for the assessment of AI concepts, practices, and perspectives

5.1 Assessing AI concepts

In assessing AI concepts, researchers should aim to measure students' grasp of background information in AI, the Five Big Ideas of AI, and interdisciplinary material covered in their curricula. Particularly for constructionist curricula, assessments should be grounded in the activities and projects that students do. Suitable assessments include project-based concept inventories, concept discussions, artifact interviews, and activity-based assessments.

Project-based concept inventories. Project-based concept inventories look for evidence of students' learning in the artifacts that students create. AI concept inventories, such as those from Rodríguez-García et al. (2021) and Lee et al. (2021), contain both general and specific AI knowledge students should understand. They include foundational ideas that students should understand, such as definitions, and check whether students hold common misconceptions. Researchers can use concept inventories to create evaluation tools such as self-assessments and project rubrics.

Many of the articles we reviewed used self-assessment tools as part of their evaluation. For example the 12-point AI competency self-assessment produced by Lassnig (2018) that included items such as “I can formalize a search problem.” Students can use self-assessment tools in a formative manner to track their learning as they progress through a curriculum. Kim et al. (2020) created embedded assessment tools such as “Sparkle Sleuth” where students and teachers used tangible notes to create portfolios of their learning. Self-assessments can help keep learning objectives at the front of students’ minds.

Project rubrics can be used to benchmark students’ application of concepts. Lao et al. (2019) allowed students to choose any project topic they wanted, then they analyzed projects by which AI concepts (e.g. multi-layer networks, transfer learning) students applied. By connecting project topics to concepts covered in a curriculum, educators can help students turn information into personally meaningful projects.

Concept discussions and artifact interviews. Concept discussions and artifact interviews allow students to reflect on and participate in constructive dialogue about their work Vartiainen et al. (2020b) and Vartiainen et al. (2021) hosted small group discussions to evaluate students’ learning. As students shared their experiences and knowledge of machine learning, they also learned from one another’s experiences and understandings. Von Wangenheim et al. (2020) used discussions to have students push one another in their understandings of ethics. Researchers can structure discussions around concepts they want to assess and take notes to evaluate learners’ understanding. Discussions are beneficial to students because they continue to learn as they get feedback on their ideas from peers.

Tseng et al. (2021) used artifact interviews to formatively assess their young participants by having them think aloud as they worked on their projects. Researchers asked students questions about concepts as students encountered them, creating a personalized assessment where students reflected on their knowledge as they applied to it. Shamir and Levin (2021) used summative artifact interviews, asking students to trace what they learned through the narrative of their project creation. Researchers can use structured artifact interviews to gather evidence about what concepts students understand and whether their knowledge has any gaps Brennan and Resnick (2012).

Activity-based assessment. Activity-based assessments like embedded assessment and classroom observation allow researchers to measure students’ learning as it happens. Activity-embedded assessment questions could include multiple-choice and open response problems to gather information about students’ grasp of ideas. This is likely the best approach for evaluating how well students understand very abstract ideas or background AI information. Ali et al. (2021b) used embedded assessment to evaluate students’ understanding of generators and discriminators in GANs while students were exploring different real-world examples of GANs. After this assessment, educators can use students’ answers as the basis for a class discussion and review, as described by Ali et al..

Classroom observation also allows researchers to gather information about what students are learning. Park et al. (2021) systematically recorded notes about students' learning, misconceptions, and struggles as they programmed. Researchers with a clear theory of learning can integrate embedded assessments and classroom observation to track concept mastery as it happens.

5.2 Evaluating AI practices

To measure students' ability to construct, analyze, and communicate about AI systems, researchers should use assessments that evaluate the process and final products of constructionist curricula. Project-based skills inventories, artifact interviews, project studios, project reports, and activity-based assessments are suitable assessments for evaluating the three subcategories of AI practices.

Project-based skill inventories. We are unaware of any published work that describes an AI skill inventory, though this kind of work is currently underway by groups such as the AI4K12 working group, `https://ai4k12.org/gradeband-progression-charts`. An AI skill inventory should include items related to constructing AI, such as "evaluating model," and analyzing AI, such as "considering the implications of an AI system." Researchers can use skill inventories to create self-assessment tools, observation schema, and project rubrics.

Students can complete skill self-assessments to reflect on how they used different skills in their projects. Kim et al. (2020) described Maker Moments, a formative self-assessment tool where students and educators note their use of key skills. As students are engaged in projects, they should see developments in their technical, ethical, and design thinking skills.

Project rubrics can be used as a quantitative benchmark of students' learning. Brennan et al. (2020) described how different educators used project rubrics to set nonrestrictive guidelines for the kinds of skills they wanted students to demonstrate in their projects. One educator accomplished this by having multiple levels of learning objectives, "mild, medium, and spicy" that students could choose from. Technical performance metrics can also be a helpful way to rate students' grasp of different skills. Sakulkueakulsuk et al. (2018) had students complete very similar projects and hosted a competition between teams to judge students' work. Researchers should create project rubrics that enable them to measure the extent to which students demonstrate technical and ethical skills in their projects.

Project studios and artifact interviews. Project studios and artifact interviews allow students to participate in constructive dialogues about their projects. In project studios, students share their work with peers to get feedback and support. Zimmermann-Niefield et al. (2019) had students discuss their projects in small groups. During these discussions, students conferred with one another on the best techniques reflected on the challenges they faced, and discovered new approaches from others' experiences. Brennan (2015) emphasized the usefulness of these discussions as part of forming a "community of practice" where individuals could increase their capacity for creativity.

In artifact interviews, students discuss their projects with researchers. Brennan and Resnick (2012) had students reflect on their process of creation and speak about how they used programming practices like debugging and iteration. Researchers can also use artifact interviews to evaluate students' ability to transfer their knowledge to new domains. In their artifact interviews, Shamir and Levin (2021) asked students to describe how they might extend their projects. Students talked through their understanding of AI practices that beyond what showed up in their projects. Researchers should structure artifact interviews such that students can reflect on skills learned in class and connect to ideas outside of class Turbak and Berg (2002).

Project reports and presentations. Assessments that include project reports and presentations let researchers evaluate students' ability to communicate about AI. Von Wangenheim et al. (2020) had students write an evaluation report about the performance of their models, putting their results in the context of the problem they were trying to solve. Project reports can also be effective for helping students communicate about ethical issues. For example, Furey and Martin (2018) had students include a summary of the ethical implications of their projects in final reports. Project reports can also be formative assessment tools: students can document their code or complete design journals as they work (Brennan and Resnick, 2012). Written reports allow students to see how AI skills relate to the broader context of projects they are interested in(Kong, 2019).

Many AI curricula included project presentations at the end of their curricula Zhang et al. (2019); Mobasher et al. (2019), but none explicitly analyzed students' presentations in their results. The benefits of presentations over artifact interviews or reports are that students have to structure their thinking and present it to a broad audience that can give them feedback. Researchers can better take advantage of these kinds of presentations as assessments of practices by using skill inventories to create a rubric. Students could use the rubric to structure their presentations and audience members (e.g., educators, peers, parents, and other community members) can use it to provide feedback.

Activity-based assessment. Activity-based assessments like programming exercises and design problems are tools researchers can use to more directly assess students' grasp of specific AI practices. To evaluate practices related to constructing artifacts, Zhang et al. (2019) gave students logic programming problems to see if they could use the techniques they learned in class on new. For practices related to analyzing artifacts, DiPaola et al. (2020) created case studies to evaluate students' ability to use ethical thinking practices like identifying stakeholders and their values. Activity-based assessments can also be used to evaluate students' ability to transfer skills to new domains. Hitron et al. (2019) and Tseng et al. (2021) asked students word problems about how they might apply their knowledge of AI to an issue that came up in their own lives. Design problems are powerful because they allow researchers to probe students' knowledge of particular practices in context, rather than from memory Brennan and Resnick (2012).

5.3 Surveying AI perspectives

Finally, measuring students' developing digital literacy, critical digital literacy, identity, and social awareness requires assessments that uncover the connections students make to the material they interact with. Assessments like surveys, discussions, debates, and activity-embedded assessments can help evaluate perspectives.

Surveys. Constructionist assessments should recognize that students make new ideas by building on their previous ones. This means that assessments should paint a picture of how students' ideas develop over time, not just where they might be at a particular moment Brennan and Resnick (2012). Kim et al. (2021) developed a student model of interest and motivation to understand "why" students want to learn about AI and, consequently, what curricula should teach. Surveys such as the one developed by Kim et al. delivered at multiple points throughout a curriculum can reveal the bigger picture about how students' digital literacy, identity, and social awareness are developing. Understanding this picture is key to developing curricula that meet students' needs, especially when students are from demographic groups that have been historically excluded and underrepresented in tech (Vachovsky et al., 2016; Lee et al., 2021).

Discussions and debates. Discussions and debates can help students clarify their beliefs about AI and make stronger arguments to support their beliefs. Discussions provide an opportunity for students to give and receive feedback on different perspectives about AI, particularly related to digital literacy and critical digital literacy. Druga et al. (2019) had students discuss their perceptions of AI in small groups as they completed a survey about their beliefs. As students discussed their perceptions, researcher observed that many of their beliefs were informed by media, their parents, and peers. Using discussions helped researchers identify not only what students' beliefs were, but where they came from. Von Wangenheim et al. (2020) had students debate different ethical issues in AI. These debates revealed students' beliefs about AI, particularly their beliefs around appropriate roles for AI given the limitations of technology. Discussions and debates allow researchers to explore where students' ideas come from while also giving students opportunities to progress in their thinking.

Activity-based assessment. Activity-based assessments of perspectives allow students to develop their understandings and put them into action. Activities around digital literacy can be helpful for both assessing and promoting students' digital literacy. Rather than using a survey to assess students' understanding of AI, Lee et al. (2021) used an activity where students worked in groups to classify different examples of technology as AI or not. These kinds of activities allow students to both show their current understanding while also supporting further learning Brennan and Resnick (2012). Ali et al. (2021a) had students put their opinions about AI into action using an activity where students advocated for different policies for regulating AI. In this way, Ali et al. allowed students to articulate their opinions about AI while also considering actions they might take to become AI activists. Activity-based assessments allow students to express and take action on their ideas.

6 Conclusion

As researchers continue to create and publish new K-12 AI literacy curricula, shared understandings of what students should learn and how researchers measure that learning will become increasingly critical. Based on published articles about K-12 AI literacy curricula, we synthesized a taxonomy of AI concepts, practices, and perspectives (CPPs) based on the computational thinking CPPs presented by Brennan and Resnick (2012) to create a shared language researchers can use to define learning objectives in their curricula. AI CPPs define the kinds of information, skills, attitudes, and beliefs that students should acquire as they engage in the construction of AI. We found that some AI topics, such as supervised machine learning, were emphasized much more than others, such as practices around communicating about AI. We hope that future AI literacy curricula will keep AI CPPs in mind as they consider what topics to emphasize.

Assessment is important for helping students and educators set and meet goals in their learning journeys. The kind of assessment that educators use implicitly defines the curriculum Rowntree (1977). Thus, curriculum designers should take care to use assessments that capture what is most beneficial to students, educators, and researchers. We reviewed different approaches researchers have taken to evaluate students' grasp of AI concepts, practices, and perspectives. Then, we used assessment methods from this review to propose specific assessment approaches that are well-aligned with a constructionist pedagogy. We hope that this proposal is a valuable contribution to curriculum designers looking for comprehensive ways to measure what students have learned.

In completing this work, we also look toward future work that would help strengthen this area of study. First, the field of K-12 AI literacy still needs guidelines about what students should learn. A comprehensive set of learning objectives would be helpful for the development of concept and skill inventories that researchers could use to develop assessment tools. Second, there needs to be more assessments of ethics, ethical thinking practice, communicating about AI, and AI perspectives. Assessments about AI concepts and constructing AI artifacts abound. However, if students are going to become citizens who think critically about technology, we will need better assessment tools for those skills. Work in this direction will help researchers develop more holistically impactful curricula.

References

M. Aki Tamashiro, M. Van Mechelen, M.-M. Schaper, and O. Sejer Iversen, "Introducing teenagers to machine learning through design fiction: An exploratory case study," in *Interaction Design and Children*, 2021, pp. 471–475.

S. Ali, D. DiPaola, and C. Breazeal, "What are gans?: Introducing generative adversarial networks to middle school students," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 17, 2021, pp. 15 472–15 479.

S. Ali, D. DiPaola, I. Lee, J. Hong, and C. Breazeal, "Exploring generative models with middle school students," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–13.

S. Ali, D. DiPaola, I. Lee, V. Sindato, G. Kim, R. Blumofe, and C. Breazeal, "Children as creators, thinkers and citizens in an ai-driven future," *Computers and Education: Artificial Intelligence*, vol. 2, p. 100040, 2021.

N. Alturayeif, N. Alturaief, and Z. Alhathloul, "Deepscratch: scratch programming language extension for deep learning education," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 7, pp. 642–650, 2020.

L. Benotti, M. C. Martínez, and F. Schapachnik, "Engaging high school students using chatbots," in *Proceedings of the 2014 conference on Innovation & technology in computer science education*, 2014, pp. 63–68.

J. P. Bigham, M. B. Aller, J. T. Brudvik, J. O. Leung, L. A. Yazzolino, and R. E. Ladner, "Inspiring blind high school students to pursue computer science with instant messaging chatbots," in *Proceedings of the 39th SIGCSE technical symposium on Computer science education*, 2008, pp. 449–453.

K.-E. K. Bilstrup, M. H. Kaspersen, and M. G. Petersen, "Staging reflections on ethical dilemmas in machine learning: A card-based design workshop for high school students," in *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, 2020, pp. 1211–1222.

K. Brennan, "Beyond technocentrism," *Constructivist Foundations*, vol. 10, no. 3, pp. 290–293, 2015.

K. Brennan and M. Resnick, "New frameworks for studying and assessing the development of computational thinking," in *Proceedings of the 2012 annual meeting of the American educational research association, Vancouver, Canada*, vol. 1, 2012, p. 25.

K. Brennan, P. Haduong, and E. Veno, "Assessing creativity in computing classrooms," *Creative Computing Lab*, 2020.

H. Burgsteiner, M. Kandlhofer, and G. Steinbauer, "Irobot: Teaching the basics of artificial intelligence in high schools," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, 2016.

E. Choi and N. Park, "Demonstration of gamification in education for understanding artificial intelligence principles at elementary school level." *Ilkogretim Online*, vol. 20, no. 3, 2021.

D. DiPaola, B. H. Payne, and C. Breazeal, "Decoding design agendas: an ethical design activity for middle school students," in *Proceedings of the Interaction Design and Children Conference*, 2020, pp. 1–10.

S. Druga and A. J. Ko, "How do children's perceptions of machine intelligence change when training and coding smart programs?" in *Interaction Design and Children*, 2021, pp. 49–61.

S. Druga, S. T. Vu, E. Likhith, and T. Qiu, "Inclusive ai literacy for kids around the world," in *Proceedings of FabLearn 2019*, 2019, pp. 104–111.

J. Estevez, G. Garate, and M. Graña, "Gentle introduction to artificial intelligence for high-school students using scratch," *IEEE access*, vol. 7, pp. 179 027–179 036, 2019.

S. Forsyth, B. Dalton, E. H. Foster, B. Walsh, J. Smilack, and T. Yeh, "Imagine a more ethical ai: Using stories to develop teens' awareness and understanding of artificial intelligence and its societal impacts," 2021.

H. Furey and F. Martin, "Introducing ethical thinking about autonomous vehicles into an ai course," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

J. Henry, A. Hernalesteen, and A.-S. Collard, "Teaching artificial intelligence to k-12 through a role-playing game questioning the intelligence concept," *KI-Künstliche Intelligenz*, vol. 35, no. 2, pp. 171–179, 2021.

T. Hitron, I. Wald, H. Erel, and O. Zuckerman, "Introducing children to machine learning concepts through hands-on experience," in *Proceedings of the 17th ACM conference on interaction design and children*, 2018, pp. 563–568.

T. Hitron, Y. Orlev, I. Wald, A. Shamir, H. Erel, and O. Zuckerman, "Can children understand machine learning concepts? the effect of uncovering black boxes," in *Proceedings of the 2019 CHI conference on human factors in computing systems*, 2019, pp. 1–11.

B. Jordan, N. Devasia, J. Hong, R. Williams, and C. Breazeal, "Poseblocks: A toolkit for creating (and dancing) with ai," in *The 11th Symposium on Educational Advances in Artificial Intelligence*, 2021.

K. M. Kahn, R. Megasari, E. Piantari, and E. Junaeti, "Ai programming by children using snap! block programming in a developing country," 2018.

M. Kandlhofer, G. Steinbauer, S. Hirschmugl-Gaisch, and P. Huber, "Artificial intelligence and computer science in education: From kindergarten to university," in *2016 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2016, pp. 1–9.

M. H. Kaspersen, K.-E. K. Bilstrup, M. Van Mechelen, A. Hjorth, N. O. Bouvin, and M. G. Petersen, "Votestratesml: A high school learning tool for exploring machine learning and its societal implications," in *FabLearn Europe/MakeEd 2021-An International Conference on Computing, Design and Making in Education*, 2021, pp. 1–10.

M. H. Kaspersen, K.-E. K. Bilstrup, and M. G. Petersen, "The machine learning machine: A tangible user interface for teaching machine learning," in *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, 2021, pp. 1–12.

S. Kim, Y. Jang, W. Kim, S. Choi, H. Jung, S. Kim, and H. Kim, “Why and what to teach: Ai curriculum for elementary school,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 17, 2021, pp. 15 569–15 576.

Y. Kim, Y. Murai, and S. Chang, “Embedded assessment tools for maker classrooms: A design-based research approach,” 2020.

S.-C. Kong, “Components and methods of evaluating computational thinking for fostering creative problem-solvers in senior primary school education,” in *Computational thinking education*, S.-C. Kong and H. Abelson, Eds. Springer Nature, 2019, ch. 8, pp. 119–143.

T. Lancaster, A. V. Robins, and S. A. Fincher, “14 assessment and plagiarism,” *The Cambridge handbook of computing education research*, p. 414, 2019.

N. Lao, I. Lee, and H. Abelson, “A deep learning practicum: Concepts and practices for teaching actionable machine learning at the tertiary education level,” 2019.

J. Lassnig, “Development of a curriculum to teach basics of artificial intelligence,” Ph.D. dissertation, Master’s thesis, Graz University of Technology, Graz, Austria, 2018.

I. Lee, S. Ali, H. Zhang, D. DiPaola, and C. Breazeal, “Developing middle school students’ ai literacy,” in *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, 2021, pp. 191–197.

P.-Y. Lin, C.-S. Chai, M. S.-Y. Jong, Y. Dai, Y. Guo, and J. Qin, “Modeling the structural relationship among primary students’ motivation to learn artificial intelligence,” *Computers and Education: Artificial Intelligence*, vol. 2, p. 100006, 2021.

P. Lin, J. Van Brummelen, G. Lukin, R. Williams, and C. Breazeal, “Zhorai: Designing a conversational agent for children to explore machine learning concepts,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 09, 2020, pp. 13 381–13 388.

D. Long and B. Magerko, “What is ai literacy? competencies and design considerations,” in *Proceedings of the 2020 CHI conference on human factors in computing systems*, 2020, pp. 1–16.

D. Long, A. Padiyath, A. Teachey, and B. Magerko, “The role of collaboration, creativity, and embodiment in ai learning experiences,” in *Creativity and Cognition*, 2021, pp. 1–10.

Z. Lyu, S. Ali, and C. Breazeal, “Introducing variational autoencoders to high school students,” *arXiv preprint arXiv:2111.07036*, 2021.

R. Mariescu-Istodor and I. Jormanainen, "Machine learning for high school students," in *Proceedings of the 19th Koli Calling International Conference on Computing Education Research*, 2019, pp. 1–9.

L. S. Marques, C. Gresse von Wangenheim, and J. C. Hauck, "Teaching machine learning in school: A systematic mapping of the state of the art," *Informatics in Education*, vol. 19, no. 2, pp. 283–321, 2020.

G. I. Melsión, I. Torre, E. Vidal, and I. Leite, "Using explainability to help children understandgender bias in ai," in *Interaction Design and Children*, 2021, pp. 87–99.

B. Mobasher, L. Dettori, D. Raicu, R. Settimi, N. Sonboli, and M. Stettler, "Data science summer academy for chicago public school students," *ACM SIGKDD Explorations Newsletter*, vol. 21, no. 1, pp. 49–52, 2019.

D. T. K. Ng, J. K. L. Leung, S. K. W. Chu, and M. S. Qiao, "Conceptualizing ai literacy: An exploratory review," *Computers and Education: Artificial Intelligence*, vol. 2, p. 100041, 2021.

N. Norouzi, S. Chaturvedi, and M. Rutledge, "Lessons learned from teaching machine learning and natural language processing to high school students," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 09, 2020, pp. 13 397–13 403.

M.-J. Ojeda-Bazaran, L. Chamba-Eras, E. Coronel-Romero, M. Labanda-Jaramillo, L. Conde-Zhingre, D. Irene-Robalino, W. Fierro-Saltos, A. Orellana-Malla, M. Romero-Flores, G. Cueva-Alvarado *et al.*, "State of the art of teaching-learning of artificial intelligence at early ages," in *2021 XVI Latin American Conference on Learning Technologies (LACLO)*. IEEE, 2021, pp. 165–172.

V. Olari and R. Romeike, "Addressing ai and data literacy in teacher education: A review of existing educational frameworks," in *The 16th Workshop in Primary and Secondary Computing Education*, 2021, pp. 1–2.

V. Olari, K. Cvejoski, and Ø. Eide, "Introduction to machine learning with robots and playful learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 17, 2021, pp. 15 630–15 639.

K. Park, B. Mott, S. Lee, K. Glazewski, J. A. Scribner, A. Ottenbreit-Leftwich, C. E. Hmelo-Silver, and J. Lester, "Designing a visual interface for elementary students to formulate ai planning tasks," in *2021 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE Computer Society, 2021, pp. 1–9.

J. D. Rodríguez-García, J. Moreno-León, M. Román-González, and G. Robles, "Evaluation of an online intervention to teach artificial intelligence with learningml to 10-16-year-old students," in *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, 2021, pp. 177–183.

G. Rosen, J. Silverman, and S. Essinger, “Inquiry-based learning through image processing,” *IEEE Signal Processing Magazine*, vol. 29, no. 1, pp. 164–169, 2011.

D. Rowntree, *Assessing students: How shall we know them?* Routledge, 1977.

A. Sabuncuoglu, “Designing one year curriculum to teach artificial intelligence for middle school,” in *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education*, 2020, pp. 96–102.

B. Sakulkueakulsuk, S. Witoon, P. Ngarmkajornwiwat, P. Pataranutaporn, W. Surareungchai, P. Pataranutaporn, and P. Subsoontorn, “Kids making ai: Integrating machine learning, gamification, and social context in stem education,” in *2018 IEEE international conference on teaching, assessment, and learning for engineering (TALE).* IEEE, 2018, pp. 1005–1010.

I. T. Sanusi, S. S. Oyelere, F. J. Agbo, and J. Suhonen, “Survey of resources for introducing machine learning in k-12 context,” in *2021 IEEE Frontiers in Education Conference (FIE).* IEEE, 2021, pp. 1–9.

M.-M. Schaper, L. Malinverni, and C. Valero, “Robot presidents: Who should rule the world? teaching critical thinking in ai through reflections upon food traditions,” in *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, 2020, pp. 1–4.

G. Shamir and I. Levin, “Transformations of computational thinking practices in elementary school on the base of artificial intelligence technologies,” in *Proceedings of EDULEARN20 Conference*, vol. 6, 2020, p. 7th.

——, “Neural network construction practices in elementary school,” *KI-Künstliche Intelligenz*, pp. 1–9, 2021.

Z. Skinner, S. Brown, and G. Walsh, “Children of color’s perceptions of fairness in ai: An exploration of equitable and inclusive co-design,” in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–8.

S. Srikant and V. Aggarwal, “Introducing data science to school kids,” in *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, 2017, pp. 561–566.

J. Su and W. Yang, “Artificial intelligence in early childhood education: A scoping review,” *Computers and Education: Artificial Intelligence*, p. 100049, 2022.

D. Tang, Y. Utsumi, and N. Lao, “Pic: A personal image classification webtool for high school students,” in *Proceedings of the International Workshop on Education in Artificial Intelligence K–12 EDUAI-20*, 2019.

M. Tedre, T. Toivonen, J. Kahila, H. Vartiainen, T. Valtonen, I. Jormanainen, and A. Pears, “Teaching machine learning in k–12 classroom: Pedagogical and technological trajectories for artificial intelligence education,” *IEEE Access*, vol. 9, pp. 110 558–110 572, 2021.

D. Touretzky, C. Gardner-McCune, F. Martin, and D. Seehorn, "Envisioning ai for k-12: What should every child know about ai?" in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 9795–9799.

T. Tseng, Y. Murai, N. Freed, D. Gelosi, T. D. Ta, and Y. Kawahara, "Plushpal: Storytelling with interactive plush toys and machine learning," in *Interaction Design and Children*, 2021, pp. 236–245.

F. Turbak and R. Berg, "Robotic design studio: Exploring the big ideas of engineering in a liberal arts environment," *Journal of Science Education and Technology*, vol. 11, no. 3, pp. 237–253, 2002.

J. Ureta and J. P. Rivera, "Using chatbots to teach stem related research concepts to high school students," in *Proceedings of the 26th International Conference on Computers in Education*, 2018.

M. E. Vachovsky, G. Wu, S. Chaturapruek, O. Russakovsky, R. Sommer, and L. Fei-Fei, "Toward more gender diversity in cs through an artificial intelligence summer program for high school girls," in *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, 2016, pp. 303–308.

J. Van Brummelen, K. Weng, P. Lin, and C. Yeo, "Convo: What does conversational programming need?" in *2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, 2020, pp. 1–5.

J. Van Brummelen, V. Tabunshchyk, and T. Heng, "" can i program you?": Student perceptions of conversational artificial intelligence before and after programming alexa," *arXiv preprint arXiv:2102.01367*, 2021.

H. Vartiainen, M. Tedre, and T. Valtonen, "Learning machine learning with very young children: Who is teaching whom?" *International journal of child-computer interaction*, vol. 25, p. 100182, 2020.

H. Vartiainen, T. Toivonen, I. Jormanainen, J. Kahila, M. Tedre, and T. Valtonen, "Machine learning for middle-schoolers: Children as designers of machine-learning apps," in *2020 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2020, pp. 1–9.

——, "Machine learning for middle schoolers: Learning through data-driven design," *International Journal of Child-Computer Interaction*, vol. 29, p. 100281, 2021.

C. G. Von Wangenheim, L. S. Marques, and J. C. Hauck, "Machine learning for all–introducing machine learning in k-12," 2020.

I. Voulgari, M. Zammit, E. Stouraitis, A. Liapis, and G. Yannakakis, "Learn to machine learn: Designing a game based approach for teaching machine learning to primary and secondary education students," in *Interaction Design and Children*, 2021, pp. 593–598.

X. Wan, X. Zhou, Z. Ye, C. K. Mortensen, and Z. Bai, "Smileycluster: supporting accessible machine learning in k-12 scientific discovery," in *Proceedings of the Interaction Design and Children Conference*, 2020, pp. 23–35.

R. Williams, "How to train your robot: Project-based ai and ethics education for middle school classrooms," in *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, 2021, pp. 1382–1382.

R. Williams, H. W. Park, and C. Breazeal, "A is for artificial intelligence: the impact of artificial intelligence activities on young children's perceptions of robots," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–11.

R. Williams, H. W. Park, L. Oh, and C. Breazeal, "Popbots: Designing an artificial intelligence curriculum for early childhood education," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 9729–9736.

C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, 2014, pp. 1–10.

S. Yoder, C. Tatar, I. Aderemi, S. Boorugu, S. Jiang, and B. Akram, "Gaining insight into effective teaching of ai problem-solving through csedm: A case study," 2021.

Y. Zhang, J. Wang, F. Bolduc, W. G. Murray, and W. Staffen, "A preliminary report of integrating science and computing teaching using logic programming," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 9737–9744.

Z. Zhang, S. Willner-Giwerc, J. Sinapov, J. Cross, and C. Rogers, "An interactive robot platform for introducing reinforcement learning to k-12 students," in *International Conference on Robotics in Education (RiE)*. Springer, 2021, pp. 288–301.

X. Zhou, J. Van Brummelen, and P. Lin, "Designing ai learning experiences for k-12: emerging works, future opportunities and a design framework," *arXiv preprint arXiv:2009.10228*, 2020.

J. Zhu and J. Van Brummelen, "Teaching students about conversational ai using convo, a conversational programming agent," in *2021 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, 2021, pp. 1–5.

A. Zimmermann-Niefield, M. Turner, B. Murphy, S. K. Kane, and R. B. Shapiro, "Youth learning machine learning through building models of athletic moves," in *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, 2019, pp. 121–132.