



Co-designing Large Language Model Tools for Project-Based Learning with K12 Educators

Prerna Ravi
Massachusetts Institute of Technology
Cambridge, MA, USA
prernar@mit.edu

John Masla
Massachusetts Institute of Technology
Cambridge, MA, USA
j_masla@mit.edu

Gisella Kakoti
Massachusetts Institute of Technology
Cambridge, MA, USA
gkakoti@alum.mit.edu

Grace C. Lin
Massachusetts Institute of Technology
Cambridge, MA, USA
gcl@mit.edu

Emma Anderson
Massachusetts Institute of Technology
Cambridge, MA, USA
eanderso@mit.edu

Matt Taylor
Massachusetts Institute of Technology
Cambridge, MA, USA
mewtaylor@gmail.com

Anastasia K. Ostrowski
Purdue University
West Lafayette, IN, USA
akostrow@purdue.edu

Cynthia Breazeal
Massachusetts Institute of Technology
Cambridge, MA, USA
cynthiab@media.mit.edu

Eric Klopfer
Massachusetts Institute of Technology
Cambridge, MA, USA
klopfer@mit.edu

Hal Abelson
Massachusetts Institute of Technology
Cambridge, MA, USA
hal@mit.edu

Abstract

The emergence of generative AI, particularly large language models (LLMs), has opened the door for student-centered and active learning methods like project-based learning (PBL). However, PBL poses practical implementation challenges for educators around project design and management, assessment, and balancing student guidance with student autonomy. The following research documents a co-design process with interdisciplinary K-12 teachers to explore and address the current PBL challenges they face. Through teacher-driven interviews, collaborative workshops, and iterative design of wireframes, we gathered evidence for ways LLMs can support teachers in implementing high-quality PBL pedagogy by automating routine tasks and enhancing personalized learning. Teachers in the study advocated for supporting their professional growth and augmenting their current roles without replacing them. They also identified affordances and challenges around classroom integration, including resource requirements and constraints, ethical concerns, and potential immediate and long-term impacts. Drawing on these, we propose design guidelines for future deployment of LLM tools in PBL.

CCS Concepts

• **Human-centered computing** → **Participatory design**; *Empirical studies in HCI*; • **Computing methodologies** → *Artificial intelligence*; • **Applied computing** → **Education**.



This work is licensed under a Creative Commons Attribution 4.0 International License.
CHI '25, Yokohama, Japan
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1394-1/25/04
<https://doi.org/10.1145/3706598.3713971>

Keywords

Generative AI, LLMs, AI for education, project-based learning, co-design, teachers, interviews

ACM Reference Format:

Prerna Ravi, John Masla, Gisella Kakoti, Grace C. Lin, Emma Anderson, Matt Taylor, Anastasia K. Ostrowski, Cynthia Breazeal, Eric Klopfer, and Hal Abelson. 2025. Co-designing Large Language Model Tools for Project-Based Learning with K12 Educators. In *CHI Conference on Human Factors in Computing Systems (CHI '25)*, April 26–May 01, 2025, Yokohama, Japan. ACM, New York, NY, USA, 25 pages. <https://doi.org/10.1145/3706598.3713971>

1 INTRODUCTION

Project-based learning (PBL) has gained prominence as a K-12 educational approach that immerses students in meaningful, real-world tasks, fostering deeper learning experiences [17, 27, 94, 96, 115]. Unlike traditional instructional methods, PBL emphasizes student-centered pedagogy, where learners actively construct knowledge through exploration, collaboration, and reflection [32, 89, 99]. By providing students with opportunities to explore topics of personal and cultural relevance, PBL supports diverse learning needs and backgrounds, making education more inclusive and accessible.

While PBL offers significant benefits to students, it also necessitates a fundamental shift in educators' roles from knowledge providers to learning facilitators [41, 112]. This requires educators to adopt new pedagogical strategies, manage complex classroom dynamics and provide ongoing, formative assessments that support student learning [5, 110]. Challenges to PBL implementation include designing standards-aligned projects that meet diverse student needs, managing planning time, transitioning students to active

roles [27, 41, 46, 119, 129, 136, 147], creating authentic driving questions and assessing interpersonal skills, which traditional assessment methods often fail to capture [3, 16, 19, 31, 80, 84, 147]. Such challenges emphasize the need for nuanced instructional strategies for PBL that better capture the scope of student learning [147].

To address these challenges, the Human-Computer Interaction (HCI) community has shown a rising interest in generative artificial intelligence (GenAI) systems for education [22, 51, 86, 118, 123, 141, 147]. Large language models (LLMs) in particular are uniquely suited to address the inherent complexities of PBL [50, 88, 146] due to their capacity to personalize learning and offer real-time custom feedback with the potential to support differentiated instruction in PBL contexts [4, 11, 67]. By automating routine tasks and augmenting project workflows, these tools could help streamline classroom management, thereby supporting teachers in delivering more engaging learning experiences [68, 147]. Given the potential of LLMs to transform PBL, it is essential that these tools are designed prioritizing educators' needs and experiences, with teachers actively involved in the design process [13, 52, 83, 97, 139].

1.1 Our Contribution

To the best of our knowledge, this is the first co-design study with K-12 teachers aimed specifically at incorporating LLMs into PBL pedagogy. Given the unprecedented and rapidly advancing nature of GenAI and its ethical considerations, our research is particularly timely. Our research objectives (RO) include:

- (1) **Explore Demands and Challenges in PBL:** We investigated specific challenges educators encounter when implementing PBL in their classrooms, focusing on how they assess student learning outcomes in areas such as teamwork, creativity, and problem-solving.
- (2) **Co-design GenAI Tools for PBL:** Next, we explored how GenAI-powered tools, co-designed with interdisciplinary educators, could address diverse needs and challenges of PBL environments. Our co-design workshops examined how these tools could alleviate teachers' workloads while supporting their current practices, pedagogical goals, and professional growth. We constructed wireframes embodying teachers' values and received iterative feedback, from teachers with diverse expertise, on their potential integration into PBL environments.
- (3) **Outline Design Considerations for LLM Tools in PBL:** Finally, we propose a set of design guidelines for LLM tools that educational technology designers and educators may find valuable for assessing student progress, setting learning goals, providing iterative and personalized feedback, and managing resources, timelines, and artifacts across different student projects and milestones.

Our findings provide valuable insights into shaping the future use of LLMs in PBL, *guiding and encouraging* its responsible implementation in classrooms. It should be noted that although we used the term GenAI for the study framing with our participants, they focused on a particular type of GenAI: LLMs for their proposed solutions. Because of this duality, both these terms are used in this paper.

2 RELATED WORK

2.1 Project-Based Learning

Rooted in constructivist learning theories, project-based learning (PBL) places students at the center of the learning process, challenging conventional education practices and redefining the roles of teaching, learning, and school organization [45, 49, 66, 142]. Unlike problem-based learning, which promotes deductive reasoning [32, 131], and involves structured evaluation based on how well a student's solution addresses a defined problem (e.g., designing a more effective trash can to improve adoption or conducting participatory action research on a social issue), project-based learning engages students in complex, real-world tasks to create a broader range of artifacts (e.g. reports, models, and presentations) that demonstrate learning [14, 27, 68, 90, 130, 131] (e.g., using photo-journalism to explore local flora and fauna or writing a comic book to analyze tropes in the hero's journey). This fosters creativity and critical thinking through open-ended driving questions (e.g. "what is the proper role of government in a democracy?") [16, 35, 95, 124]. Research has consistently demonstrated the numerous benefits of PBL across various educational contexts. Unlike traditional instruction, PBL (1) encourages students to delve deeply into the subject matter, fostering a more profound understanding of the material [17, 27, 94, 115]; (2) enhances mastery of subjects and supports the development of critical 21st-century skills like collaboration and communication [27, 32, 89, 99]; (3) boosts motivation, and engagement [15, 16, 55, 58, 60, 61]; and (4) develops metacognitive skills by helping students assess their progress and continuously improve [15, 40, 129, 131, 134].

While PBL is inherently student-centered, the success of this approach is heavily influenced by the role of teachers as facilitators, guiding and scaffolding student learning [14]. This shift in roles and professional identity requires teachers to adapt their pedagogical approaches and develop new skills, often leading to greater ownership, self-efficacy, and confidence in their teaching [29, 54, 101]. However, transitioning to PBL often requires targeted professional development (PD) that provides teachers with peer collaboration, reflection, and ongoing support during PBL implementation [2, 16, 37, 102, 143]. Given the pivotal role of teachers in harnessing PBL benefits, our research reveals their needs and explores how LLMs could support them in successfully implementing PBL and foster continued professional growth.

To maximize the effectiveness of these tools, it is essential to consider challenges teachers encounter during PBL implementation, especially around generating authentic driving questions, managing time and group work, and balancing instructor-led guidance with student-directed learning [85, 130, 147]. Students' discomfort with the cognitive and social demands of PBL can lead to frustration, particularly among high-achieving students accustomed to traditional instruction [32]. In addition, teachers face challenges initiating inquiry, facilitating dialogue, and scaffolding learning [62, 69, 103, 105, 106]. Assessing student learning in PBL is also notably difficult, as traditional tests often fail to capture the depth of understanding that PBL aims to develop. Performance-based assessments are hard to implement reliably [10, 34, 56, 85], student artifacts can be difficult to score consistently, and teachers lack the time to provide personalized student feedback [53]. Integrating

technology is also challenging due to broader institutional factors such as limited resources, district mandates, and lack of school tech maintenance and support [147]. Our study unpacks these PBL challenges in the literature through the nuanced experiences of teachers in interdisciplinary settings and explores how LLMs impact their roles and instructional practices.

2.2 AI Tools to Support PBL Pedagogical Needs

AI's capacity to personalize learning is beneficial in PBL settings where differentiated instruction is crucial. AI tools can support targeted instruction by tailoring lessons for diverse learners, encouraging student iteration on artifacts, and adapting materials to individual student strengths and weaknesses [4, 11, 67]. Helping students manage time and materials efficiently, tools like Trello and Cronofy, integrated with AI plug-ins, can predict resource needs [127]. AI-assisted design tools like Autodesk Dreamcatcher and intelligent project management software such as Asana and Monday.com can facilitate student project management, allowing them to focus on creative and critical thinking [127].

Considering project collaboration and knowledge construction [11, 38, 127], research scholars have discussed how AI can evaluate group performance by analyzing interaction patterns and predicting outcomes based on academic and behavioral data, enabling more effective grouping strategies [24, 113]. AI systems can also guide students toward productive problem-solving, suggesting activities on collaborative styles, and pedagogical interventions to improve group work [1, 39, 70]. Considering grading in PBL, AI-powered assessment systems, such as Automated Essay Scoring (AES) and Automated Written Corrective Feedback (AWCF), can offer real-time, continuous feedback, helping students to refine their work iteratively [33, 109]. These systems not only reduce the teacher's workload but also enhance the accuracy and efficiency of grading, allowing teachers to focus on more meaningful interactions with students [93]. AI can also support formative assessment practices in PBL by providing insights into students' progress, helping teachers guide and support students more effectively [74].

Despite potential benefits, integrating LLMs into PBL presents challenges, such as biased algorithms, that necessitates careful design and human oversight to ensure fairness and accuracy in assessment and feedback [114]. Comprehensive PD is needed for teachers to use LLM tools responsibly and equitably [9, 74], understand their long-term effects across age groups and subject areas, and address ethical concerns like data privacy [11, 140, 144]. In our current work, we aim to overcome some of these challenges by co-designing GenAI-powered support systems with *teachers* in PBL settings. This is unlike recent studies that have focused on co-designing these tools with *students* in the higher education contexts [50, 88, 146, 147]. We also examine challenges more unique to K-12 PBL, that emphasizes scaffolding and skill-building [32].

2.3 Co-designing Educational Tools with Teachers

Co-design is a collaborative, iterative process where teachers, researchers, and developers jointly design, prototype, and evaluate educational tools [108]. This approach leverages stakeholders'

expertise to address concrete educational needs, making it well-suited for creating technology-enhanced learning environments that are contextually relevant and practical for real-world classrooms [20, 30, 78]. Studies have documented the co-design of mobile science applications, scripted wiki environments, and other educational technologies that support authentic scientific inquiry and other pedagogical practices [30, 98, 121, 145].

Involving teachers in the co-design process enhances development efficiency, increases teacher agency and ownership, and promotes higher adoption rates and sustained use of educational tools beyond the initial study [13, 52, 77, 83, 97, 139]. Co-design can also serve as a form of professional development, enabling teachers to deepen their understanding of new technologies and explore ways to incorporate them into instructional strategies [13, 137, 139]. Because co-design fosters a more reciprocal and participatory approach [6], it has the potential to upend the traditional, unidirectional process of educational technology [128] and enables teachers to engage with complex topics beyond their expertise [36]. Furthermore, co-design has the potential to influence the broader sociocultural context of schools and drive systemic change in educational technology design and implementation by reshaping power dynamics and promoting an ethics of care within educational research and design [57, 81, 138]. Scholars have however noted that involving teachers in the human centered design process happens less often in K-12 settings compared to higher education [132].

3 METHODS

We conducted two studies: the first with expert PBL teachers and the second with teachers who had varying levels of PBL expertise. We explain our definition for expertise in section 3.1.1. By involving both groups, we sought to address the varied challenges and opportunities in PBL classrooms. Our studies were approved by the Institutional Review Board (IRB) at Massachusetts Institute of Technology. We obtained informed consent from all participants.

3.1 STUDY 1

In this study, we engaged interdisciplinary teachers with high levels of PBL experience and comfort. We worked with experts to leverage their deep pedagogical knowledge and refined understanding of effective PBL strategies, accumulated through years of trial, reflection, and adaptation. They participated in semi-structured interviews, a two-step co-design process, and feedback sessions on wireframes. We followed guidelines put forth in the literature [79, 91] by incorporating a mix of both divergent and convergent design thinking opportunities for participants. The divergent stages facilitated the generation of numerous ideas and concepts, while convergent stages focused on refining and narrowing down these ideas [91]. By offering diverse tools and activities, we enabled participants to discover the most effective methods for articulating their thoughts and generating new ideas [91].

3.1.1 Participants and Recruitment. Eleven participants were recruited via mailing lists, direct emails, collaborations, and social media across the United States. We specifically recruited participants with high levels of experience and comfort with PBL. This was the only precondition to participate in the study. Teachers were

compensated \$200 for participation (of approximately 10 hours) on this study.

Teachers completed a pre-survey covering demographic information and details regarding their past PBL and GenAI experience as part of the recruitment process. We received over 50 responses from interested teachers. We narrowed this list down to select our final 11 participants for the study [23]. We aimed to have a diverse sample by selecting teachers from multiple U.S. states, school systems (rural/urban/suburban areas as well as public/private schools with varied resource access [47]), subject backgrounds, varying levels of GenAI experience, and diverse student age groups. We also sought teachers from both STEM and non-STEM classrooms, those who work with historically marginalized students, and those who have taught PBL online [42, 104, 133, 135]. We wanted to intentionally design GenAI PBL tools with the above factors in mind from the beginning to ensure inclusive educational practices and avoid exacerbating existing disparities [42].

Table 1 breaks down the participants recruited (from the pre-survey responses). The teachers instructed various subjects, including computer science ($n = 3$), history ($n = 2$), English ($n = 1$), math ($n = 2$), and design, tech, and engineering ($n = 3$). Additionally, P06 served as a tech integration specialist, collaborating with other teachers to design and implement technology-enhanced projects. P05 and P07 taught PBL in online settings. Teachers, on average, had 5+ years of experience implementing PBL and reported an average comfort level of 4.6/5. In this study, we define PBL expertise as a combination of years of experience and comfort levels (showing self-efficacy) resulting from focused PBL implementation [116]. Comfort, in particular, is influenced not only by prior experience but also by the school systems teachers operate within. For instance, P03 described themselves as a PBL expert despite having relatively few years of experience, attributing their confidence to the rigorous training and practice received from working at a school strongly focused on experiential learning [73].

3.1.2 Data Collection.

Interviews. Between the months of March and April 2024, we conducted 11 semi-structured interviews [72]. All interviews were audio recorded with participant consent, ranged from 60-90 minutes each, and were conducted remotely via Zoom. We used these interviews as an opportunity to build rapport with our participants prior to co-design. The narrative style of the interview encouraged participants to engage in divergent thinking, allowing them to respond to open-ended questions freely with minimal redirection from the interviewers [91].

During the interviews, we asked participants about their teaching contexts and experiences with PBL, starting with its integration into the curriculum, alignment with other methods, technology use, and contextual factors like teacher-student ratios. We then explored strategies for maintaining coherence across projects, fostering student engagement, and tracking progress. A significant portion of the interviews focused on assessments, including approaches to grading, rubric use, and challenges with differentiation and progress tracking. We also asked teachers about helping students transition from traditional learning to PBL, and their views on scaffolding techniques and the open-ended nature of PBL. The recruitment survey responses from our 11 participants informed

these interview questions (see Table 1). In addition, the broader topics in the protocol were decided based on the discussions outlined in the literature review by Condliffe et al. [32].

Co-design Workshops. We conducted two co-design workshops (each lasting 2.5-3 hours) with teachers to explore how GenAI tools could support diverse PBL classroom needs without disrupting current practices. Due to teachers' limited availability and varying schedules, each workshop was divided into three smaller group sessions with three to four teachers, enabling focused, interactive discussions in a manageable online setting. The sessions, which included teachers from different subject areas to ensure diverse perspectives, were conducted remotely via Zoom and recorded with participants' consent. Below, we outline the structure and activities of each workshop:

Workshop 1. The first co-design workshop aimed to identify key PBL challenges and explore how GenAI could help alleviate teacher workload. Each session included three to four teachers, a facilitator, and a notetaker. The make-up of each group was:

- Group 1: P11 (Computer Science), P06 (Design, Tech, Engineering), P03 (History), P07 (Mathematics)
- Group 2: P05 (Design, Tech, Engineering), P02 (Computer Science), P04 (Mathematics)
- Group 3: P01 (English), P08 (Computer Science), P10 (Design, Tech, Engineering), P09 (History)

Workshop participants engaged in hands-on tasks below to foster creative thinking and group collaboration, with discussions throughout to explore their ideas:

- (1) After establishing workshop norms [91], participants used Miro boards for three five-minute rounds of **open brainstorming**. They filled post-it notes identifying (1) key challenges in PBL, shared (2) past strategies and tools they had used to address those challenges, and (3) envisioned dream tool features. After brainstorming rounds, participants reviewed the post-it notes, labeled promising ideas, and linked them to past experiences and potential concerns. They then grouped these ideas into categories on the Miro boards, which helped reinforce the ideas that emerged from our interviews.
- (2) We then gave a **presentation about GenAI** with examples of the technology's use cases in K-12 education. We explained how ChatGPT works and encouraged participants to use this knowledge to explore its potential and limitations in classrooms. Teachers also shared their experiences using these tools in their teaching. Our teachers gravitated towards using LLMs during subsequent stages of the study, even though they may have referred to GenAI during discussions as it was a more commonly recognized term for them.
- (3) Finally, participants expanded on the initial problem scoping to ideate where GenAI (particularly LLMs) could be incorporated into their selected category from the board. They created **storyboards** [21] of concrete scenarios depicting PBL implementation challenges and a concept of how a fictional LLM tool could serve as a solution in their everyday teaching context. These storyboards were then shared and discussed within the group.

Participant	Gender	Available School Details	Subjects	Student groups	Age	Years of experience with PBL	Comfort levels with PBL	GenAI Tool Experience
P01	F	public, urban, socioeconomically disadvantaged students	English	9th-12th grade, higher ed, adults		More than 5 years	5	5
P02	M	public, suburban	Computer Science	9th-12th		More than 5 years	5	3
P03	M	public, experiential learning school	History	9th-12th		1-2 years	4	2
P04	M	public, urban, socioeconomically disadvantaged students	Mathematics	9th-12th		3-5 years	4	3
P05	F	public, online teaching	Design, Tech, and Eng	9th-12th, higher ed, adults		More than 5 years	5	4
P06	F	public, rural	Design, Tech, and Eng	3rd-5th, 6th-8th		More than 5 years	4	5
P07	M	online teaching	Mathematics	3rd-5th grade, 6th-8th, 9th-12th		More than 5 years	5	4
P08	F	private	Computer Science	3rd-5th, 6th-8th, 9th-12th, higher ed, adults		3-5 years	4	3
P09	M	private	History	kindergarten, 1st-2nd, 3rd-5th, 6th-8th		More than 5 years	5	4
P10	F	private, urban	Design, Tech, and Eng	9th-12th		More than 5 years	5	4
P11	F	public, suburban	Computer Science	9th-12th		More than 5 years	5	2

Table 1: Study 1 participants' details, N=11

Written material from the participants and notetakers, and audio recordings of the small group discussions were collected online for analysis.

Workshop 2. We summarized the brainstorming boards and storyboards from Workshop 1 to guide activities for Workshop 2 that developed conceptual design prototypes for PBL GenAI tools. Each group session consisted of three to four teachers, a primary facilitator, and a notetaker (similar to Workshop 1). We tried to reorganize participant groups to encourage a wider range of discussions across the two workshops:

- Group 1: P11 (Computer Science), P05 (Design, Tech, Engineering), P09 (History)
- Group 2: P01 (English), P08 (Computer Science), P10 (Design, Tech, Engineering)
- Group 3: P03 (History), P06 (Tech Integration: all subjects), P02 (Computer Science), P04 (Mathematics)

In this workshop, participants developed conceptual prototypes based on the storyboards from Workshop 1. They discussed their tools' LLM roles, its ethical impact on stakeholders, and classroom

integration constraints by collaboratively completing worksheets with specific prompts using Google Slides, suitable for the remote setting. An example of blank worksheets from some of the steps below is shown in Figure 1. We summarize this process below:

- (1) **Define Final Tool Idea:** Participant groups first selected 1-2 storyboards with overlapping themes for further exploration, wrote a detailed description, and ensured all members agreed on the concept, which became the basis for their final prototype.
- (2) **Identify Stakeholders:** Participants then identified direct and indirect stakeholders involved in their chosen scenarios and their specific roles and interaction with the tool.
- (3) **Identify How the GenAI Prototype Is Used:** Next, they detailed the role of the GenAI in the selected scenarios. Groups created a list of envisioned features, considered their data-sharing requirements, preferred devices for deployment (phone, tablet, or laptop), and potential integration with existing Learning Management Systems (LMS) like Canvas, Google Classroom, etc.

Step 4: Stakeholders Matrix

Values, benefits, harms for stakeholders from [ethical implications handout](#)

List the values, benefits, and harms for each stakeholder for your tool below

Values: core principles or standards that stakeholders hold important. These are the foundational beliefs that guide their expectations and interactions with the tool.

Benefits: positive outcomes or advantages that stakeholders gain from using the tool. **Harms:** potential negative outcomes or disadvantages that stakeholders might experience due to the tool's implementation.

Stakeholders	Values (what do they care about?)	Benefits (how will this tool help them)	Harms (how might this tool harm them)

Step 5: Ethical impact of tool

Think of short-term and long-term effects of this tool from [handout](#)

Based on the [handout](#) and stakeholders from previous slide, answer the following. Think of concerns they may raise around ethics, personal agency, and privacy for teachers and students:

What are the immediate short-term benefits and harms of such a tool in your classroom?

What are the long-term benefits and harms of such a tool in the PBL community?

How will you ensure transparency in communicating the tool's use to students?

Figure 1: Examples of blank worksheets distributed to participants in workshop 2 outlining stakeholder impacts and ethical implications of PBL-LLM tools

- (4) **Ethical Implications:** We then led participants through discussions of their designs' societal impacts. Participants received an ethical implications handout to identify values, benefits, and potential harms for each stakeholder group from Step 2 [7, 8, 75]. They also investigated the immediate short-term and long-term consequences of using this tool in their classrooms and the PBL community.
- (5) **Classroom Constraints:** Finally, participants considered practical aspects of integrating their prototype in their schools, including resource requirements, tool training and maintenance, anticipated challenges (technical, logistical, pedagogical), and strategies to address these challenges effectively.

Prototype worksheets, facilitator notes, and audio recordings of discussions were collected for analysis.

PBL GenAI Wireframes Walkthrough. Using data from interviews and co-design sessions, we created wireframes outlining valuable design features for teachers using PBL, focusing on curriculum support (project brainstorming and lesson planning), assessment support (rubric creation with differentiation and grading), and progress tracking (monitoring progress at the individual student, group, and class levels). These wireframes and support areas, described further in our results section, were reviewed in one-on-one sessions with teachers, who assessed the usability of the tool's

wireframes within their classroom contexts and unique pedagogical practices. Sessions were conducted remotely via Zoom and recorded with consent. We expanded on this feedback by including other teachers outside Study 1 (Section 3.2.1). We administered a post-survey at the end to understand participants' perceptions of GenAI and PBL and gather their reflections on Study 1's activities.

3.1.3 Data Analysis. We transcribed all interviews and took an inductive, iterative approach to thematic analysis [18]. Initial codes were identified from the transcripts and refined into broader themes, which were continuously reviewed to accurately capture participants' insights. Two researchers then coded the transcripts, achieving a Cohen's kappa of 0.906 on 40% of the data, indicating substantial interrater reliability [82, 122]. We then deductively applied these themes to the rest of our co-design and wireframe testing data [12].

3.2 STUDY 2

The bulk of the co-design process was conducted with the expert teachers in Study 1. Following this, we conducted a second study involving teachers with varying levels of experience and comfort with PBL (and hence varying expertise as per our definition in 3.1.1). This workshop aimed to gather a wider range of perspectives from a

Pre-survey metric	Collated participant responses
Participants' gender	14 female and 16 male
Subjects taught (multi-select)	Science (15), Mathematics (7), Computer Science (6), Engineering and Technology (4), STEAM (3), English (1), Business (1), and Teacher Education (1)
Students' age groups (multi-select)	Kindergarten (2), 1st-5th grade (7), 6th-8th grade (14), 9th grade and above (21)
Experience with PBL	46% for more than 5 years, 23% for 3-5 years, and 31% for a year or less.
Comfort levels with PBL (scale 1-5)	5: 40%, 4: 26.7%, 3: 16.7%, 2: 10%, 1: 6.7%
Describe past PBL experience	Difficult (Significant challenges, poor engagement, and outcomes): 23% Neutral (mix of challenges and successes): 46% Very successful (Highly positive, engaged students, and excellent outcomes): 33%
Experience with GenAI tools like ChatGPT	1: 20%, 2: 26.7%, 3: 26.7%, 4: 20%, 5: 6.7%

Table 2: Study 2 participants' details, N=30

diverse group of teachers, ensuring the tool wireframes could adapt to different teaching styles and classroom environments (as per RO2). This approach also evaluated the tool's potential to reduce barriers of entry to PBL and allow educators to focus more on experiential learning and holistic development.

3.2.1 Participants and Recruitment. Thirty STEM teachers from diverse U.S. states and school contexts were recruited in collaboration with the MIT Scheller Teacher Education Program and MIT Alumni Clubs. These teachers, admitted to the Science and Engineering Program for Teachers (SEPT) program, participated in a 2-hour workshop on GenAI for project-based and collaborative learning. We distributed a pre-survey to explore the backgrounds, subjects, and experiences of workshop participants (see Table 2). Teachers primarily taught middle and high school students and had different levels of PBL experience (average of 3.4 years) and comfort (average of 3.8/5) resulting in varying expertise. Participants were equally divided between those who had found success with PBL and those who had not.

3.2.2 Data Collection. Study 2's workshop started with brief discussions on participants' perceptions of GenAI and collaborative learning. Following these, participants were divided into three groups based on their pedagogical interests: curriculum planning, assessment, and progress monitoring. Similar to our procedure at the end of section 3.1.2 of Study 1, we walked Study 2 participants in our small group sessions through the same set of wireframes, which focused on curriculum support, assessment support, and progress tracking. After a brief review, discussions centered on the specific support area participants in the small group had selected. Participants explored the wireframes, provided feedback, and assessed usability in their classroom contexts. They also identified tool tasks best kept human-driven, and their impact on teaching creativity, well-being, workload, and stress. We followed this with a post-survey to gauge participants' excitement and concerns on GenAI, interest in PBL, desired additional features for the wireframes, and onboarding resources needed for classroom use.

3.2.3 Data Analysis. Discussions from all three groups were audio recorded with consent from participants for the purposes of

transcription. We then deductively applied the same themes from Study 1 to code our data from Study 2 [12]. **For the purposes of this paper, we focus on Study 1's interviews and co-design workshops, as well as the PBL GenAI wireframes and their feedback from Study 1 and Study 2. Note that the feedback on the wireframes has been analyzed and discussed by combining the results from both studies, as they together shaped the design guidelines and future directions for the tool.**

4 RESULTS

Table 3 shows themes and their corresponding definitions that emerged from our data. Section 4.1 delves into Category A themes that set the stage for foundational PBL needs and Category B themes regarding challenging aspects associated with addressing these needs. Section 4.2 highlights Category C themes that suggest pathways for integrating GenAI (specifically LLMs) into PBL, drawing on key insights from our two co-design workshops. Finally, Section 4.3 presents wireframes for our proposed PBL LLM tools, synthesizing themes from all three categories. This mapping of research objectives, study stages, and themes is shown in Figures 2 and 3.

4.1 Exploring Demands and Challenges in Project-Based Learning (Category A and B themes)

We present findings from our interviews in Study 1, focusing on themes highlighting teachers' core beliefs about the demands of high-quality PBL implementation (Category A in 4.1.1) and the challenges for them and their students (Category B in 4.1.2). These answer RO1. We recognize that some of the findings from these interviews align with themes already discussed in the literature. However, our primary aim was to build relationships with our teachers by gaining a nuanced, in-depth understanding of the types of projects they implement within their distinct subject areas, along with the specific challenges they face. By centering teachers' voices and lived experiences, we capture and communicate the authentic context of their pedagogy, enabling readers to better appreciate how the subsequent co-design artifacts are shaped by and respond to their unique classroom demands.

Category	Themes	Definitions
(A) Set stage for PBL needs	school context	details on subjects taught, grades, public/private school, student backgrounds, teacher experience, the way PBL is structured in the school
	specific project examples	specific examples that teachers gave on projects they have implemented in their classrooms. Include details about the project so we know what is, not just in reference to other parts pedagogy.
	student agency	how teachers give students autonomy, choice, freedom on projects, including group formation
	student differentiation	how material/grading/instruction is adapted for different groups/needs of students, including personalized support. Supports for some students.
	student scaffolding	any and all tools/strategies for scaffolding different parts of PBL for students. Supports all students get.
	teacher role	role of teacher in PBL setting
	personalized feedback	how individual and personal feedback is given to every student
	goals, checklists, progress tracking	anything around setting goals, milestones, checklists, deadlines, monitoring regular progress etc. (exit tickets)
	online learning platforms and tools	types of online learning platforms and grading platforms used (e.g. Canvas, Google Classroom)
	rubrics/grading scales	details on rubrics and grading schemes/strategies used for evaluating different parts of projects
	self and peer reflections	any details on self and peer reflections/feedback as part of projects
	managing student groups	details on managing student groups and dynamics, interpersonal skills and interactions. conflict resolution, assigned roles, leveraging different strengths etc.
	remote PBL	experiences and challenges of PBL in online settings. Includes structure and procedures for conducting class
	PBL inclusion	details on gender gaps, STEM, equity, etc.
(B) Challenging aspects	PBL challenges for students	specific details on challenges encountered by students in PBL
	PBL challenges for teachers	specific details on PBL challenges encountered by teachers
	PBL teacher perceptions	teacher perceptions of PBL
	PBL student perceptions	student perceptions of PBL
(C) AI + PBL	AI perceptions	any details on teacher's perceptions/details of AI, including GenAI
	current AI uses/challenges in classroom	details on use cases and challenges from using AI in classrooms
	PBL wish list	details on things teachers wish for in the future in their classrooms

Table 3: Final set of themes and their definitions from our data analysis. These span three main types: (A) those that set the stage for PBL needs, (B) those that describe challenging aspects of PBL implementation, and (C) those that show AI integration in PBL

4.1.1 Demands in Project-Based Learning. Teachers in our study emphasized the distinct elements of PBL that had to be addressed to foster engagement and promote deeper learning. P10 stressed the importance of focusing on the big picture while managing granular project details, helping students see the connections between project components. Teachers underscored the interdisciplinary nature of PBL, requiring them to develop the expertise to integrate subjects like physics, engineering, and design, to foster holistic learning. The **adaptability and improvisational aspects of PBL** were recurrent themes, as illustrated by P06:

“Last year, I was ready to throw my 3D printer out the window because students couldn’t even name their projects properly despite detailed instructions. I had to rethink my approach and make tweaks...Flexibility and

recognizing that you don’t have everything figured out is key to successful PBL.” – P06

Teachers shared how **prior PD experiences**, including workshops and training, were crucial for implementing PBL effectively and staying current with evolving educational practices and technologies. They also spoke about their roles and responsibilities as **“expert facilitators and coaches”**, helping students navigate projects while promoting independence. P02 noted that much of the time was spent observing, asking questions, and offering support as needed, allowing students to explore and experiment. Handling multiple projects required effective **context-switching and organization**. P07 stressed building student confidence, particularly in challenging subjects like math, and creating a supportive classroom

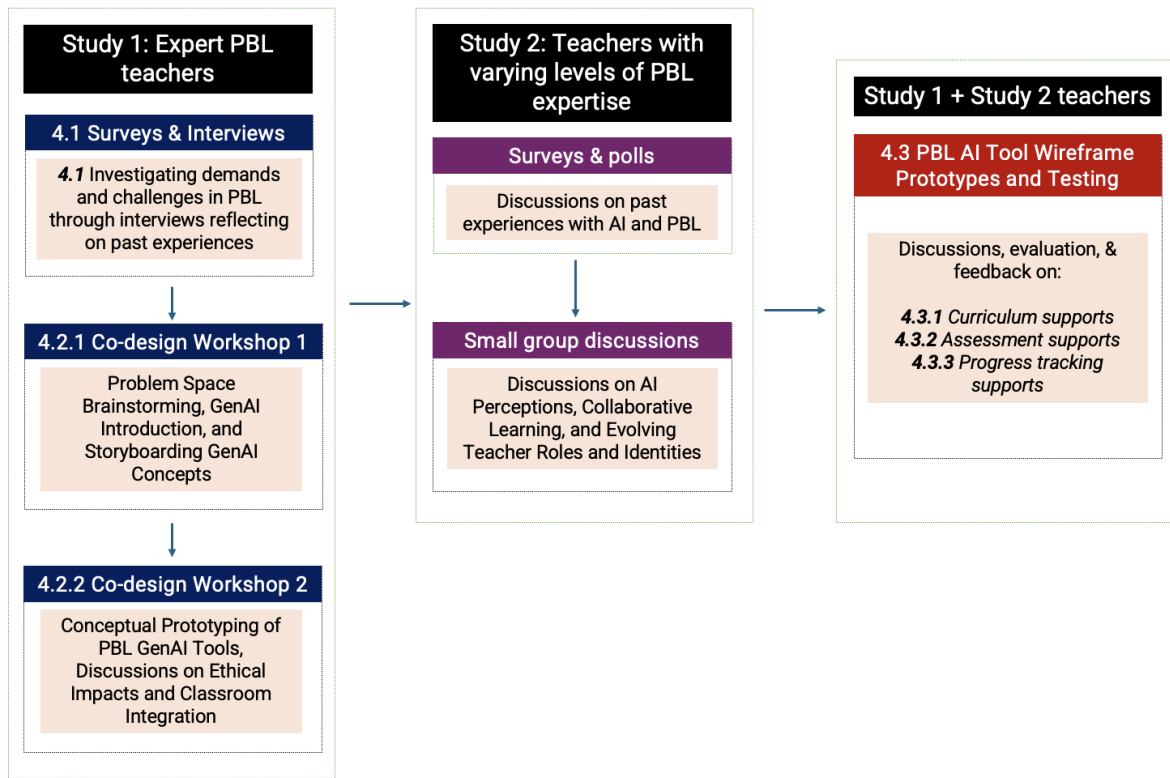


Figure 2: A timeline illustrating the various stages of our study alongside the corresponding paper section numbers where their findings are discussed. For Study 2, note that we present findings from the final wireframe testing phase only, as mentioned in Section 3.2.3

Research Objective	Study Stage	Coded Themes
R01: Explore demands & challenges in PBL	Study 1: Expert PBL teachers <ul style="list-style-type: none"> 4.1 Study 1: Surveys and Interviews 	Category A: Setting the stage for PBL needs + Category B: Challenging aspects of PBL
R02: Co-design AI Tools for PBL	<ul style="list-style-type: none"> 4.2.1 Study 1: Co-design Workshop 1 4.2.2 Study 1: Co-design Workshop 2 	Category C: AI integration in PBL
	Study 2: Teachers with varying levels of PBL expertise <ul style="list-style-type: none"> Study 2: Surveys and Polls Study 2: Small group discussions 	
	Study 1 + Study 2 teachers 4.3 PBL AI Tool Wireframe Prototypes and Testing	Category A: Setting the stage for PBL needs + Category B: Challenging aspects of PBL + Category C: AI integration in PBL
R03: Outline Design Considerations for AI PBL Tools	5.2 Discussion: AI Design Recommendations for Project-based Learning	N/A

Figure 3: A figure displaying the mapping of research objectives to the different stages of the study and their results' coded themes

where mistakes were part of learning. Our teachers relied on standardized rubrics and project descriptions to ensure fairness, but they had to **adapt the content to different abilities**, necessitating the creation of multiple versions of resources. Teachers also understood that some students preferred hands-off approaches while others needed frequent support, and they balanced these needs through annotated monitoring and **personalized interactions**:

"Some students embrace it and love the challenge and freedom to explore in their own way. Others resist and rebel against it, preferring a more traditional approach with worksheets, lectures, and tests. And then there's everything in between. Some are willing but don't know how to do it." –P03

Teachers used **scaffolding tools like checklists** to help students adapt to PBL's self-directed approach, keeping them organized and focused. They mapped rubric items to content standards and other criteria, such as creativity and sustainability, refining these annually to match evolving goals. To manage group projects, teachers proactively defined tasks and deadlines for individual members, conducted **regular check-ins**, and **encouraged reflection**, ensuring equitable contributions and progress. These strategies fostered a structured yet flexible learning environment, enhancing students' time management, self-assessment, problem-solving, and metacognitive skills.

4.1.2 Challenges in Project-Based Learning. Implementing PBL presented organizational difficulties for teachers, balancing content depth and breadth, and managing diverse student needs. The shift from traditional assessments to more dynamic, project-based evaluations also demanded innovative assessment methods. P02 highlighted the challenges of **evaluating open-ended projects without additional support** staff in schools. Teachers grappled with balancing the depth of knowledge required for specific projects with the breadth of content mandated by the curriculum. This often meant that students developed an in-depth understanding of their chosen project topic but lacked comprehensive knowledge of other areas, impacting their performance on traditional assessments. Integrating technology also added complexity, with students' enthusiasm for new tools sometimes **diverting focus from project goals**. Additionally, the availability of resources and the need to create or adapt materials presented ongoing struggles. Teachers often felt constrained by their curriculum's scope and sequence, which limited their ability to fully embrace PBL's potential. **Fair grading** was particularly challenging due to the subjective nature of projects and the need for continuous feedback. P10 highlighted the **workload** involved:

"It's a lot of work. I'm constantly grading. I don't get a break because it's a continuous cycle of work. I give iterative feedback throughout, and my turnaround is as quick as possible, but it means grading all weekend." –P10

Maintaining student engagement was another recurring challenge in PBL. P01 noted that some students (e.g. struggling readers) found it hard to stay on task, requiring additional prompts and support to remain focused. **The transition from lecture-based learning to PBL** was difficult, with students initially resisting

the shift from teacher-directed instruction. P05 recounted how some students found projects daunting and needed the **tasks broken into smaller, manageable steps**. This step-by-step approach helped students gradually build confidence and see their progress, ultimately leading to successful project completion. Teachers also stressed the importance of social-emotional development, noting the pandemic's negative impact on students' ability to collaborate and support each other. Flexibility and empathy were essential in helping students manage PBL challenges, particularly when overwhelmed or facing mental health issues. P09 observed that students often struggled to apply memorized concepts to projects, requiring a structured approach with front-loaded skills before diving into projects. The hands-on nature of PBL demanded continuous support to bridge the gap between conceptual knowledge and practice. P11 noted challenges with **unequal team participation**, affecting grades and contributions. Ensuring fairness required careful monitoring and regular teacher intervention.

These challenges in student engagement compounded in **remote settings**, with many students preferring chat over cameras or microphones. P05 used breakout rooms for small group discussions and maintained engagement through Zoom polls and exit tickets, despite inconsistent Wi-Fi and varying school district permissions. **Logistical barriers, like inconsistent access to resources**, further complicated remote PBL.

4.2 Integrating AI in Project-Based Learning (Category C themes)

Building on interviews, we present findings from the co-design process, where teachers explored customizing LLM tools to meet PBL's specific demands (Category C). These help answer RO2.

4.2.1 Co-design Workshop 1: The first workshop aimed to identify challenges teachers faced with implementing PBL and map scenarios in which LLMs tools could help.

Brainstorming- Challenges, Current Strategies, and "Magic wand" features: Across our three groups and rounds of brainstorming, we saw similarities in the categories that emerged.

Participants discussed strategies for **setting goals and tracking progress** in PBL. One suggestion was a *"dashboard to monitor progress with a traffic light system (red, yellow, green)"* (Group 1), for visual tracking. Another idea involved using a checklist in the form of a *"bingo board for final projects"* (Group 3). However, these tools sometimes failed when items were checked off without actual completion. Frequent check-ins and breaking projects into manageable tasks were also mentioned, though this process was noted to be labor-intensive due to the need for continuous tracking, timely feedback, and adjustments to lesson plans. Participants promoted **student autonomy** by *"maintaining momentum"* (Group 3) during classroom time, *"keeping project ideas fresh"* (Group 1), and organizing *"fun, creative tasks within projects to reduce cognitive and social discomfort"* (Group 1). Allowing students to *"pick goals with teacher feedback"* (Group 1), and *"outline the top three criteria they wish to be graded on"* (Group 3) gave them more control over their learning. However, balancing this autonomy with meeting curriculum standards remained a challenge.

Participants emphasized the need for **differentiation and personalization** by "challenging all students where they are" (Group 1) and providing "feedback targeted at their current level of maturity, technical skill, and aspirations" (Group 1). There was a call for tools that offer "additional resources tailored to students' strengths and weaknesses" (Group 3) and help navigating complex materials. Ensuring "equitable [opportunities] for students of diverse backgrounds and learning levels." (Group 3), along with having alternate plans if needed, was seen as crucial for student success.

Managing group dynamics was another area of focus within the brainstorming boards, such as dealing with "students who give minimal effort to the team" (Group 1) and "ensuring that everyone in the group understands the tasks" (Group 2). "Assessing [group] projects for the group as well as individuals" (Group 2) was noted as a difficulty, highlighting the need for fair rubrics. **Time management** also surfaced as a significant concern. While self-reflecting on incremental progress was seen as valuable for students, it was also time-consuming. Integrating technology required time to train students, and providing meaningful feedback posed a challenge when balancing thoroughness with efficiency.

Participants discussed various **grading strategies and challenges**, such as the use of external automated tools for scaffolded templates like "Repl.it, Codecheck.it, CodingRooms, Google Colab Notebooks" (Group 2), which were helpful but time-consuming to set up. There was debate over balancing and "assessing [final] product vs. [learning] process" (Group 2) within projects. Other concerns included integrating standards-based grading into rubrics while managing the pressure from parents to convert these into traditional letter grades for transcripts. Fairness in grading was also highlighted, given the complexity of creating rubrics for different project parts, student needs, and interdisciplinary topics. Some educators suggested involving students in the evaluation process through "self-grading along with teacher grading" (Group 3), which promoted student self-awareness and provided deeper insights to educators for personalized feedback.

Storyboarding Areas of Opportunity for GenAI Tools in PBL: Participants expanded on initial problem scoping by creating storyboards illustrating scenarios in which a hypothetical LLM tool could help address challenges in their daily teaching practices. Our analysis revealed three major areas of support: #1: curriculum and lesson planning, #2: assessments and grading, and #3: managing group dynamics and progress tracking.

[Support area #1]- Curriculum & Lesson Planning + [Support area #2]- Assessments & Grading: Our participants created storyboards where teachers and students used the LLM system to collaboratively brainstorm project ideas, generate lesson materials, and customize these for diverse student needs. They also envisioned the LLM aiding in creating assessments (rubrics), and automating grading with teacher input. The themes from support areas #1 and #2 frequently co-occurred in these scenarios. For example:

P11 highlighted the challenges of defining projects and designing assessments in PBL, which can be time-consuming (upwards of 40 hours). They noted difficulty in finding comprehensive resources online suited to various instructional needs. To address these issues, P11 envisioned a LLM-powered chatbot that assists in **project**

creation by generating tailored project plans based on class information and goals provided by the teacher (Figure 4). The tool would offer suggestions for student tasks, sentence starters, and **help design rubrics and checklists**, accommodating **differentiated instruction** by considering students' physical, academic, and behavioral needs. Additionally, it could generate documents like timelines, visual progress indicators, and follow-up exams. Other teachers suggested integrating an open-source database with both tested and untested unit plans for cross-referencing into this design. This storyboard streamlines labor-intensive tasks in PBL, allowing teachers to adapt projects instead of starting from scratch.

"For formative assessments, we need status checks to ensure students are on track during a project. For example, in a 3-day project, students might stay focused because the deadline is near. But in a longer project, they might get off track. The chatbot could provide feedback on their progress, motivation, and any issues they're facing, helping them stay on course... For summative assessments, I would want the chatbot to suggest feedback and explain its reasoning, allowing me to modify it as needed. Feedback should be age-appropriate and level-appropriate. For instance, some students need positive feedback, while others might be motivated by challenges. The chatbot should offer tailored suggestions and allow me to adjust them before finalizing." – P11

P05's storyboard (Figure 5) envisioned LLMs helping **create alternative or customized assessments** to accommodate different learning styles. They highlighted the challenge of creating varied assessment formats, such as multiple versions of vocabulary tests, to meet diverse needs and prevent sharing of test content between different class periods. P05 suggested LLMs could quickly generate similar essay prompts or multiple-choice questions, **maintaining fairness** across classes and compiling these into a centralized question bank. They also saw LLMs as a **brainstorming partner for generating activity ideas** when time is limited, allowing teachers to focus more on individualized instruction and feedback. However, they noted concerns about accessibility, particularly in online settings (pertinent to P05), where account restrictions limit access to GenAI tools like Google's Gemini or Microsoft's Copilot. Other participants discussed the learning curve in using GenAI tools like ChatGPT, noting the difficulty of recalling effective prompts and the need for prompt-sharing among educators for P05's tool.

[Support area #3]- Managing Group Dynamics and Progress Tracking: We observed storyboards illustrating LLM systems to monitor individual and group progress, streamline communication with students and parents, document project artifacts, track individual contributions within groups, ensure equitable participation, and resolve group conflicts.

For example, P10 envisioned an LLM system to **support teachers monitoring at-risk students** (Figure 6) by offering a comprehensive visual representation of a student's academic performance, highlighting key issues like missed assignments. The system would generate detailed reports and **automate communication with colleagues, parents, and administrators**. The LLM could also

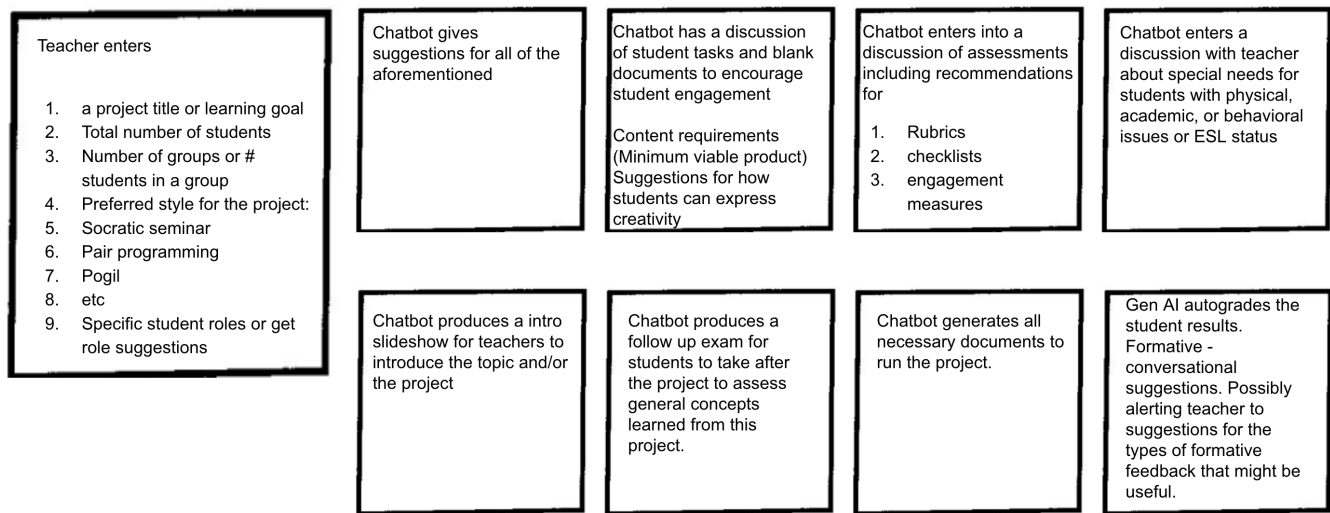


Figure 4: Storyboard (P11) mapping a step-by-step teacher-LLM interaction for project and assessment ideation

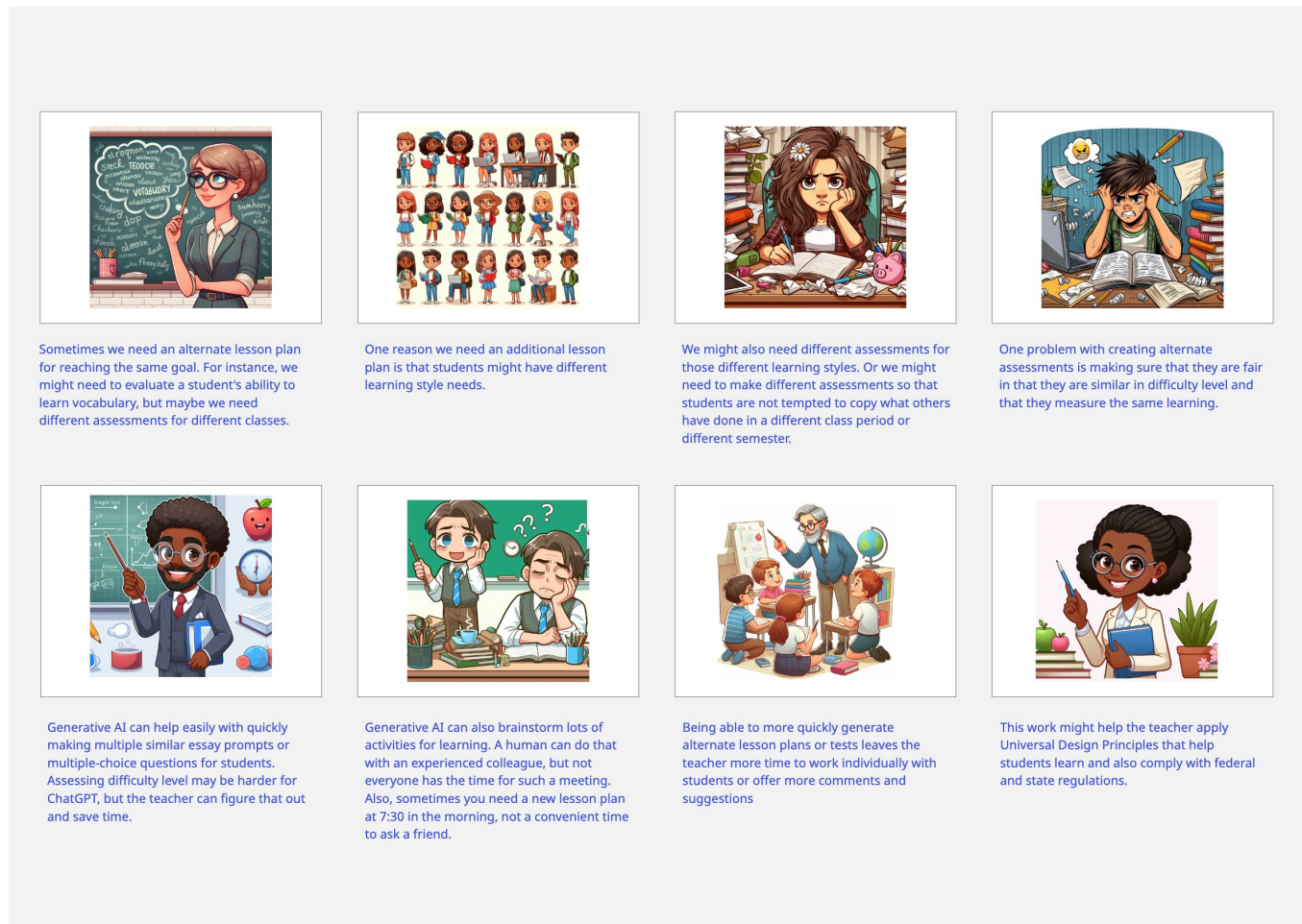


Figure 5: Storyboard (P05) showing an LLM tool for alternative assignments and lesson plans

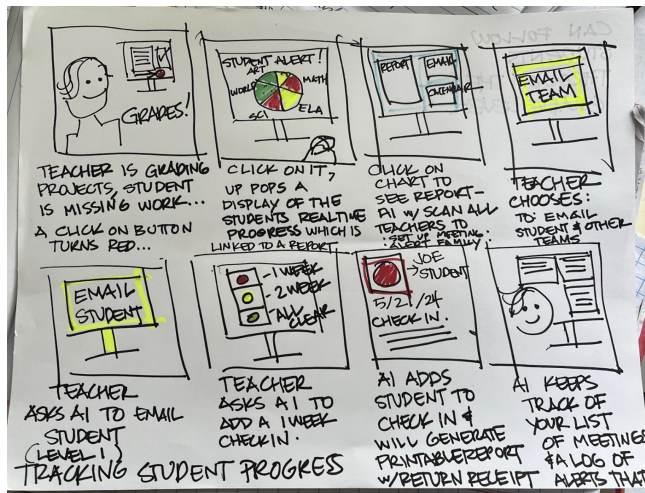


Figure 6: Storyboard (P10) depicting an LLM assistant to monitor at-risk students

help schedule meetings by coordinating calendars and setting reminders for follow-ups to facilitate timely interventions. P10 emphasized the value of LLMs in maintaining continuity of support as students progress through different grade levels, analyzing patterns such as chronic lateness or incomplete projects to enable proactive support. Other participants suggested that the tool could also be adapted for student leaders to **track team submissions** and manage project delays, to avoid adding to the teacher's workload.

4.2.2 Co-design Workshop 2: In this workshop, participants built a conceptual prototype based on storyboard ideas from Workshop 1. These prototypes aligned with the three support areas from the storyboards above. We present the discussions on tool stakeholders, LLM data sharing, classroom integration, and ethical impacts below.

Tool Stakeholders: Participants identified key groups involved in implementing LLM tools in PBL, outlining their values, potential benefits and harms: **Classroom teachers, co-teachers, and new/pre-service teachers** could benefit from reduced administrative workload and support for lesson planning but would be concerned about over-reliance on LLMs and loss of control. **Special education teachers** emphasize meeting diverse student needs and Individualized Education Program (IEP) goals, with worries about data privacy and maintaining personal connections with students. Participants also discussed involving **IT teams, the Board of Education, students, and parents** as stakeholders, each valuing student progress, data security, and educational equity, but wary of privacy concerns, technological divides, and the potential impact on teacher-student interactions.

Data sharing requirements of LLM prototypes: Participants suggested that the LLM tool should **access teacher-specific information**, such as state/local standards, project goals, past assignments, rubrics, guiding questions, and preferred instructional methods (e.g., Socratic seminars, Process Oriented Guided Inquiry Learning (POGIL)). This data would help the LLM align its outputs

with educational objectives and teaching styles. To personalize content, educators also stipulated that the system have **access to comprehensive student profiles**, including standardized test scores, reading and math levels, and accommodations from IEPs. This would enable the LLM to tailor materials to match students' learning levels and accommodations, such as adjusted deadlines for students who require extra time.

Participants also emphasized the importance of **integrating the LLM tool with existing Learning Management Systems (LMS)** to allow access to student demographic data, schedules, performance records, and adaptation of outputs between teachers. This would also facilitate communication with students and families regarding academic progress and interventions. They acknowledged the need for data to enhance LLM functionality but raised concerns about privacy, especially regarding sensitive data like student learning needs or disabilities. Authorization and **compliance with data privacy laws** (Ed Law 2d, COPPA, FERPA, HIPAA) was also emphasized. They also discussed challenges in ensuring the LLM's ability to respect privacy during data access and **teacher training to use the tool ethically** and responsibly.

Classroom Integration and Constraints. Teachers also explored the practical considerations for implementing their prototypes in their specific school and classroom contexts. They emphasized needing **reliable infrastructure**, including strong internet and compatible devices. **High-quality PD and ongoing support** were deemed essential for ethical and practical tool use. A continuous feedback mechanism for user experience and feature requests was also highlighted, along with clear expectations and accountability for its implementation, potentially overseen by tech committees or school boards.

However, teachers anticipated several challenges accompanying this integration, including **obtaining approval and managing costs**, navigating budget constraints, and ensuring affordability and accessibility. There were concerns about **resistance to change from educators accustomed to traditional teaching methods**, especially if the tool changes established workflows. **Time constraints** were also noted, as teachers already face significant demands and may struggle to find time for training. **Communicating effectively with parents** and addressing their concerns about LLM tools was deemed crucial, with an emphasis on building trust and educating them on benefits and potential harms.

Ethical Impact of GenAI Prototypes: Benefits and Harms. Participants were enthusiastic about the tool's potential to support teachers by streamlining the creation of projects and rubrics based on educational standards, goals, and **differentiated practices**, thereby improving lesson planning flexibility and fostering equity in the classroom. Additionally, the tool could enhance student learning by making lessons more interactive, personalized, and connected to real-world problems, potentially **increasing motivation and participation**. Participants, however, expressed concerns about the tool potentially reducing teacher and student autonomy through **over-reliance** on LLM-generated recommendations, which could undermine human judgment and diminish valuable teacher-student interactions crucial for trust and understanding. They also worried about the tool **disrupting existing workflows** during its initial

Figure 7: PBL LLM tool supports for brainstorming project ideas

integration and **exacerbating digital inequalities** due to varying student access to reliable internet.

Participants saw the tool as a way to make **PBL more approachable** for teachers, encouraging continued use and innovation by simplifying the process. They also noted its potential to maintain consistency across classes and **align educational materials with state standards**, potentially enhancing outcomes. On the other hand, they were concerned that over-reliance on LLMs could stifle

creativity, resulting in generic responses and reducing originality in classroom activities. They also raised issues about potential **copyright violations**, loss of creative ownership in education, and the invasiveness of the technology, especially in material sourcing. Participants emphasized the need for transparency in LLM tool use, including clear communication with parents, administrators, and other stakeholders. They stressed understanding the LLM’s data sources and algorithms to avoid issues surrounding its **"black box"** nature and suggested incorporating warnings into interfaces to **verify LLM-generated content** before usage.

4.3 PBL GenAI Tool Wireframe Prototypes and Testing (all Category themes)

From these discussions, we created wireframes for a teacher dashboard spanning the three categories of support (curriculum, assessments, and progress tracking). These wireframes explicitly link their features to the core PBL challenges identified (from category B themes), and translate the identified needs (from category A themes) into actionable LLM-PBL solutions (in category C themes). **Below, we present these wireframes and the combined feedback received from teachers in both Study 1 (section 3.1.1) and Study 2 (section 3.2.1). By gathering perspectives from teachers with varying levels of PBL expertise, we wanted to ensure adaptability of LLM tool wireframes across diverse teaching methods and environments (addressing RO2).** For brevity, we use the term CAIL (Collaborative AI for Learning) for the system and will refer to it as such throughout this section.

4.3.1 Curriculum Supports. We focused on two key needs: brainstorming new project ideas with CAIL and co-implementing weekly or unit lesson plans based on these finalized ideas.

The project brainstorming wireframes (Figure 7) allow educators to expand on initial ideas gathered from classroom experiences, augmenting teacher creativity from the outset. Teachers input their learning goals and standards, ensuring **alignment with curricular objectives**. The tool integrates a curated list of U.S. national and state standards and allows teachers to **define expected project outputs** (e.g., slide decks, podcasts, portfolios), group size, project duration, and upload preferred lesson planning templates. CAIL offers multiple LLM-generated ideas, enabling teachers to refine and expand on them and provide students with **varied project options**.

Our participants suggested adding adjustable project durations, with the ability to **accommodate different class schedules**. Live collaboration features were also requested, enabling interdisciplinary teamwork among educators for formulating ideas. Teachers expressed interest in uploading custom standards and explored the idea of the LLM crosswalking standards to identify and address gaps. Combining multiple LLM-generated project ideas into a **choice board** for students was highlighted as a potential feature that could enhance differentiation.

Teachers can then use CAIL for lesson planning (Figure 8) to implement selected project ideas. This includes creating formative assessments ("knowledge checks"), exit tickets, hooks, and scaffolded activities like games, discussions, peer learning, and problem-solving tasks. CAIL offers teachers a **structured lesson planning template** aligned with the standards they selected for

Week 1 for Project 1

Lesson Plan for week 1

Return to Project 1 Dashboard

NGSS Lesson Planning Template

Grade/Grade Band: Topic: Lesson # _____ in a series of _____ lessons

Brief Lesson Description:

Performance Expectation(s):

Specific Learning Outcomes:

Narrative / Background Information:

Prior Student Knowledge:

Science & Engineering Practices:

Disciplinary Core Ideas:

Crosscutting Concepts:

Possible Preconceptions/Misconceptions:

Select different template

Preview/Edit Lesson Template

CAIL has filled out parts of this template with the info you provided. You can expand on it further.

Create Rubrics for Learning Activities

Submit to CAIL

Project milestone for week 1

Project Overview

Project Name: Project Manager:

Compiled By/Date: Executive Sponsor:

Business Case:

[Summarize project and value of initiating it for the organization.]

Opportunity/Problem:

[Explain what the opportunity that the project will exploit or the problem it will solve.]

Preview/Edit Project Milestone

Create Rubrics for Week 1 Milestone

Submit to CAIL

Lesson Planning: Week 1 for Project 1

Lesson Plan Document for week 1: Review and edit

Return to Week 1

NGSS Lesson Planning Template

Grade/Grade Band: 5-6 grade Topic: MS Structure and Properties of Matter Lesson # _____ in a series of _____ lessons

Brief Lesson Description:

Performance Expectation(s):

Specific Learning Outcomes: Develop models to describe the atomic composition of simple molecules and extended structures.

Narrative / Background Information:

Prior Student Knowledge:

Science & Engineering Practices:

Disciplinary Core Ideas:

Crosscutting Concepts:

Possible Preconceptions/Misconceptions:

LESSON PLAN - 5-6 Model

ENGAGE: Opening Activity - Assess Prior Learning / Stimulate Interest / Generate Questions:

EXPLORE: Lesson Description - Materials Needed / Probing or Clarifying Questions:

EXPLAIN: Concepts Explained and Vocabulary Defined:

Vocabulary:

ELABORATE: Applications and Extensions:

EVALUATE:

Formative Monitoring (Questioning / Discussion):

Summative Assessment (Quiz / Project / Report):

Elaborate Further / Reflect: Enrichment:

To be filled

To be filled

To be filled

To be filled

Generate ideas for warm up activities

Generate ideas for hook activities

Generate ideas for guiding questions

Brainstorm ideas for scaffolding concepts

Brainstorm activities for knowledge checks

Generate exit tickets

Lesson Planning: Week 1 for Project 1

Lesson Plan Document for week 1: Review and edit

Return to Week 1

Generate exit tickets

What topics/concepts should this week's exit tickets cover?

Would you like to include any self or peer reflection questions? What would you like to check for?

Select question types you would like to include

Generate questions

Q1: How would you define xyz?

Q2: What is one thing that went well in today's lesson?

Q3: What is one thing you struggled with?

Q4: How did your group activity go today?

Add to document

Project Milestone: Week 1 for Project 1

Project milestone for week 1

Return to Week 1

What would you like to cover this week for the project?

Generate milestone document

Project Overview

Project Name: Project Manager:

Compiled By/Date: Executive Sponsor:

Business Case:

[Summarize project and value of initiating it for the organization.]

Opportunity/Problem:

[Explain what the opportunity that the project will exploit or the problem it will solve.]

Goals/Objectives:

[What is the business value to be gained and how the business environment will be improved?]

In/Out of Scope:

[What is within the scope of the project and what is outside of the project's scope?]

Preview & Edit Milestone Description Document

Personalize for project groups

Preview personalized document

Additional Resources

Resource 1

Resource 2

Resource 3

Add to document

Deliverable checklist

Item 1 due MM/DD

Item 2 due MM/DD

Item 3 due MM/DD

Add to document

Upload milestone for CAIL

Figure 8: PBL LLM tool supports for lesson planning

their project ideas. For this wireframe, we picked an NGSS (Next Generation Science Standards) example template that guides teachers through performance expectations, learning outcomes, and reflection elements. It provides targeted lesson plan suggestions, reducing planning time and offering guidance for both new and experienced PBL teachers. It maintains **institutional memory** by gathering data to inform future practices. Additionally, CAIL

assists teachers in **creating milestone documents** and tailoring resources, such as checklists and schedules, to guide student groups through project phases. Participants recognized the tool's long-term time-saving potential:

Create Rubrics

Rubric Generator

Select knowledge check/
Project milestone:

Select from dropdown menu

Upload knowledge check
instructions PDF

List of artifacts for this knowledge check/milestone

1. Item 1: PowerPoint Photo Gallery

2. Item 2: Pdf of Essay Outline

3. Item 3: Pdf of Brainstorm Sketch

Pre-populate rubric competencies/items below

Pre-populated competencies (rows of the rubric) including interpersonal skills. Edit competencies and describe an acceptable standard for each item

Add/edit competency to measure

Develop models to describe the atomic composition of simple molecules and extended structures

Gather and make sense of information to describe that synthetic materials come from natural resources and impact society.

Develop a model that predicts and describes changes in particle motion, temperature, and state of a pure substance when thermal energy is added or removed

Enter acceptable standard for competency

Select rubric type:

1-point

3-point

4-point

5-point

Generate

Preview student-facing rubric

Differentiate

Differentiation for Rubrics

Let's create multiple versions of this rubric

Original rubric preview:

	Below Standard	Approaching Standard	At Standard	Above Standard
Explanation of Ideas & Information	• does not present information, arguments, data, or findings clearly, concisely, and logically; evidence cannot follow the line of reasoning	• presents information, findings, arguments and supporting evidence in a way that is not clearly or concisely	• presents information, findings, arguments and supporting evidence clearly, concisely, and logically; evidence can follow the line of reasoning	• presents information, findings, arguments and supporting evidence clearly, concisely, and logically; evidence can follow the line of reasoning
Organization	• does not meet requirements for what should be included in the presentation	• meets most requirements for what should be included in the presentation	• meets all requirements for what should be included in the presentation	• meets all requirements for what should be included in the presentation

Do any of your students have IEP/504 requirements?

Select students

Enter IEP requirement

Generate differentiated rubric

Would you like to modify (add/remove/edit) any competencies for some students or groups?

Select students/groups

Modify competencies

Generate differentiated rubric

Would you like to personalize the rubric for every project group?

Generate personalized rubric

Preview & Edit for every group

Approve

Return to rubric creation

View all versions of the rubric

Differentiation for Rubrics

Let's create multiple versions of this rubric

Modify competencies for differentiated rubric

This allows you to modify rubrics for some students or groups if they are behind or way ahead of the curve. To keep things fair, teachers can only modify the rows/competencies/items measured for every student – not the columns (acceptable standards for each item are same across all students)

Add/edit competency to measure	Below standard	Approaching standard	At standard	Above standard?
Develop models to describe the atomic composition of simple molecules and extended structures	View only	View only	View only	
Gather and make sense of information to describe that synthetic materials come from natural resources and impact society.	View only	View only	View only	
Develop a model that predicts and describes changes in particle motion, temperature, and state of a pure substance when thermal energy is added or removed	View only	View only	View only	

Save rubric

Would you like to personalize the rubric for every project group?

Generate personalized rubric

Preview & Edit for every group

Approve

Return to rubric creation

All rubric versions at a glance

View and approve all versions of the rubric you created

Original rubric

View/edit selected students

Edit rubric

IEP/504 requirements rubric

View/edit selected students

Edit rubric

Modified competencies rubric

View/edit selected students/groups

Edit rubric

Preview personalized group rubrics (if any)

Return to rubric creation

Approve all rubrics

Figure 9: PBL LLM tool supports for rubric creation and differentiation

“These are the ‘money’ tools for me. These buttons that help with generating these smaller activities that support the bigger project would be a huge time and brain saver.” –P01

However, they stressed the need for **thorough PD and microlearning modules** during rollout to help educators fully understand the tool’s benefits, suggesting tutorials no longer than

20 minutes to avoid overwhelming them. Participants highlighted the importance of clear learning outcomes for each activity, aligning with competency-based teaching, and appreciated the iterative refinement of lesson plans. They also suggested that LLMs estimate activity time requirements while acknowledging their current limitations in precise time predictions.

Figure 10: PBL LLM tool supports for grading

4.3.2 Assessment Supports. The assessment support section of the wireframes received mixed reactions from teachers, as discussed below.

Teachers can use CAIL to **develop rubrics** (Figure 9) for formative assignments and project milestones. They can input specific student artifacts for assessment and automatically populate rubric rows with competencies from the selected curriculum standards. Educators can further refine the criteria by editing acceptable standards for each rubric item, and **selecting a rubric type** (e.g., 1-point, 3-point, 4-point, or 5-point) to suit their classroom. The tool also supports differentiation by generating multiple rubric versions tailored to individual student needs. Teachers can input **IEP/504 requirements** and pick competencies for specific students and groups, ensuring fair assessment by maintaining consistent standards while accommodating individual learning trajectories.

Participants suggested including examples to explain the different rubric types to help new teachers, **student work samples to set expectations**, and using visual highlights on rubrics to make feedback more accessible, noting that students often overlooked detailed feedback. Participants raised **privacy concerns** about using sensitive student IEP data, suggesting anonymized profiles and LMS integration for secure access. Participants emphasized the need for both **teacher- and student-facing rubrics**, with LLM assistance in simplifying language for students and using "I can" (e.g., "I can analyze data from a chart to make informed conclusions.") statements, to make these more empowering.

Figure 10 shows the CAIL grading page, which simplifies evaluating formative assessments and project milestones by tracking submissions, generating response summaries, and sending reminders for pending work. It allows teachers to upload scored examples as benchmarks, helping LLMs contextualize assessments. CAIL maps

rubrics to relevant tasks, with LLM-generated scores complemented by required teacher feedback. An optional "live scoring" feature provides instant feedback to students if enabled by teachers.

Teachers were cautious about relying solely on LLMs for grading, particularly for subjective or creative work, emphasizing the **need for human oversight** to ensure feedback is personalized. Concerns were raised about biases in LLM grading, as well as its ability to handle atypical responses and physical artifacts that require nuanced understanding. Some participants like P05 (who taught PBL in online settings) did not want grading to be handled by the LLM at all:

"I understand why people want auto grading...but I think you learn so much about the individual students by interacting with their work. How else are you interacting with them? By seeing what they did and reading it. I wouldn't feel good about having AI grade it." – P05

Some teachers suggested using live scoring features to boost engagement and prompt revisions, **minimizing grading iterations**. However, others preferred the option to review and adjust LLM-generated comments before students received them and supported integrating custom feedback based on **classroom observations to complement LLM scores**.

4.3.3 Progress Tracking Supports. The progress tracking section of the wireframes (Figure 11) provides educators with an overview of student performance across individual, group, and classroom levels. The **Exit Tickets Summary** highlights topics students enjoy, areas needing help, and interests for future exploration, enabling targeted interventions; the **Knowledge Check Performance Summary** displays **class strengths and areas of concern** on competencies, allowing follow-up with struggling students; and the **Groups Project and Contributions** Tracking monitors group members' progress and individual contributions for effective interventions.

Participants recommended making progress data accessible to students individually and anonymously at the class level to enhance motivation and self-awareness. They emphasized **cross-disciplinary tracking for a holistic view of performance** and coordinated support across subjects. Although teachers found the progress data valuable for differentiation and scaffolding, they were concerned **about administrators using the data to evaluate their own performance**. They suggested creating an opt-in database to share projects and best practices, searchable using LLMs, based on student progress.

5 DISCUSSION

5.1 Tensions in Designing PBL LLM Tools

Our findings suggest that LLMs hold significant potential to enhance PBL practices by alleviating several teacher concerns. However, their use involves trade-offs that must be carefully managed for long-term effectiveness and sustainable use in the classroom. We present three tensions critical for the success of PBL-LLM tools: (1) technology choice and feasibility, (2) teacher agency, creativity, PD, and learning, and (3) balancing educational depth with reducing administrative workload.

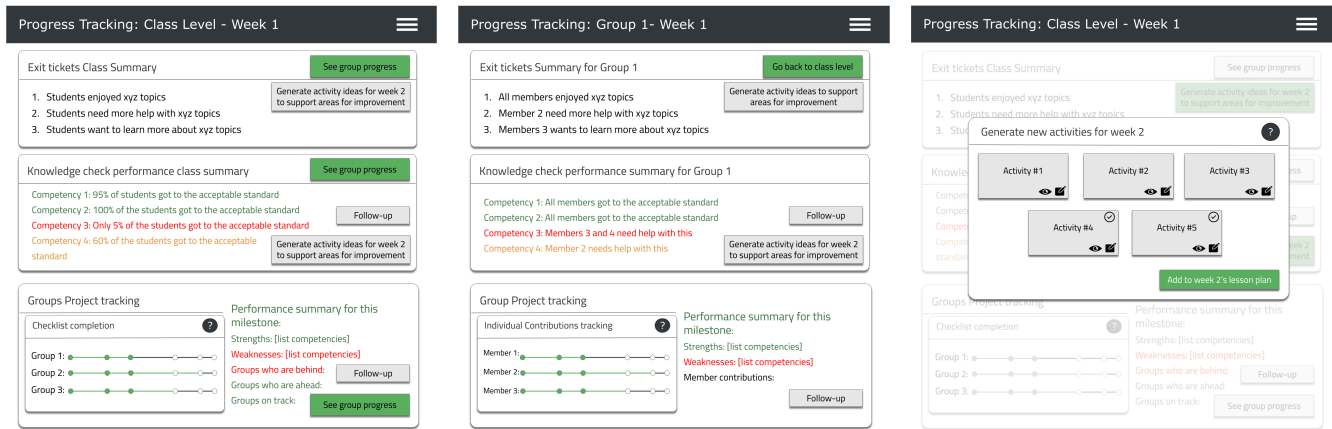


Figure 11: PBL LLM tool supports for progress tracking

5.1.1 Technology Choice and Feasibility: The selective use of LLMs is crucial to their effectiveness and sustainability in PBL. While LLMs can offer powerful capabilities in generating educational content and personalized learning materials [74, 93, 141, 147], teachers in our co-design workshops raised issues related to the accuracy and bias inherent in LLM-generated content, as seen in prior work [93, 114, 144]. LLMs could also undermine the fairness and inclusivity of educational practices due to this bias [43, 109, 114]. Our teachers expressed doubts about the reliability of LLMs in handling tasks that require a deep understanding of context or the subtle nuances involved in student expressions and creative multimodal artifacts [109, 114]. This skepticism reflects broader concerns about the limits of GenAI's interpretive abilities and risks associated with delegating too much of the evaluative process to technology [109, 147].

The feasibility of implementing LLMs in diverse educational settings was another significant issue. Our study highlighted concerns around logistical constraints, such as limited access to high-speed internet, adequate digital infrastructure, and computational resources, which can vary greatly across schools and districts [144]. These disparities raise concerns about digital equity, as schools with fewer resources may find themselves unable to leverage advanced LLM tools, potentially widening the gap in educational opportunities [74, 93]. Despite these challenges, teachers in our study were open to strategic and scoped integration of PBL-LLM tools by focusing on PBL-specific tasks best suited for LLM optimization to ensure their effectiveness in diverse contexts.

5.1.2 Teacher Agency, Creativity, Professional Development, and Learning: Maintaining teacher agency and fostering creativity when integrating LLM tools were recurring areas of discussion in our work. Teachers expressed a strong preference for tools that align with their pedagogical goals, support creativity, and promote critical thinking among students [74]. However, they also voiced concerns that excessive automation could diminish their role [33], particularly for novice PBL educators still learning the nuances of implementing high-quality PBL pedagogy.

To address these concerns, our findings presented opportunities for LLM tools to offer teachers real-time guidance during resource

creation while allowing for their active input. Rather than replacing teachers, the LLM could act as a co-educator and brainstorming partner, supporting their ability to design and implement activities [87]. Our wireframes suggest opportunities for integrating LLM-driven PD modules into PBL tools, helping teachers build new skills and strategies. Drawing on perspectives of teachers in our study, this training could also focus on helping PBL educators understand the tool's long-term benefits and customization options, ensuring it integrates smoothly into existing workflows. The in-tool LLM support could serve as an additional learning resource for novice PBL teachers by providing scaffolded guidance during lesson planning, aiding them in developing the skills needed to design and implement effective PBL lessons. Teachers in our study were particularly enthusiastic about the prospect of learning prompt engineering and other LLM-specific skills, seeing these as new ways to enhance their creativity and adapt technology to their classroom needs [120]. To realize these benefits, the design of LLM tools must involve ongoing collaboration among developers, teachers, and PD designers.

5.1.3 Balancing Educational Depth with Reducing Administrative Workload: LLM tools have the potential to significantly reduce administrative burdens in PBL settings, such as managing group projects, tracking student progress, and providing individualized feedback [25] that can free up teachers to focus on more meaningful instructional activities [111]. However, our results caution against allowing efficiency gains to oversimplify the complex aspects of PBL, such as developing students' interpersonal skills, cultivating creativity, and supporting critical thinking—areas that require nuanced, human-centered guidance [74]. Similar to previous work, our teachers also feared that excessive automation could diminish opportunities for spontaneous, in-the-moment learning experiences and the development of soft skills that are crucial for student growth [28, 59]. There is thus a pressing need to identify which administrative tasks can be effectively automated without compromising the integrity of the learning experience.

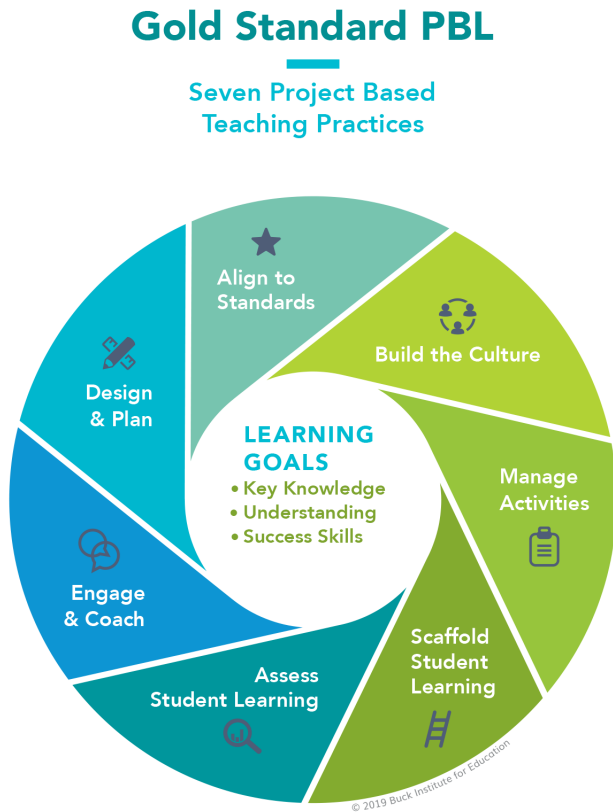


Figure 12: Gold Standard Framework: Project Based Teaching Practices

5.2 LLM Design Recommendations for Project-Based Learning

Keeping the above tensions in mind, we present design recommendations for educational technology designers and educators when integrating LLMs into PBL (addressing RO3). We draw on the “**Gold Standard PBL: Project Based Teaching Practices**” framework from the Buck Institute for Education that guides teachers in implementing high-quality PBL [96]. We briefly describe the framework’s seven PBL teaching practices (Figure 12) and unpack potential ways LLM systems can mitigate implementation barriers. **We focus on recommendations that alleviate teachers’ administrative workload in implementing high-quality PBL, allowing them to dedicate more time to the creative and fulfilling aspects of teaching.** We advocate for these discussions to actively include new teachers, curriculum coaches, school boards, admins, information technology (IT) and special education committees — stakeholders our teachers emphasized as vital but often underrepresented in education tool design [100]. We structure the two sub-sections below by mapping a combination of these practices to the wireframe supports we developed and the design recommendations they informed.

5.2.1 Recommendations for Project Ideation, Curriculum Planning, and Management. -

Gold Standard PBL Teaching Practice #1: Design and Plan: “Teachers create or adapt a project for their context and students, and plan its implementation from launch to culmination while allowing for some degree of student voice and choice.”

Gold Standard PBL Teaching Practice #2: Align to Standards: “Teachers use standards to plan the project and make sure it addresses key knowledge and understanding from subject areas to be included.”

Gold Standard PBL Teaching Practice #3: Manage Activities: “Teachers work with students to organize tasks and schedules, set checkpoints and deadlines, find and use resources, create products and make them public.”

Gold Standard PBL Teaching Practice #4: Scaffold Student Learning: “Teachers employ a variety of lessons, tools, and instructional strategies to support all students in reaching project goals.”

Based on the feedback we received on project ideation and curriculum planning supports, we outline the following design considerations for Practices 1, 2, 3, and 4:

Recommendation 1: Augment Teacher Creativity with LLM Tools

Design LLM tools to augment and elevate teacher creativity by empowering educators to lead the project brainstorming process from the outset while leveraging LLMs to support mundane, peripheral tasks. Build on established pedagogical practices of PBL educators, reducing the need for creating resources from scratch.

Our results showed evidence of PBL teachers grappling with balancing the depth of knowledge required for specific projects with the breadth of content mandated by the curriculum. Teachers also raised concerns about the excessive contextual information required by current LLM systems and their ability to effectively prompt the LLM for specific tasks. A proposed solution is structured input: teachers specify contextual details such as specific learning goals, standards, desired final project artifacts, and duration of implementation, and LLM tools generate ideas aligned with curricular objectives and tailored to classroom needs. This structured input approach not only supports experienced educators but also provides novice PBL teachers with foundational knowledge for constructing high-quality PBL units.

Recommendation 2: Provide Teacher-Directed Scaffolding with Flexibility

Design LLM tools for PBL that intentionally prioritize and mandate core teacher inputs, while offering optional customization features.

This recommendation is rooted in two ideas from our research: (1) the need to promote high-quality pedagogy by ensuring that key LLM functionalities are integrated into teaching practices, while also providing flexibility for educators to choose supplementary features that support their unique classroom needs, (2) ensuring the LLM system has sufficient contextual information from the required inputs to generate high-quality resources. For example, requiring teachers to input or select educational standards for their project ideas should be mandatory, ensuring alignment with curriculum

requirements and learning outcomes [71]. In contrast, optional features can accommodate varying school contexts, as our results showed that teachers sometimes chose not to use extended features depending on their specific needs.

Recommendation 3: Ensure Flexibility in Scheduling for Diverse Classroom Needs

Design LLM tools to accommodate specific temporal structures of diverse educational environments.

This includes accounting for various class schedules, including block scheduling (a method of dividing time into distinct blocks and assigning each block to a specific task or activity) [27, 112]. Since methods for chunking projects varied vastly across teachers of different grade levels and disciplines in our workshops' data [66], features to parameterize project duration based on hours, days, or weeks are necessary to provide the flexibility teachers require when brainstorming ideas for projects.

Recommendation 4: Promote Differentiation with LLM-Generated Project Ideas

Design LLM tools to generate multiple project ideas that teachers can combine into choice boards, thereby enhancing differentiation and student agency.

Implementing choice boards with multiple project options was a common practice among our teachers (Figure 13), but proved challenging due to time constraints and difficulties with inventing novel, creative ideas each time. To address this, the tool should allow for continuous modification and regeneration of LLM-generated content, enabling teachers to refine and align projects with their educational goals and classroom needs.

Recommendation 5: Streamline Lesson Planning with LLM-powered Standards Templates

Embed LLM supports within existing lesson planning templates from common educational standards to streamline the process of generating engaging instructional content.

This can reduce the time and cognitive load for teachers when aligning PBL activities with academic standards. Consistent with other scholars, our research revealed that teachers often struggled with the complexity of designing knowledge-check activities, driving questions, and other project components while ensuring alignment with standards [41, 112]. The LLM system could provide immediate, context-sensitive suggestions and resources, tailored to specific lesson goals and standards, allowing teachers to focus on customizing their lesson plans rather than building them from scratch. This feature is especially beneficial for less experienced teachers, easing their adoption of PBL principles, while also providing experienced educators with tools to document and share their expertise.

5.2.2 Recommendations for Progress Tracking and Assessments. -

Gold Standard PBL Teaching Practice #5: Build the Culture:
“Teachers explicitly and implicitly promote student independence and growth, open-ended inquiry, team spirit, and attention to quality.”

Gold Standard PBL Teaching Practice #6: Assess Student Learning:
“Teachers use formative and summative assessments of knowledge, understanding, and success skills, and include self and peer assessment of team and individual work.”

Gold Standard PBL Teaching Practice #7: Engage and Coach:
“Teachers engage students in their learning and work alongside them to identify when they need skill-building, redirection, encouragement, and celebration.”

Based on the feedback we received for our assessments and progress tracking supports, we outline the following design considerations for Practices 5, 6, and 7:

Recommendation 6: Develop Equitable and Actionable Rubrics

Incorporate a default rubric format within the LLM tool that promotes fair and equitable grading practices, aligning with high-quality PBL pedagogy.

Our research data revealed the importance of having educators collaborate with LLMs to set actionable and easily understandable expectations within rubrics that meet required competencies. The LLM tool should also facilitate the creation of two separate rubric sets: one for teachers and administrators, focusing on competencies from established standards, and another for students, emphasizing clear, actionable expectations tied to project outputs.

- **Recommendation 6.1: Implement single-point rubrics as the default rubric type within LLM tools for PBL.** Single-point rubrics use a single column for feedback instead of multiple performance levels. Research shows that such rubrics are quicker and easier for teachers to create since they don't need to anticipate all possible ways students might “fail expectations” [48]. For students, these rubrics are simpler to understand, focusing on clear target expectations. They also promote higher-quality feedback, as teachers highlight specific strengths and areas for improvement, fostering a growth mindset and iterative learning.
- **Recommendation 6.2: Augment rubric items with detailed teacher feedback.** Rubric items must be augmented with detailed feedback from teachers. Designers of the LLM tools should explore the possibility of providing a starting point for feedback from student submissions, allowing teachers to elaborate and personalize their notes [51, 125].
- **Recommendation 6.3: Integrate LLM-assisted positive feedback to complement constructive criticism alongside rubrics.** Our teachers mentioned struggling to provide granular positive feedback due to time constraints. Praise and celebration, bolstered by relevant specifics, can instill student confidence and promote balanced assessments.

Recommendation 7: Ensure Privacy in LLM-Driven Differentiation

Ensure that data processing and LLM interactions for differentiating student resources occur locally on the teacher's device or within a secure school network.

Teachers in our study expressed significant concerns about privacy when using LLMs to differentiate rubrics for students with

ANCIENT CIVILIZATIONS PROJECT MENU

Choose one of the projects from the table below to inform readers about the legacy of your ancient civilization (Mesopotamia, Ancient Egypt, or the Israelite Kingdoms). Be sure to refer to the rubric as you plan your project.

Book Creator Use Book Creator to create a book for younger students highlighting the features of your civilization. CODE: 922V9WJ23	Green Screen Write a script for a news broadcast to be recorded in front of a green screen. Choose related images to enhance your script.	Podcast Write and record a podcast about your civilization. You will need to include all of the topics outlined on your rubric.
Google Drawings Informational Poster Create a poster with all the necessary information to showcase your civilization.	Front Page of a Newspaper A variety of tools can be used to create the front page of a newspaper from ancient times. You will be able to address all of the required areas using headlines, articles, and advertisements related to your civilization.	

COLONIAL ADVERTISEMENT PROJECT MENU

Choose one of the projects from the table below to inform future settlers about your colony. Be sure to refer to the rubric as you plan your project.

Book Creator Use Book Creator to create a book for younger students highlighting the features of your colony.	Google Drawings Informational Poster Create a poster with all the necessary information to showcase your colony.	Podcast Write and record a podcast about your colony. You will need to include all of the topics outlined on your rubric.	Green Screen Write a script for a news broadcast to be recorded in front of a green screen. Choose related images to enhance your script.
Play Love to write? This is the one for you! Write an original play using the information you've learned about your colony. You will have the opportunity to share your play through a live or recorded performance.	Journal Entries Travel back in time and take on the role of an original colonial settler. Write a series of journal entries to introduce future settlers to your colony. A variety of tools can be used for this including, Book Creator, Google Slides or Doc, and more!	Interview Interview one of the notable people from your colony. Your interview will be written and recorded.	Front Page of a Newspaper A variety of tools can be used to create the front page of a newspaper from colonial times. You will be able to address all of the required areas using headlines, articles, and advertisements related to your colony.

Reading Response Choice Board

Learning Objective	 Create a Vocabroo with your book recommendation. Include at least 3 reasons why you do or do not recommend the book using evidence from the text. Post the link here:	 Create a Powtoon retelling the story from another character's point of view. Tutorial Post the link here:	 Create a timeline of at least 6 major events from your story using Google Drawings Post the link here:
 Use Read and Write for Google Extension to highlight 3 unfamiliar or interesting words. Use the extension to create a vocabulary Google Doc and write each word in a new sentence. Post the link here:	Start Here <ul style="list-style-type: none">Read the book in your Storia account.Complete 3 more activities. (Fill in with green point)Turn in your work to Google Classroom when finished.  Post the link here:		 Create a 2 minute Book Trailer to get other students interested in your book using iMovie. Upload video into your Google Drive. Post the link here:
 Create a Google Form with 3 important questions you wonder about from the story. Your classmates will respond later. Post the link here:	 Complete a Google Sheet listing the main character's internal and external character traits. Provide 2 pieces of evidence for each trait. Post the link here:	 Create a Google Slideshow with 4 connections from the story. Text-Text Text-Self Text-Movie Text-World Post the link here:	

Created by @joliboucher www.flippedtechcoaching.com

Figure 13: Examples of choice boards used by our teachers to promote student agency

IEPs, particularly regarding the sharing of sensitive student information. To address these concerns, LLM tools should either operate locally or be integrated securely within existing Learning Management Systems (LMS) as an app or extension [140, 144]. This approach ensures alignment with privacy regulations such as COPPA (Children’s Online Privacy Protection Rule). Pseudonyms can also automatically replace identifying student information before being used for differentiation.

Recommendation 8: Integrate Project Management Scaffolds and Supports

- **Recommendation 8.1: PBL LLM tools should include student-facing project management features that complement teacher supports.** Teachers in our study found balancing open-ended problem-solving with personalized guidance challenging, especially in managing group projects. The LLM can serve as a brainstorming partner and offer scaffolding to guide students, providing support when teachers are unavailable [26, 64, 65, 65, 118, 126]. LLM conversational agents could potentially capture key discussion points and synthesize them for review, acting as self-check tools to ensure groups complete daily tasks and take appropriate next steps [22, 74, 127]. Additionally, these agents can analyze

and monitor participation frequency, and assess the quality of contributions to ensure equitable participation within groups [92, 117, 126].

- **Recommendation 8.2: Design LLM tools that assist teachers in providing personalized feedback during active class time.** By alerting teachers to groups needing immediate attention and tracking group progress, LLMs can help prioritize the teacher’s time and provide notes for asynchronous feedback [38, 63]. This support allows teachers to efficiently manage multiple groups without overwhelming their workload, ensuring that all student groups receive necessary guidance.
- **Recommendation 8.3: Implement robust LLM-powered portfolio supports that enable students to submit evidence of their processes and reflections, documenting their learning journey from ideation to project completion.** Prior research has underscored the importance of processes (*how* students did things throughout the project, from ideation to prototyping, for example, including challenges faced and *what* they did about them), and reflection (growth or trajectory over time) [44]. Our teachers used Padlets for students to submit their daily work (e.g., snapshots, videos, responses) and as a reflective tool for exit tickets (Figure 14). LLMs can support the creation of individual portfolios to showcase personal skills and group portfolios to document

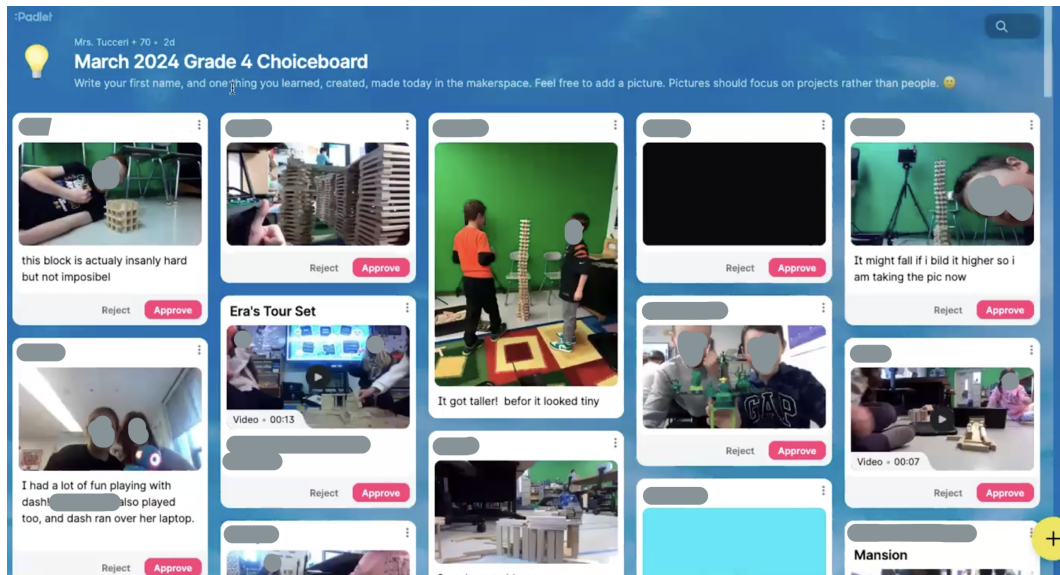


Figure 14: Examples of Padlets used by our teachers to document student progress and self-reflections

project development [76]. Additionally, LLMs should incorporate self and peer reflection activities, as well as conflict resolution tools, drawn from toolkits like “Beyond Rubrics”, to ensure meaningful, ongoing assessment [66, 107]. By embedding these elements into portfolios, LLM can facilitate both formative and summative assessments, offering a comprehensive view of student learning.

- Recommendation 8.4: Design LLM tools to support rather than replace teacher involvement in the grading process by organizing and presenting data that tracks student progress.** Our study revealed that teachers highly value their agency in the grading process as it enables them to deeply understand student thinking, provide personalized feedback, and make informed decisions. Teachers also expressed concerns about the quality of LLM-generated feedback and the subjectivity involved in grading PBL student multimodal artifacts. By organizing data on milestones, monitoring group performance, and curating exit ticket responses, LLM tools can alleviate administrative burdens and provide a structured starting point for teachers. This ensures that teachers remain at the center of the grading process, using LLMs as a means to enhance their ability to deliver nuanced and targeted feedback, ultimately leading to more effective and personalized student assessments.

6 LIMITATIONS AND FUTURE WORK

Our study has limitations that suggest avenues for future research. First, our focus on PBL in the context of the United States limits the generalizability of the findings; future research should explore the applicability of LLM-driven PBL tools in culturally diverse educational settings globally. We also acknowledge the small sample size and absence of collected racial demographics information as limitations for generalizability. Additionally, our study involved

a self-selected group of teachers who were likely more interested in GenAI and PBL, potentially biasing the findings toward a more positive outlook. Future work could include longer-term studies that investigate how such tools might engage teachers who are more resistant to PBL or GenAI integration. The wireframes developed in this study are only a starting point; building and testing a working product through iterative design cycles is crucial to evaluate the feasibility of GenAI for the proposed features. Future research should focus on co-designing student-facing supports that align with teacher tools, fostering an integrated flow of information and feedback to create a balanced ecosystem within the LLM-driven PBL experience.

Acknowledgments

We would like to thank Robert Parks and Raechel Walker for their invaluable assistance with the workshops. We also sincerely appreciate pivotal input received from Christina Bosch during key stages of the study. Finally, this study would have been impossible without the 41 remarkable educators, all deeply passionate about project-based learning, who brought their enthusiasm to every step of the process.

References

- [1] David Adamson, Gregory Dyke, Hyeju Jang, and Carolyn Penstein Rosé. 2014. Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education* 24 (2014), 92–124.
- [2] L Aitken. 2019. Teacher perspectives on using project-based learning. https://www.teachermagazine.com/au_en/articles/teacher-perspectives-on-using-project-based-learning.
- [3] Maija Aksela, Outi Haatainen, et al. 2019. Project-based learning (PBL) in practise: Active teachers’ views of its’ advantages and challenges. *Integrated Education for the Real World* 34 (2019).
- [4] Ashraf Alam. 2023. Intelligence unleashed: An argument for AI-enabled learning ecologies with real world examples of today and a peek into the future. In *AIP Conference Proceedings*, Vol. 2717. AIP Publishing.

- [5] Mark A Albanese and Georgia L Hinman. 2019. Types and Design of Assessment in PBL. *The Wiley Handbook of Problem-Based Learning* (2019), 389–409.
- [6] Riordan Alfredo, Vanessa Echeverria, Yueqiao Jin, Lixiang Yan, Zachari Swiecki, Dragan Gašević, and Roberto Martínez-Maldonado. 2024. Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence* (2024), 100215.
- [7] Safinah Ali, Vishesh Kumar, and Cynthia Breazeal. 2023. AI audit: a card game to reflect on everyday AI systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 15981–15989.
- [8] Matthew Arnold, Rachel KE Bellamy, Michael Hind, Stephanie Houde, Sameep Mehta, Aleksandra Mojsilović, Ravi Nair, K Natesan Ramamurthy, Alexandra Olteanu, David Piorkowski, et al. 2019. FactSheets: Increasing trust in AI services through supplier's declarations of conformity. *IBM Journal of Research and Development* 63, 4/5 (2019), 6–1.
- [9] Nursultan Askarbekuly and Nenad Aničić. 2024. LLM examiner: automating assessment in informal self-directed e-learning using ChatGPT. *Knowledge and Information Systems* (2024), 1–18.
- [10] Sinem Aslan and Charles M Reigeluth. 2015. Examining the challenges of learner-centered education. *Phi delta kappan* 97, 4 (2015), 63–68.
- [11] Andi Asrifan et al. 2024. Integrating AI in Project-Based Learning for Differentiated English Language Instruction: A Scoping Review. *Journal of English Education and Teaching* 8, 3 (2024), 586–608.
- [12] Theophilus Azungah. 2018. Qualitative research: deductive and inductive approaches to data analysis. *Qualitative research journal* 18, 4 (2018), 383–400.
- [13] Marie AB Bakah, Joke M Voogt, and Jules M Pieters. 2012. Updating polytechnic teachers' knowledge and skills through teacher design teams in Ghana. *Professional development in education* 38, 1 (2012), 7–24.
- [14] Brigid Barron and Linda Darling-Hammond. 2008. Teaching for Meaningful Learning: A Review of Research on Inquiry-Based and Cooperative Learning. Book Excerpt. *George Lucas Educational Foundation* (2008).
- [15] William N Bender. 2012. *Project-based learning: Differentiating instruction for the 21st century*. Corwin Press.
- [16] Phyllis C Blumenfeld, Elliot Soloway, Ronald W Marx, Joseph S Krajcik, Mark Guzdial, and Annemarie Palincsar. 1991. Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational psychologist* 26, 3-4 (1991), 369–398.
- [17] Jo Boaler. 1998. Open and closed mathematics: Student experiences and understandings. *Journal for research in mathematics education* 29, 1 (1998), 41–62.
- [18] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.
- [19] Jonathan Brinkerhoff and Krista Glazewski. 2004. Support of expert and novice teachers within a technology enhanced problem-based learning unit: A case study. *International Journal of Learning Technology* 1, 2 (2004), 219–230.
- [20] Ann L Brown. 1992. Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The journal of the learning sciences* 2, 2 (1992), 141–178.
- [21] Bill Buxton. 2010. *Sketching user experiences: getting the design right and the right design*. Morgan kaufmann.
- [22] Zhenyao Cai, Seehee Park, Nia Nixon, and Shayan Doroudi. 2024. Advancing Knowledge Together: Integrating Large Language Model-based Conversational AI in Small Group Collaborative Learning. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–9.
- [23] Kelly Caine. 2016. Local standards for sample size at CHI. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 981–992.
- [24] Ling Cen, Dymitr Ruta, Leigh Powell, Benjamin Hirsch, and Jason Ng. 2016. Quantitative approach to collaborative learning: Performance prediction, individual assessment, and group composition. *International Journal of Computer-Supported Collaborative Learning* 11 (2016), 187–225.
- [25] Cecilia Ka Yuk Chan and Katherine KW Lee. 2023. The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and millennial generation teachers? *Smart learning environments* 10, 1 (2023), 60.
- [26] Catherine C Chase, Doris B Chin, Marily A Oppezzo, and Daniel L Schwartz. 2009. Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of science education and technology* 18 (2009), 334–352.
- [27] Cheng-Huan Chen and Yong-Cih Yang. 2019. Revisiting the effects of project-based learning on students' academic achievement: A meta-analysis investigating moderators. *Educational Research Review* 26 (2019), 71–81.
- [28] Jennifer J Chen and Jasmine C Lin. 2024. Artificial intelligence as a double-edged sword: Wielding the POWER principles to maximize its positive effects and minimize its negative effects. *Contemporary Issues in Early Childhood* 25, 1 (2024), 146–153.
- [29] Junghee Choi, Ju-Ho Lee, and Booyuel Kim. 2019. How does learner-centered education affect teacher self-efficacy? The case of project-based learning in Korea. *Teaching and Teacher Education* 85 (2019), 45–57.
- [30] Rebecca Cober, Esther Tan, Jim Slotta, Hyo-Jeong So, and Karen D Könings. 2015. Teachers as participatory designers: Two case studies with technology-enhanced learning environments. *Instructional Science* 43 (2015), 203–228.
- [31] Kabba E Colley. 2006. Understanding ecology content knowledge and acquiring science process skills through project-based science instruction. *Science activities* 43, 1 (2006), 26–33.
- [32] Barbara Condliffe. 2017. Project-Based Learning: A Literature Review. Working Paper. *MDRC* (2017).
- [33] Bill Cope, Mary Kalantzis, and Duane Sears. 2021. Artificial intelligence for education: Knowledge and its assessment in AI-enabled learning ecologies. *Educational philosophy and theory* 53, 12 (2021), 1229–1245.
- [34] Linda Darling-Hammond, Frank Adamson, and Jamal Abedi. 2010. *Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning*. Stanford Center for Opportunity Policy in Education.
- [35] W Diehl, T Grobe, H Lopez, and C Cabral. 1999. Project-based learning: A strategy for teaching and learning.
- [36] Betsy DiSalvo and Kayla DesPortes. 2017. Participatory design for value-driven learning. In *Participatory design for learning*. Routledge, 175–188.
- [37] Kyle Dunbar and Aman Yadav. 2022. Shifting to student-centered learning: Influences of teaching a summer service learning program. *Teaching and Teacher Education* 110 (2022), 103578.
- [38] Srijani Dutta, Sakshi Ranjan, Sushruta Mishra, Vandana Sharma, Pradeep Hewage, and Celestine Iwendi. 2024. Enhancing Educational Adaptability: A Review and Analysis of AI-Driven Adaptive Learning Platforms. In *2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM)*. IEEE, 1–5.
- [39] Gregory Dyke, David Adamson, Iris Howley, and Carolyn Penstein Rosé. 2013. Enhancing scientific reasoning and discussion with conversational agents. *IEEE Transactions on Learning Technologies* 6, 3 (2013), 240–247.
- [40] Mary C English and Anastasia Kitsantas. 2013. Supporting student self-regulated learning in problem-and project-based learning. *Interdisciplinary journal of problem-based learning* 7, 2 (2013), 6.
- [41] Peggy A Ertmer and Krista D Simons. 2006. Jumping the PBL implementation hurdle: Supporting the efforts of K–12 teachers. *Interdisciplinary Journal of Problem-based learning* 1, 1 (2006), 40–54.
- [42] Virginia Eubanks. 2018. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- [43] Xiao Fang, Shangkun Che, Minjia Mao, Hongzhe Zhang, Ming Zhao, and Xiaohang Zhao. 2024. Bias of AI-generated content: an examination of news produced by large language models. *Scientific Reports* 14, 1 (2024), 5224.
- [44] Deborah Fields, Debora Lui, Yasmin Kafai, Gayithri Jayathirtha, Justice Walker, and Mia Shaw. 2023. Communicating about computational thinking: understanding affordances of portfolios for assessing high school students' computational thinking and participation practices. In *Assessing Computational Thinking*. Routledge, 112–146.
- [45] Miranda S Fitzgerald. 2020. Overlapping opportunities for social-emotional and literacy learning in elementary-grade project-based instruction. *American Journal of Education* 126, 4 (2020), 573–601.
- [46] Shelagh A Gallagher. 1997. Problem-based learning: Where did it come from, what does it do, and where is it going? *Journal for the Education of the Gifted* 20, 4 (1997), 332–362.
- [47] Debbie Goh and Ugur Kale. 2016. The urban–rural gap: project-based learning with Web 2.0 among West Virginian teachers. *Technology, Pedagogy and Education* 25, 3 (2016), 355–376.
- [48] Jennifer Gonzalez. 2015. Know Your Terms: Single-Point Rubric. <https://www.culttopedagogy.com/single-point-rubric/>. Accessed: 2024-08-19.
- [49] Kai Guo, Yuchun Zhong, Danling Li, and Samuel Kai Wah Chu. 2023. Effects of chatbot-assisted in-class debates on students' argumentation skills and task motivation. *Computers & Education* 203 (2023), 104862.
- [50] Jerry Ryan David Gustafson, Gaganpreet Jhaji, Xiaokun Zhang, and Fuhua Oscar Lin. 2025. Enhancing Project-Based Learning With a GenAI Tool Based on Retrieval: Augmented Generation and Knowledge Graphs. In *AI Applications and Strategies in Teacher Education*. IGI Global, 161–194.
- [51] Ariel Han, Xiaofei Zhou, Zhenyao Cai, Shenshen Han, Richard Ko, Seth Corrigan, and Kylie A Peppler. 2024. Teachers, Parents, and Students' perspectives on Integrating Generative AI into Elementary Literacy Education. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.
- [52] Adam Handelzalts. 2019. *Collaborative curriculum development in teacher design teams*. Springer International Publishing.
- [53] John Hattie and Mark Gan. 2011. Instruction based on feedback. In *Handbook of research on learning and instruction*. Routledge, 263–285.
- [54] William Havice, Pamela Havice, Chelsea Waugaman, and Kristin Walker. 2018. Evaluating the effectiveness of integrative STEM education: Teacher and administrator professional development. *Journal of Technology Education* 29, 2 (2018), 73–90.
- [55] Pedro Hernández-Ramos and Susan De La Paz. 2009. Learning history in middle school by designing multimedia in a project-based learning experience. *Journal of Research on Technology in Education* 42, 2 (2009), 151–173.
- [56] Nancy B Hertzog. 2007. Transporting pedagogy: Implementing the project approach in two first-grade classrooms. *Journal of Advanced Academics* 18, 4 (2007), 530–564.

- [57] Denise Higgins, Adelaide Dennis, Angela Stoddard, Alexander G Maier, and Susan Howitt. 2019. 'Power to empower': conceptions of teaching and learning in a pedagogical co-design partnership. *Higher Education Research & Development* 38, 6 (2019), 1154–1167.
- [58] Margaret Holm. 2011. Project-based instruction: A review of the literature on effectiveness in prekindergarten. *River academic journal* 7, 2 (2011), 1–13.
- [59] James Hutson and Jason Ceballos. 2023. Rethinking education in the age of AI: the importance of developing durable skills in the Industry 4.0. *Journal of Information Economics* 1, 2 (2023).
- [60] Intel Education Initiative et al. 2007. Designing effective projects: Characteristics of projects. *Benefits of Project-based Learning*. Santa Clara, CA: Intel Corporation (2007).
- [61] Stavroula Kaldi, Diamanto Filippatou, and Christos Govaris. 2011. Project-based learning in primary schools: Effects on pupils' learning and attitudes. *Education* 3–13 39, 1 (2011), 35–47.
- [62] Yael Kali, Marcia C Linn, et al. 2008. Technology-enhanced support strategies for inquiry learning. *Handbook of research on educational communications and technology* (2008), 145–161.
- [63] Md Kabin Hasan Kanchon, Mahir Sadman, Kaniz Fatema Nabila, Ramisa Taranum, and Riasat Khan. 2024. Enhancing personalized learning: AI-driven identification of learning styles and content modification strategies. *International Journal of Cognitive Computing in Engineering* 5 (2024), 269–278.
- [64] Hanseob Kim, Bin Han, Jieun Kim, Muhammad Firdaus Syawaludin Lubis, Gerard Joungyun Kim, and Jae-In Hwang. 2024. Engaged and Affective Virtual Agents: Their Impact on Social Presence, Trustworthiness, and Decision-Making in the Group Discussion. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.
- [65] Soomin Kim, Jinsu Eun, Changhoon Oh, Bongwon Suh, and Joohnwan Lee. 2020. Bot in the bunch: Facilitating group chat discussion by improving efficiency and participation with a chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [66] Dimitra Kokotsaki, Victoria Menzies, and Andy Wiggins. 2016. Project-based learning: A review of the literature. *Improving schools* 19, 3 (2016), 267–277.
- [67] Siu-Cheung Kong, Man-Yin William Cheung, and Olson Tsang. 2024. Developing an artificial intelligence literacy framework: Evaluation of a literacy course for senior secondary students using a project-based learning approach. *Computers and Education: Artificial Intelligence* 6 (2024), 100214.
- [68] Joseph S Krajcik and Phyllis C Blumenfeld. 2006. *Project-based learning*. na.
- [69] Joseph S Krajcik and Kongju Mun. 2014. Promises and challenges of using learning technologies to promote student learning of science. In *Handbook of research on science education, volume II*. Routledge, 337–360.
- [70] Rohit Kumar and Carolyn P Rose. 2014. Architecture for building conversational agents that support collaborative learning. *IEEE Transactions on Learning Technologies* 4, 1 (2010), 21–34.
- [71] Gila Kurtz, Meital Amzalag, Nava Shaked, Yanay Zaguri, Dan Kohen-Vacs, Eran Gal, Gideon Zailer, and Eran Barak-Medina. 2024. Strategies for Integrating Generative AI into Higher Education: Navigating Challenges and Leveraging Opportunities. *Education Sciences* 14, 5 (2024), 503.
- [72] Steinar Kvale and Svend Brinkmann. 2009. *Interviews: Learning the craft of qualitative research interviewing*. sage.
- [73] Shui-fong Lam, Rebecca Wing-yi Cheng, and Harriet C Choy. 2010. School support and teacher motivation to implement project-based learning. *Learning and instruction* 20, 6 (2010), 487–497.
- [74] Yu-Ju Lan and Nian-Shing Chen. 2024. Teachers' agency in the era of LLM and generative AI. *Educational Technology & Society* 27, 1 (2024), I–XVIII.
- [75] Q Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: informing design practices for explainable AI user experiences. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–15.
- [76] G. C. Lin. 2023. *Internal Memo/Report for Project-Based Learning Assessment: Portfolio Approach*. Technical Report. Scheller Teacher Education Program | The Education Arcade.
- [77] Phoebe Lin and Jessica Van Brummelen. 2021. Engaging teachers to co-design integrated AI curriculum for K-12 classrooms. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–12.
- [78] Andreas Lingnau, Andreas Harter, Markus Kuhn, and H ULRICH HOPPE. 2007. Empowering teachers to evolve media enriched classroom scenarios. *Research and Practice in Technology Enhanced Learning* 2, 02 (2007), 105–129.
- [79] Y-C Liu, Amaresh Chakrabarti, and Thomas Bligh. 2003. Towards an 'ideal' approach for concept generation. *Design studies* 24, 4 (2003), 341–355.
- [80] Anette Markula and Maija Aksela. 2022. The key characteristics of project-based learning: how teachers implement projects in K-12 science education. *Disciplinary and Interdisciplinary Science Education Research* 4, 1 (2022), 2.
- [81] Camillia Matuk, Rebecca Martin, Veena Vasudevan, Kim Burgas, Kim Chaloner, Ido Davidesco, Sushmita Sadhukha, Yury Shevchenko, Engin Bumbacher, and Suzanne Dikker. 2021. Students learning about science by investigating an unfolding pandemic. *Aera Open* 7 (2021), 23328584211054850.
- [82] Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica* 22, 3 (2012), 276–282.
- [83] Susan McKenney, Ferry Boschman, Jules Pieters, and Joke Voogt. 2016. Collaborative design of technology-enhanced learning: What can we learn from teacher talk? *TechTrends* 60 (2016), 385–391.
- [84] Gale A Mentzer, Charlene M Czerniak, and Lisa Brooks. 2017. An examination of teacher understanding of project based science as a result of participating in an extended professional development program: Implications for implementation. *School Science and Mathematics* 117, 1-2 (2017), 76–86.
- [85] John R Mergendoller and John W Thomas. 2005. Managing project based learning: Principles from the field. Retrieved June 14 (2005), 2005.
- [86] Michael Muller, Lydia B Chilton, Anna Kantosalo, Charles Patrick Martin, and Greg Walsh. 2022. GenAICHI: generative AI and HCI. In *CHI conference on human factors in computing systems extended abstracts*. 1–7.
- [87] Simon Nagy, Richard McInnes, and Laura Airey. 2023. Gen-AI: A Transformative Partner in Collaborative Course Development. *International Journal on Innovations in Online Education* 7, 2 (2023).
- [88] Sasha Nikolic and Karley Beckman. [n. d.]. Supporting Engineering Project-Based Learning through the Use of ChatGPT and Generative AI: A Case Study. In *Artificial Intelligence Applications in Higher Education*. Routledge, 215–232.
- [89] Pedro Noguera, Linda Darling-Hammond, and Diane Friedlaender. 2015. Equal opportunity for deeper learning. (2015).
- [90] Ayse Oguz-Unver and Sertac Arabacioglu. 2014. A comparison of inquiry-based learning (IBL), problem-based learning (PBL) and project-based learning (PBJL) in science education. *Academia Journal of Educational Research* 2, 7 (2014), 120–128.
- [91] Anastasia K Ostrowski, Cynthia Breazeal, and Hae Won Park. 2021. Long-term co-design guidelines: empowering older adults as co-designers of social robots. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 1165–1172.
- [92] Fan Ouyang, Mian Wu, Luyi Zheng, Liyin Zhang, and Pengcheng Jiao. 2023. Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education* 20, 1 (2023), 4.
- [93] Valentine Joseph Owan, Kingsley Bekom Abang, Delight Omoji Idika, Eugene Onor Etta, and Bassey Asuquo Bassey. 2023. Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education* 19, 8 (2023), em2307.
- [94] Mookdaporn Panasan and Prasart Nuangchalerms. 2010. Learning outcomes of project-based and inquiry-based learning activities. *Online Submission* 6, 2 (2010), 252–255.
- [95] Walter Parker, Susan Mosborg, John Bransford, Nancy Vye, John Wilkerson, and Robert Abbott. 2011. Rethinking advanced high school coursework: Tackling the depth/breadth tension in the AP US Government and Politics course. *Journal of Curriculum Studies* 43, 4 (2011), 533–559.
- [96] Buck Institute of Education PBLWorks. 2019. What is Project-based Learning? <https://www.pblworks.org/what-is-pbl>.
- [97] William R Penuel, Jeremy Roschelle, and Nicole Shechtman. 2007. Designing formative assessment software with teachers: An analysis of the co-design process. *Research and practice in technology enhanced learning* 2, 01 (2007), 51–74.
- [98] Vanessa L Peters and James D Slotta. 2009. Co-designing curricula to promote collaborative knowledge construction in secondary school science. (2009).
- [99] BW Peterson. 2012. Uncovering the progressive past: The origins of project-based learning. *UnBoxed: A Journal of Adult Learning in Schools* 8 (2012).
- [100] Dimitrios Pnevmatikos, Panagiotis Christodoulou, and Nikolaos Fachantidis. 2020. Stakeholders' Involvement In Participatory Design Approaches of Learning Environments: A Systematic Review of the Literature. *EDULEARN20 proceedings* (2020), 5543–5552.
- [101] Ashley Seidel Potvin, Alison G Boardman, and Kristina Stamatis. 2021. Consequential change: Teachers scale project-based learning in English language arts. *Teaching and Teacher Education* 107 (2021), 103469.
- [102] Janet Quint and Barbara Condliffe. 2018. Project-Based Learning: A Promising Approach to Improving Student Outcomes. Issue Focus. *MDRC* (2018).
- [103] Chris Quintana, Brian J Reiser, Elizabeth A Davis, Joseph Krajcik, Eric Fretz, Ravit Golan Duncan, Eleni Kyza, Daniel Edelson, and Elliot Soloway. 2018. A scaffolding design framework for software to support science inquiry. In *Scaffolding*. Psychology Press, 337–386.
- [104] Prerna Ravi, Azra Ismail, and Neha Kumar. 2021. The pandemic shift to remote learning under resource constraints. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–28.
- [105] Brian Reiser. 2006. Making authentic practices accessible to learners: Design challenges and strategies. In *The Cambridge handbook of the learning sciences*. Cambridge University Press, 335–354.
- [106] Brian J Reiser. 2018. Scaffolding complex learning: The mechanisms of structuring and problematizing student work. In *Scaffolding*. Psychology Press, 273–304.
- [107] Lucas Education Research. 2019. *Project-Based Learning Increases Science Achievement in Elementary School and Advances Social and Emotional Learning*. Technical Report. George Lucas Education Foundation.

- [108] Jeremy Roschelle, William Penuel, and Nicole Shechtman. 2006. Co-design of innovations with teachers: Definition and dynamics. (2006).
- [109] Jürgen Rudolph, Samson Tan, and Shannon Tan. 2023. ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of applied learning and teaching* 6, 1 (2023), 342–363.
- [110] Soumendra Sahoo, Kay Thi Myint, Htoo Htoo Kyaw Soe, and Simerjit Singh. 2013. Formative assessment of ophthalmology problem-based learning tutorials during undergraduate medical training. *The Asia-Pacific Journal of Ophthalmology* 2, 5 (2013), 282–285.
- [111] Agariadne Dwinggo Samala, Soha Rawas, Tianchong Wang, Janet Marie Reed, Jinhee Kim, Natalie-Jane Howard, and Myriam Ertz. 2024. Unveiling the landscape of generative artificial intelligence in education: a comprehensive taxonomy of applications, challenges, and future prospects. *Education and Information Technologies* (2024), 1–40.
- [112] John R Savery. 2015. Overview of problem-based learning: Definitions and distinctions. *Essential readings in problem-based learning: Exploring and extending the legacy of Howard S. Barrows* 9, 2 (2015), 5–15.
- [113] Bertrand Schneider and Roy Pea. 2014. Toward collaboration sensing. *International Journal of Computer-Supported Collaborative Learning* 9, 4 (2014), 371–395.
- [114] Johannes Schneider, Bernd Schenk, Christina Niklaus, and Michaelis Vlachos. 2023. Towards llm-based autograding for short textual answers. *arXiv preprint arXiv:2309.11508* (2023).
- [115] Rebecca M Schneider, Joseph Krajcik, Ronald W Marx, and Elliot Soloway. 2002. Performance of students in project-based science classrooms on a national measure of science achievement. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching* 39, 5 (2002), 410–422.
- [116] Jonathan Schweig. 2019. *Measuring Teacher Effectiveness: Understanding Common, Uncommon, and Combined Methods*. RAND Corporation, Santa Monica, CA. doi:10.7249/RR4312.4
- [117] Boran Sekeroglu, Kamil Dimililer, and Kubra Tuncal. 2019. Student performance prediction and classification using machine learning algorithms. In *Proceedings of the 2019 8th international conference on educational and information technology*. 7–11.
- [118] Orit Shaer, Angelora Cooper, Osnat Mokryn, Andrew L Kun, and Hagit Ben Shoshan. 2024. AI-Augmented Brainwriting: Investigating the use of LLMs in group ideation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.
- [119] Krista D Simons, James D Klein, and Thomas R Brush. 2004. Instructional strategies utilized during the implementation of a hypermedia, problem-based learning environment: A case study. *Journal of Interactive Learning Research* 15, 3 (2004), 213–233.
- [120] Hyo-Jeong So, Chohui Lee, Sujin Kim, and Nur Arifah Drajiati. 2024. Enhancing Teacher Competency for Creative Problem Solving with Generative AI. In *EdMedia+ Innovate Learning*. Association for the Advancement of Computing in Education (AACE), 277–284.
- [121] Daniel Spikol, Marcelo Milrad, Heidy Maldonado, and Roy Pea. 2009. Integrating co-design practices into the development of mobile science collaboratories. In *2009 ninth IEEE international conference on advanced learning technologies*. IEEE, 393–397.
- [122] Steven E Stemler. 2019. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research, and Evaluation* 9, 1 (2019), 4.
- [123] Yuan Sun, Eunhae Jang, Fenglong Ma, and Ting Wang. 2024. Generative AI in the Wild: Prospects, Challenges, and Strategies. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–16.
- [124] Vanessa Svihla and Richard Reeve. 2020. Facilitating problem framing in project-based learning. *Interactional research into problem-based learning* (2020), 129–164.
- [125] Dattatray G Takale, Parikshit N Mahalle, and Bipin Sule. 2024. Assessing Generative AI-Generated Assignments: A Framework Based on the 3R. *Journal of Data Mining and Management* 9, 2 (2024), 11–16.
- [126] Seng Chee Tan, Alwyn Wven Yen Lee, and Min Lee. 2022. A systematic review of artificial intelligence techniques for collaborative learning over the past two decades. *Computers and Education: Artificial Intelligence* 3 (2022), 100097.
- [127] Xiaoqi Tanga, Xinyan Mac Xu Dingb, Siman Zhangd, and Junfeng Diaoe. 2024. An Exploration of Project-Based Learning Supported by Artificial Intelligence. In *2024 5th International Conference on Big Data and Informatization Education (ICBDIE 2024)*. Atlantis Press, 220–230.
- [128] Leah Teeters, A Susan Jurow, and Molly Shea. 2016. The challenge and promise of community co-design. In *Design as Scholarship*. Routledge, 41–54.
- [129] John W Thomas. 1998. Project based learning overview. *Novato, CA: Buck Institute for Education* 3 (1998).
- [130] John W Thomas. 2010. A review of research on project-based learning. 2000. *The Autodesk Foundation: San Rafael* (2010).
- [131] John W Thomas, John R Mergendoller, and Andrew Michaelson. 1999. *Project based learning: A handbook for middle and high school teachers*. Buck Institute for Education.
- [132] Paraskevi Topali, Alejandro Ortega-Arranz, María Jesús Rodríguez-Triana, Erkan Er, Mohammad Khalil, and Gökhan Akçapınar. 2024. Designing human-centered learning analytics and artificial intelligence in education solutions: a systematic literature review. *Behaviour & Information Technology* (2024), 1–28.
- [133] Torrey Trust and Jeromie Whalen. 2021. Emergency remote teaching with technology during the COVID-19 pandemic: using the whole teacher lens to examine educator’s experiences and insights. *Educational Media International* 58, 2 (2021), 145–160.
- [134] Kuo-Hung Tseng, Chi-Cheng Chang, Shi-Jer Lou, and Wen-Ping Chen. 2013. Attitudes towards science, technology, engineering and mathematics (STEM) in a project-based learning (PjBL) environment. *International Journal of Technology and Design Education* 23 (2013), 87–102.
- [135] Ruben Vanderlinde and Johan van Braak. 2010. The gap between educational research and practice: Views of teachers, school leaders, intermediaries and researchers. *British educational research journal* 36, 2 (2010), 299–316.
- [136] Elina Viro, Daranee Lehtonen, Jorma Joutsenlahti, and Ville Tahvanainen. 2020. Teachers’ perspectives on project-based learning in mathematics and science. (2020).
- [137] Joke Voogt, Therese Laferriere, Alain Breuleux, Rebecca C Itow, Daniel T Hickey, and Susan McKenney. 2015. Collaborative design as a form of professional development. *Instructional science* 43 (2015), 259–282.
- [138] Susan J Wake and Chris Eames. 2013. Developing an “ecology of learning” within a school sustainability co-design project with children in New Zealand. *Local Environment* 18, 3 (2013), 305–322.
- [139] Xiaoyu Wan, Jingwan Tang, Xiaofei Zhou, and Zhen Bai. 2021. Exploratory Process Analysis of Teacher Learning of AI Integration through Collaborative Design. (2021).
- [140] Shan Wang, Fang Wang, Zhen Zhu, Jingxuan Wang, Tam Tran, and Zhao Du. 2024. Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications* 252 (2024), 124167.
- [141] Justin D Weisz, Jessica He, Michael Muller, Gabriela Hoefler, Rachel Miles, and Werner Geyer. 2024. Design Principles for Generative AI Applications. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–22.
- [142] Peter J. Woods, Avneet Hira, Emma Anderson, and Amelia M. Farid. 2024. Towards a Framework for Collaborative Design in PBL-centered Instructional Coaching. *Under review* (2024).
- [143] Shin Park Young. 2018. How to equip students to be problem solvers through STEAM The emphasis of scientific thinking, computational thinking, and emotional thinking. *JSSE Research Report* 32, 8 (2018), 3–6.
- [144] Siyu Zha, Yuehan Qiao, Qingyu Hu, Zhongsheng Li, Jiangtao Gong, and Yingqing Xu. 2024. Designing Child-Centric AI Learning Environments: Insights from LLM-Enhanced Creative Project-Based Learning. *arXiv preprint arXiv:2403.16159* (2024).
- [145] BaoHui Zhang, Chee-Kit Looi, Peter Seow, Gean Chia, Lung-Hsiang Wong, Wenli Chen, Hyo-Jeong So, Elliot Soloway, and Cathie Norris. 2010. Deconstructing and reconstructing: Transforming primary science learning via a mobilized curriculum. *Computers & Education* 55, 4 (2010), 1504–1523.
- [146] Chengbo Zheng, Zeyu Huang, Shuai Ma, and Xiaojuan Ma. 2024. SelfGauge: An Intelligent Tool to Support Student Self-assessment in GenAI-enhanced Project-based Learning. In *Adjunct Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–3.
- [147] Chengbo Zheng, Kangyu Yuan, Bingcan Guo, Reza Hadi Mogavi, Zhenhui Peng, Shuai Ma, and Xiaojuan Ma. 2024. Charting the Future of AI in Project-Based Learning: A Co-Design Exploration with Students. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–19.