

Desiderata for AI Tutors as Knowledge-Building Partners

Hari Subramonyam, Mayank Sharma, Roy Pea
Graduate School of Education, Stanford University, Stanford, CA, USA
[harihars,masharma,roypea]@stanford.edu

Reframing LLM tutors from knowledge-passing systems to constructivist collaborators

Abstract. Current Large Language Model (LLM) tutors mimic traditional teachers, focusing on knowledge-passing. We propose reframing them as knowledge-building partners, shifting to a constructivist approach that empowers learners as active knowledge creators. We examine six key dimensions: cognitive engagement, cultural responsiveness, formative assessment, power dynamics, accountability, and metacognition, outlining desiderata for LLM tutors in each area. By reimagining LLMs' role, we can potentially transform education while addressing longstanding challenges.

Keywords. Intelligent Tutoring Systems, Large Language Models, Constructivism, Knowledge Building, Learning Conversations

1 Introduction

Large language models (LLMs) are driving the evolution of intelligent tutoring systems towards personalized and adaptive education [1]. For instance, Khanmigo¹, developed by Khan Academy in collaboration with OpenAI, can tutor students through Socratic dialogues, guiding them towards answers with targeted questions, offering personalized explanations and instant work feedback. Although current LLM tutors facilitate one-on-one tutoring [2], they are primarily directed toward mimicking teachers and their *knowledge-passing* tendencies and mitigating concerns about academic integrity [3]. Specific instructions to LLM tutors include withholding answers to encourage critical thinking, imposing roleplay constraints to maintain academic integrity, enforcing topic restrictions to prevent misuse of the technology, and limiting the LLM's ability to generate specific types of content such as sensitive or inappropriate material [4, 5], etc. While such strategies may help provide immediate guidance and convey knowledge, they fall short of realizing the full potential of LLMs in education.

Instead, we propose that LLM tutors should be geared towards fostering *knowledge-building* conversations within learning communities [6] as *knowledge-building partners*. Our perspective is based on constructivist teaching methods [7] and aims to empower learners to take an *active role* in developing and improving ideas, moving beyond passive retention of information to engage in meaningful, collaborative learning experiences. For instance, in learning about why dinosaurs disappeared, instead of simply stating that an asteroid hit

¹<https://www.khanacademy.org/khan-labs>

the planet and caused their extinction, teachers can engage students by asking, “What ideas do you have about why dinosaurs may have disappeared? [8, 9, 10]” Student talk moves could include suggestions such as extreme temperatures, disappearing food sources, dangerous predators, etc. The teacher can then facilitate a discussion on developing these ideas into theories, considering different perspectives that have been put forward on the mass extinction event, and exploring articles seeking evidence for and against alternative theories. Such knowledge generation activities are crucial to productive learning, but unfortunately, they are challenging to enact in current classroom settings [11].

In current classrooms, traditional teaching methods, rigid curricula, assessment practices, classroom dynamics and its culture frequently prohibit knowledge-building discourse [12, 13]. Some teachers view the teaching process as a formal knowledge transfer from expert instructors to inexperienced learners [14]. Consequently, students are positioned as mere recipients of information and this learning environment design fails to capitalize on the diverse perspectives and experiences that students bring to the classroom. AI tutors have potential to overcome these limitations by creating decentralized, open, and interactive learning environments essential for engaging students in meaningful and collaborative knowledge-building activities [15]. LLMs with their vast knowledge base, advanced natural language processing capabilities, and contextual understanding, can facilitate rich, interactive dialogues that prompt critical thinking, encourage questioning, and help students connect ideas, develop theories, and build their knowledge [16].

Given these capabilities, we identify desiderata for LLM tutors to address longstanding issues in K-20 education by examining current constraints to knowledge-building ‘learning conversations.’ Specifically, we conjecture how LLM tutors can address challenges in cognitive engagement, cultural responsiveness, formative assessment, power dynamics, accountability, and metacognition within knowledge-building conversations. For each dimension, we identify needs and operationalization challenges in current classroom settings and propose requirements for LLM tutors that reframe their role in education. By shifting from knowledge-passing to knowledge-building discourse, LLM tutors can create more engaging, effective, and transformative learning experiences. These advancements will be important not only for K-20 formal education but for professional learning among adults [17].

2 Current Challenges in Knowledge-Building

Decades of educational research and studies in the learning sciences have surfaced a variety of learning environment dimensions that are central to their effective functioning as knowledge-building communities. Here, we provide briefs into six vital dimensions and critically consider how LLMs might contribute to their embodiment in learning environments by productively exploiting their unique affordances as augmentation engines for human intelligence.

2.1 Cognitive Engagement

Cognitive engagement is crucial for effective learning, defined in terms of the depth of processing and the quality of thinking strategies students use in their learning tasks [18, 19]. Studies show it leads to better retention, more nuanced understanding, and improved ability to apply knowledge in new contexts [20]. Cognitive engagement is the difference between memorizing facts for a test and genuinely engaging with concepts to develop lasting comprehension [21]. Yet many teachers struggle to consistently create learning environments to promote high levels of cognitive engagement for all students [22, 23].

Designing activities sufficiently challenging yet accessible to students with diverse abilities and backgrounds is complex. This challenge is rooted in the concept of the Zone of Proximal Development (ZPD), introduced by Vygotsky [24] - the space between what a learner can do without assistance and with adult guidance or collaboration with more able peers. Educators must create tasks that are within each student’s ZPD,

which varies widely in diverse classrooms [25]. It is crucial to consider the diverse types of educational conversations that can foster deep learning. For instance, critical literacy discussions can promote analysis of underlying messages and biases in texts, developing students' competence in questioning societal and cultural assumptions [26]. Similarly, problem-based learning (PBL) discussions engage students with complex, real-world problems, enhancing research and collaborative skills [27]. However, effectively facilitating these varied conversation types requires significant educator skill, and traditional assessment methods often fall short in capturing the nuanced learning outcomes of such diverse dialogic interactions [28, 29].

LLMs may offer promising solutions to these challenges by supporting a wide range of discourse patterns and approaches, adapting to various conversational styles. They can simulate different roles (aka 'persona': [30]) in role-playing conversations, assist in brainstorming sessions, and provide instant feedback in reciprocal teaching scenarios [16]. Moreover, LLMs can help address the challenge of personalization in diverse classrooms by tailoring conversations to individual students' interests, cultural backgrounds, and learning paces.

Desiderata for LLM Tutors:

- **Discourse Versatility:** LLMs should be capable of facilitating various types of educational conversations, including critical literacy discussions and problem-based learning dialogues, adapting their communication style accordingly.
- **Generative Questioning:** The system should generate thought-provoking questions that stimulate critical thinking and encourage students to make connections across concepts and domains.

2.2 Cultural Responsiveness

Any teacher knows about the tremendous variability among their students in cultural heritage and many other aspects of their identities the teacher needs to understand to provide effective instruction. Teachers can make efforts to throw cultural relevance bridges during classroom interactions concerning topics being learned to the different students, leveraging their out-of-school experiences and background cultural knowledge. For instance, Ladson-Billings [31, 32] observes that "Teachers who foster cultural competence understand that they must work back and forth between the lives of their students and the life of school." Au and Jordan [33] called teacher pedagogy in a Hawaiian school "culturally appropriate" when it incorporated into reading instruction features of students' cultural backgrounds. Students used their Native Hawaiian children talk-story interaction style on standardized reading tests, achieving at higher than predicted levels. Mohatt and Erickson's study of teacher-learner "culturally responsive" interactions and participation structures for Native American students discovered teachers using language interaction patterns like students' home cultural patterns were better at improving student academic performance [34, 35, 36].

However, implementing culturally responsive pedagogy presents significant challenges. In diverse classrooms, balancing the needs and cultural backgrounds of multiple student groups simultaneously can overwhelm them. Teachers struggle to spontaneously draw upon students' cultural knowledge and experiences to illustrate academic concepts, especially when working with diverse classrooms [37]. This difficulty stems from the complex interplay of content knowledge, pedagogical skills, and cultural understanding required to create these bridges in real time.

LLMs could be engineered to be attentive to and respectful of each student's cultural identities and funds of knowledge [38], potentially overcoming some of the challenges human teachers face in implementing culturally responsive practices. LLMs could be programmed to generate culturally relevant examples, analogies, and connections on the fly, drawing from various cultural contexts to illustrate academic concepts [37]. This capability could help address the difficulty teachers often face in spontaneously creating cultural bridges

during classroom interactions. But it is essential to note that LLMs currently lack a valid theory of mind – the ability to track other people’s mental states – about the learner with which they are interacting—although technical progress is being made in this regard [39].

Desiderata for LLM Tutors:

- **Cultural Knowledge Integration:** LLM tutors should generate and use culturally appropriate analogies, metaphors, and examples that resonate with individual students’ backgrounds.
- **Cultural Competence Development:** Encourage students to explore and articulate their own cultural identities while fostering understanding and appreciation of others’ cultures.

2.3 Formative Assessment

Formative assessment is essential to the regulation of the learning process [40, 41]. It refers to the ongoing interactive evaluation employed by both educators and learners to assess learning progress as it occurs [42]. Pellegrino compellingly describes it as ‘assessment *for* learning’ in contrast to ‘assessment *of* learning’ [43]. To effectively provide formative assessments, teachers must (1) notice student behavior and engagement during the learning process, (2) recognize or interpret the observations in the context of learning goals and expected progress, and (3) respond by adjusting instructions, asking probing questions or providing targeted feedback, etc [2, 40, 44, 45]. Unfortunately, in traditional classrooms, teachers don’t always administer formative assessments because lessons are often planned with specific content objectives to be covered in a set time frame [45, 46, 47]. Additionally, creating varied, engaging, and effective formative assessments that meet individual student needs while maintaining consistency of instructions can be challenging [48, 49]. Studies show that students may find assessment questions inappropriate and irrelevant or experience a misalignment between what they know and need to know [50, 51, 52, 53]. Consequently, evaluation is often done at the lesson end in the form of summative assessments limiting opportunities for teachers or students to address misconceptions or knowledge gaps in a timely manner [54].

In contrast, knowledge-building approaches require regular use of formative assessments to recognize and foster emergent student understanding, facilitate inquiry discussions, and evaluate metacognitive skills during knowledge-building discourse. LLMs should be able to significantly enhance this process through real-time analysis of student contributions, identifying key concepts, novel ideas, and potential misconceptions as they emerge [55, 56]. By generating thought-provoking questions based on the history of an ongoing discourse, LLMs can push students to explore ideas further or consider alternative perspectives, deepening the collective inquiry. Their capacity to analyze the language and structure of student contributions also allows for the evaluation of crucial metacognitive aspects such as self-reflection, critical thinking, and the ability to build on others’ ideas, supporting both teachers and students in the collaborative learning process.

Desiderata for LLM Tutors:

- **Continuous Assessment:** Operate as an ongoing process throughout the knowledge-building dialog.
- **Iterative Improvement:** Assess current understanding against previous states to measure knowledge growth and enhancement of ideas over time
- **Synthesis Capability:** Evaluate learners’ ability to link concepts and integrate diverse ideas into coherent knowledge.

2.4 Power Dynamics

Power dynamics in educational settings are complex and fluid, as highlighted by Foucault’s relational concept of power [57]. Rather than being a top-down force or an inherent attribute of specific individuals or groups, power manifests itself in the intricate web of interactions among students, teachers, and even the subject matter itself [58, 59]. Shah and Lewis’s [60] empirical study in a high school computer science classroom provides valuable insights into these dynamics. They conceptualize power relationships as a dynamic balance, constantly shifting in response to various situational factors. Concretely, classroom power dynamics comprise at least three key dimensions, including (1) *ownership of ideas*, (2) *partisanship*, and (3) *persuasive discourse*. Ownership of ideas in a discourse as described by Engle and Conant [61], reflects the power relationship between individuals and concepts. In educational settings, the perception of who owns knowledge—be it teachers, textbooks, students themselves, or peers—significantly influences students’ relationships with both the ideas and other classroom participants. This ownership perception affects cognitive engagement and shapes learning interactions. Traditional classrooms often reinforce teacher authority through authoritative speech patterns and evaluative questioning [62, 63], implicitly asserting the teacher’s ownership of content.

Regarding partisanship, Hatano and Inagaki [64] highlight how power dynamics emerge through ‘taking sides’ in classroom interactions. Students form alliances around concepts or during discussions, creating a network of power relationships that influence how ideas are received, supported, or challenged within the learning community. Third, persuasive discourse, rooted in Bakhtin’s work [65], focuses on how communication styles themselves shape power dynamics. It encompasses both externally authoritative discourse, where power stems from position or perceived expertise, and internally persuasive discourse, where influence is gained through the strength and resonance of arguments. The interplay of idea ownership, partisanship, and persuasive discourse significantly shapes how knowledge is constructed, shared, and validated within a classroom.

LLMs can play a significant role in managing and potentially rebalancing power dynamics in classroom knowledge-building conversations. LLMs can facilitate more equitable participation by monitoring engagement patterns and prompting less vocal students, while also mediating persuasive discourse to encourage consideration of multiple perspectives. They can help mitigate biases arising from social or academic status, providing a more neutral ground for idea evaluation. However, it is crucial to implement these capabilities thoughtfully, ensuring that the LLM itself doesn’t become a new locus of authority that students overly defer to.

Desiderata for LLM Tutors:

- **Equitable Idea Attribution:** LLMs should accurately track and acknowledge the origin of ideas, ensuring proper credit is given to students who contribute novel thoughts or insights, thus democratizing idea ownership.
- **Adaptive Authority:** LLMs should dynamically adjust their level of directiveness, gradually ceding authority to students as they demonstrate increased capability in managing productive knowledge-building discourse.
- **Transparency of Influence:** The system should make its own role in shaping conversations explicit, helping students understand and critically evaluate the AI’s impact on their discussions.

2.5 Accountability

Accountability in education refers to the systemic responsibility for achieving learning outcomes, involving standard-setting, performance measurement, and holding stakeholders responsible for educational goals [66]. Accountable Talk, a concept developed by Resnick, Michaels and O’Connor [67], operationalizes accountability at the classroom level through *structured discourse*. It encompasses three key dimensions: accountability

to the learning community, to accurate and appropriate knowledge, and to rigorous thinking. To facilitate Accountable Talk, educators often employ specific discourse strategies called “talk moves,” which include techniques such as revoicing, asking students to apply their own reasoning to someone else’s idea, and prompting students to extend or elaborate on their thoughts [68].

Unfortunately, in traditional classrooms and many current intelligent tutoring systems, accountability is often centered on individual performance and adherence to predefined standards. While these systems may attempt to structure conversations, they typically fail to foster truly accountable talk [69, 70, 71]. Intelligent tutors, designed within this paradigm, usually emphasize correct answers and fixed solution paths rather than open-ended exploration. For instance, a science tutor might quiz students on the stages of photosynthesis but rarely encourage them to formulate and test hypotheses about how changing environmental factors might affect the process. Accountability to knowledge in these systems is often limited to verifying facts against a static knowledge base, rather than encouraging students to seek out and evaluate new information critically for expansive learning [72].

In knowledge-building conversations, tutors must facilitate and assess higher-order cognitive engagement [73]. Moreover, such an LLM-augmented learning environment would need to dynamically assess the quality of student contributions not only for factual accuracy, but for depth of reasoning, use of evidence, and engagement with others’ ideas, etc. LLMs can enhance accountability and rigor in knowledge-building conversations by analyzing discourse patterns, tracking idea development, and prompting for evidence-based reasoning. They can identify when claims lack support, recognize logical inconsistencies, and encourage students to engage more deeply with each other’s ideas.

Desiderata for LLM Tutors:

- **Evidence Tracking:** LLM tutors can track emerging claims and prompt students to provide evidence for unsupported claims, highlighting when new information contradicts or supports previous assertions.
- **Conceptual Depth Gauging:** LLM tutors should evaluate the depth of conceptual understanding demonstrated in the conversation and challenge surface-level discussions by asking probing questions to require students explain underlying principles, make connections between ideas, and apply concepts to new situations.
- **Participation monitoring:** LLM tutors should actively track and promote the student’s sustained, high-quality participation by ensuring conversation remains student-driven, and remedy patterns of passive responses.

2.6 Metacognition

Metacognition, a fundamental aspect of effective learning, encompasses the ability to reflect on, understand, and regulate one’s cognitive processes. Flavell [74] defined it as “knowledge or cognition that takes as its object or regulates any aspect of cognitive endeavor.” Metacognition is divided into two main components: knowledge about cognition and regulation of cognition [75, 76]. The knowledge component involves understanding one’s own cognitive abilities and learning strategies, while the regulatory component focuses on actively managing and optimizing learning processes. Metacognition’s knowledge aspect can be differentiated into three types: declarative (knowing about oneself and strategies), procedural (knowing how to use strategies), and conditional (knowing when and why to use strategies). The regulatory aspect involves subprocesses of planning, information management, comprehension monitoring, debugging strategies, and evaluation [77].

In the context of knowledge-building learning conversations, fostering metacognition presents unique

Dimension	Knowledge-Passing Workflow	Knowledge-Building Workflow
Cognitive Engagement	Information provider; focuses on memorization of facts.	Cognitive partner and facilitator; emphasizes deep processing and high-quality thinking strategies.
Cultural Responsiveness	Limited consideration of diverse cultural backgrounds.	Integrates students' cultural identities and funds of knowledge; uses culturally relevant examples and analogies.
Formative Assessment	Often limited to end-of-lesson summative assessments.	Continuous, interactive evaluation throughout the learning process; focuses on progress and process.
Power Dynamics	Teacher-centered; reinforces teacher authority.	Collaborative; balances power among students, teachers, and subject matter.
Accountability	Focused on individual performance and adherence to standards.	Emphasizes accountability to the learning community, knowledge accuracy, and rigorous thinking.
Metacognition	Limited explicit focus on metacognitive strategies.	Actively promotes reflection on cognitive processes and regulation of learning.

Table 1: Comparison of knowledge-passing and knowledge-building workflows.

challenges. The dynamic and collaborative nature of these interactions requires students to simultaneously engage with content, monitor their understanding, and respond to the evolving ideas of their peers. This cognitive multitasking can be overwhelming, particularly for novice learners [78]. Furthermore, assessing the quality of one’s contributions in relation to the collective knowledge advancement of the group adds another layer of complexity to metacognitive processes [79]. LLMs can potentially enhance metacognition by offering prompts encouraging students to reflect on their thinking processes, assess their understanding, and consider alternative perspectives. Additionally, LLMs could model metacognitive strategies by “thinking aloud” about complex problems, demonstrating how to approach, monitor, and evaluate cognitive processes. By analyzing discourse patterns, LLMs might identify moments when metacognitive interventions could be most beneficial, such as when a student’s contribution shows a shift in understanding or when the group’s discussion reaches a critical juncture.

Desiderata for LLM Tutors:

- **Reflective Prompting:** LLMs should generate timely, context-sensitive prompts that encourage students to reflect on their thinking processes and the evolution of their ideas within the conversation.
- **Metacognitive Modeling:** LLMs should demonstrate metacognitive strategies by “thinking aloud” about complex problems, showing students how to plan, monitor, and evaluate their cognitive processes.

3 Discussion

Our perspective that LLM tutors should be modeled as knowledge-building partners represents a significant paradigm shift for AI’s roles in education. We are moving away from traditional knowledge-passing instantiations of intelligent tutors to a more dynamic, constructivist model of learning. In terms of benefits, such a shift can better prepare students for increasingly complex real-world challenges such as climate change, in which they need to navigate ambiguity, synthesize diverse perspectives, and tackle problems that require collaborative knowledge building. LLMs have the capacity to facilitate rich interactive dialogues, promote critical thinking, encourage questioning, adapt to various conversational styles, and support a wide range of

discourse patterns that could help address longstanding challenges in education.

Although such a transformation offers exciting possibilities, it also exposes substantial challenges that must be carefully addressed. In addition to hallucination and accuracy concerns [80], a significant threat is the potential for “knowledge collapse [81].” LLMs’ vast but sometimes inconsistent knowledge base could lead to information homogenization, marginalizing non-mainstream perspectives, or reinforcing misconceptions. Another challenge is the potential for AI over-reliance. Students might develop a habit of turning to LLM tutors for quick answers rather than engaging in the productive struggle [82, 83] often crucial for deeper understanding. This raises questions about how to strike the right balance between AI assistance and independent learning. In addition, LLMs as knowledge-building partners will necessitate a shift in the role of educators. Teachers must develop new skills in AI literacy, data interpretation, and adaptive teaching strategies [84]. They will need to learn how to effectively integrate LLM tutors into their teaching practice, using these tools to enhance rather than replace human instruction [85]. This transformation may challenge long-held notions of teacher authority and expertise but also presents opportunities for educators to focus on higher-order aspects of teaching – fostering knowledge building, creativity, emotional intelligence, and ethical reasoning. By thoughtfully addressing the challenges and leveraging the unique capabilities of LLMs, we can work towards creating learning environments that truly empower students as active builders of knowledge.

References

- [1] Silvia García-Méndez, Francisco de Arriba-Pérez, and María del Carmen Somoza-López. A review on the use of large language models as virtual tutors. *Science & Education*, pages 1–16, 2024.
- [2] Benjamin S Bloom et al. Handbook on formative and summative evaluation of student learning. 1971.
- [3] Mike Perkins. Academic integrity considerations of ai large language models in the post-pandemic era: Chatgpt and beyond. *Journal of university teaching & learning practice*, 20(2):07, 2023.
- [4] Robin Schmucker, Meng Xia, Amos Azaria, and Tom Mitchell. Ruffle&riley: Insights from designing and evaluating a large language model-based conversational tutoring system. *arXiv preprint arXiv:2404.17460*, 2024.
- [5] Shen Wang, Tianlong Xu, Hang Li, Chaoli Zhang, Joleen Liang, Jiliang Tang, Philip S Yu, and Qingsong Wen. Large language models for education: A survey and outlook. *arXiv preprint arXiv:2403.18105*, 2024.
- [6] Marlene Scardamalia and Carl Bereiter. A brief history of knowledge building. *Canadian Journal of Learning and Technology/La revue canadienne de l'apprentissage et de la technologie*, 36(1), 2010.
- [7] Brent Gayle Wilson. Constructivist learning environments: Case studies in instructional design. 1996.
- [8] Michelene TH Chi and Randi Daimon Koeske. Network representation of a child’s dinosaur knowledge. *Developmental psychology*, 19(1):29, 1983.
- [9] Michelene TH Chi, Jean E Hutchinson, and Anne F Robin. How inferences about novel domain-related concepts can be constrained by structured knowledge. *Merrill-Palmer Quarterly (1982-)*, pages 27–62, 1989.
- [10] Camilla Gobbo and Michelene Chi. How knowledge is structured and used by expert and novice children. *Cognitive development*, 1(3):221–237, 1986.
- [11] Marlene Scardamalia and Carl Bereiter. Computer support for knowledge-building communities. *The journal of the learning sciences*, 3(3):265–283, 1994.

- [12] Robert C Pianta, Jay Belsky, Renate Houts, and Fred Morrison. Opportunities to learn in america's elementary classrooms. *Science*, 315(5820):1795–1796, 2007.
- [13] Andrew J Mashburn, Robert C Pianta, Bridget K Hamre, Jason T Downer, Oscar A Barbarin, Donna Bryant, Margaret Burchinal, Diane M Early, and Carollee Howes. Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child development*, 79(3):732–749, 2008.
- [14] Michael J Reddy. The conduit metaphor: A case of frame conflict in our language about language. *Metaphor and thought*, 2:164–201, 1979.
- [15] Kurt VanLehn, Hugh Burkhardt, Salman Cheema, Seokmin Kang, Daniel Pead, Alan Schoenfeld, and Jon Wetzel. Can an orchestration system increase collaborative, productive struggle in teaching-by-eliciting classrooms? *Interactive Learning Environments*, 29(6):987–1005, 2021.
- [16] Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günemann, Eyke Hüllermeier, et al. Chatgpt for good? on opportunities and challenges of large language models for education. *Learning and individual differences*, 103:102274, 2023.
- [17] Harm Tillema, Gert J van der Westhuizen, and Martijn P van der Merwe. Knowledge building through conversation. In *Mentoring for learning*, pages 1–19. Brill, 2015.
- [18] Phyllis C. Blumenfeld, Toni M. Kempler, Joseph S. Krajcik, and Phyllis Blumenfeld. Motivation and cognitive engagement in learning environments. In R. Keith Sawyer, editor, *The Cambridge Handbook of the Learning Sciences*, pages 475–488. Cambridge University Press, New York, 2006.
- [19] Michelene TH Chi and Ruth Wylie. The icap framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, 49(4):219–243, 2014.
- [20] Judith L Meece, Phyllis C Blumenfeld, and Rick H Hoyle. Students' goal orientations and cognitive engagement in classroom activities. *Journal of educational psychology*, 80(4):514, 1988.
- [21] Paul R Pintrich and Barbara Schragben. Students' motivational beliefs and their cognitive engagement in classroom academic tasks. Routledge, 2012.
- [22] Michelene TH Chi, Joshua Adams, Emily B Bogusch, Christiana Bruchok, Seokmin Kang, Matthew Lancaster, Roy Levy, Na Li, Katherine L McEldoon, Glenda S Stump, et al. Translating the icap theory of cognitive engagement into practice. *Cognitive science*, 42(6):1777–1832, 2018.
- [23] Megan Pedler, Suzanne Hudson, and Tony Yeigh. The teachers' role in student engagement: A review. *Australian Journal of Teacher Education (Online)*, 45(3):48–62, 2020.
- [24] Lev Semenovich Vygotsky and Michael Cole. Mind in society: Development of higher psychological processes. Harvard university press, 1978.
- [25] Carol Ann Tomlinson. The differentiated classroom: Responding to the needs of all learners. Ascd, 2014.
- [26] Mitzi Lewison, Amy Seely Flint, and Katie Van Sluys. Taking on critical literacy: The journey of newcomers and novices. *Language arts*, 79(5):382–392, 2002.
- [27] Cindy E Hmelo-Silver. Problem-based learning: What and how do students learn? *Educational psychology review*, 16:235–266, 2004.
- [28] Robin Alexander. Developing dialogic teaching: Genesis, process, trial. *Research papers in education*,

- 33(5):561–598, 2018.
- [29] Gordon Wells. Dialogic inquiry in education. *Vygotskian perspectives on literacy research*, pages 51–85, 2000.
- [30] Tiancheng Hu and Nigel Collier. Quantifying the persona effect in llm simulations. *arXiv preprint arXiv:2402.10811*, 2024.
- [31] Gloria Ladson-Billings. Toward a theory of culturally relevant pedagogy. *American educational research journal*, 32(3):465–491, 1995.
- [32] Gloria Ladson-Billings. Asking the right questions. *Studying diversity in teacher education*, pages 385–398, 2011.
- [33] Kathryn Au, Cathie Jordan, et al. Teaching reading to hawaiian children: Finding a culturally appropriate solution. *Culture and the bilingual classroom: Studies in classroom ethnography*, pages 139–152, 1981.
- [34] Gerald Mohatt, Frederick Erickson, et al. Cultural differences in teaching styles in an odawa school: A sociolinguistic approach. *Culture and the bilingual classroom: Studies in classroom ethnography*, 105, 1981.
- [35] CB Cazden and EL Leggett. Culturally responsive education: A response to remedies. *Culture and the Bilingual Classroom: Studies in classroom ethnography*, pages 69–86, 1981.
- [36] Frederick Erickson and Gerald Mohatt. Cultural organization of participation structures in two classrooms of indian students. *Doing the ethnography of schooling*, 132:174, 1982.
- [37] Geneva Gay. Preparing for culturally responsive teaching. *Journal of teacher education*, 53(2):106–116, 2002.
- [38] Luis Moll, Cathy Amanti, Deborah Neff, and Norma Gonzalez. Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. Routledge, 2006.
- [39] James WA Strachan, Dalila Albergo, Giulia Borghini, Oriana Pansardi, Eugenio Scaliti, Saurabh Gupta, Krati Saxena, Alessandro Rufo, Stefano Panzeri, Guido Manzi, et al. Testing theory of mind in large language models and humans. *Nature Human Behaviour*, pages 1–11, 2024.
- [40] Bronwen Cowie and Beverley Bell. A model of formative assessment in science education. *Assessment in Education: Principles, Policy & Practice*, 6(1):101–116, 1999.
- [41] Paul Black and Dylan Wiliam. Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of personnel evaluation in education)*, 21:5–31, 2009.
- [42] M Scriven. The methodology of evaluation (washington, dc, american educational research association). *ScrivenThe Methodology of Evaluation*1967, 1967.
- [43] James W Pellegrino. Assessment of and for learning. pages 410–421. Routledge, 2018.
- [44] D Royce Sadler. Formative assessment and the design of instructional systems. *Instructional science*, 18(2):119–144, 1989.
- [45] John Pryor and Barbara Crossouard. A socio-cultural theorisation of formative assessment. *Oxford review of Education*, 34(1):1–20, 2008.
- [46] David Carless, Gordon Joughin, and Magdalena Mok. Learning-oriented assessment: principles and practice. *Assessment and evaluation in Higher Education*, 31(4):395–398, 2006.

- [47] Jannette Elwood. Formative assessment: Possibilities, boundaries and limitations. *Assessment in Education: Principles, Policy & Practice*, 13(2):215–232, 2006.
- [48] Margaret Heritage. Formative assessment: What do teachers need to know and do? *Phi Delta Kappan*, 89(2):140–145, 2007.
- [49] Randy Elliot Bennett. Formative assessment: A critical review. *Assessment in education: principles, policy & practice*, 18(1):5–25, 2011.
- [50] Alexis A Lopez, Danielle Guzman-Orth, Diego Zapata-Rivera, Carolyn M Forsyth, and Christine Luce. Examining the accuracy of a conversation-based assessment in interpreting english learners’ written responses. *ETS Research Report Series*, 2021(1):1–15, 2021.
- [51] Diego Zapata-Rivera, Blair Lehman, Jesse R Sparks, Han-Hui Por, and Kofi James. Identifying and addressing unexpected responses in conversation-based assessments. *ETS Research Memorandum Series*, 18:13, 2018.
- [52] Carol L Barry, S Jeanne Horst, Sara J Finney, Allison R Brown, and Jason P Kopp. Do examinees have similar test-taking effort? a high-stakes question for low-stakes testing. *International Journal of Testing*, 10(4):342–363, 2010.
- [53] Hanna Eklöf. Skill and will: Test-taking motivation and assessment quality. *Assessment in Education: Principles, Policy & Practice*, 17(4):345–356, 2010.
- [54] Wynne Harlen and Mary James. Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in education: Principles, policy & practice*, 4(3):365–379, 1997.
- [55] Zahra Abbasiantaeb, Yifei Yuan, Evangelos Kanoulas, and Mohammad Aliannejadi. Let the llms talk: Simulating human-to-human conversational qa via zero-shot llm-to-llm interactions. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, pages 8–17, 2024.
- [56] Clayton Cohn, Nicole Hutchins, Tuan Le, and Gautam Biswas. A chain-of-thought prompting approach with llms for evaluating students’ formative assessment responses in science. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 23182–23190, 2024.
- [57] C Gordon and M Foucault. selected interviews and other writings 1972- 1977, 1980.
- [58] Lindsay L Cornelius and Leslie Rupert Herrenkohl. Power in the classroom: How the classroom environment shapes students’ relationships with each other and with concepts. *Cognition and instruction*, 22(4):467–498, 2004.
- [59] Indigo Esmonde and Jennifer M Langer-Osuna. Power in numbers: Student participation in mathematical discussions in heterogeneous spaces. *Journal for Research in Mathematics Education*, 44(1):288–315, 2013.
- [60] Niral Shah and Colleen M Lewis. Amplifying and attenuating inequity in collaborative learning: Toward an analytical framework. *Cognition and instruction*, 37(4):423–452, 2019.
- [61] Randi A Engle and Faith R Conant. Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and instruction*, 20(4):399–483, 2002.
- [62] Courtney B Cazden. Classroom discourse: The language of teaching and learning. ERIC-ED288206, 1988.

- [63] Hugh Mehan. Learning lessons: Social organization in the classroom. Harvard University Press, 1979.
- [64] Giyoo Hatano and Kayoko Inagaki. Sharing cognition through collective comprehension activity. 1991.
- [65] Mikhail Bakhtin. The dialogic imagination (trans. c. emerson and m. holquist). *Austin: Univer*, 1981.
- [66] Henry M Levin. A conceptual framework for accountability in education. *The School Review*, 82(3):363–391, 1974.
- [67] Sarah Michaels, Catherine O’Connor, and Lauren B Resnick. Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in philosophy and education*, 27:283–297, 2008.
- [68] Catherine O’Connor and Sarah Michaels. Supporting teachers in taking up productive talk moves: The long road to professional learning at scale. *International Journal of Educational Research*, 97:166–175, 2019.
- [69] Catherine A. Stein. Let’s talk: Promoting mathematical discourse in the classroom. *The Mathematics Teacher*, 101(4):285 – 289, 2007.
- [70] Arthur C Graesser, Kurt VanLehn, Carolyn P Rosé, Pamela W Jordan, and Derek Harter. Intelligent tutoring systems with conversational dialogue. *AI magazine*, 22(4):39–39, 2001.
- [71] Ryan Shaun Baker, Albert T Corbett, and Kenneth R Koedinger. Detecting student misuse of intelligent tutoring systems. In *Intelligent Tutoring Systems: 7th International Conference, ITS 2004, Maceió, Alagoas, Brazil, August 30-September 3, 2004. Proceedings 7*, pages 531–540. Springer, 2004.
- [72] Yrjö Engeström. Studies in expansive learning: Learning what is not yet there. Cambridge University Press, 2016.
- [73] Robert A Sottolare, Arthur C Graesser, Xiangen Hu, and Anne M Sinatra. Design recommendations for intelligent tutoring systems: Volume 6-team tutoring. volume 6. US Army Research Laboratory, 2018.
- [74] John H Flavell. Monitoring social cognitive enterprises: Something else that may develop in the area of social cognition. *Social cognitive development: Frontiers and possible futures*, 11:272–287, 1981.
- [75] Ann L Brown. Metacognition, executive control, self-regulation, and other more mysterious mechanisms. *Metacognition, motivation, and understanding*, pages 65–116, 1987.
- [76] John H Flavell. Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, 34(10):906, 1979.
- [77] Gregory Schraw and Rayne Sperling Dennison. Assessing metacognitive awareness. *Contemporary educational psychology*, 19(4):460–475, 1994.
- [78] Cindy E Hmelo-Silver and Howard S Barrows. Facilitating collaborative knowledge building. *Cognition and instruction*, 26(1):48–94, 2008.
- [79] Marlene Scardamalia and Carl Bereiter. Knowledge building: Theory, pedagogy, and technology. na, 2006.
- [80] Yue Zhang, Yafu Li, Leyang Cui, Deng Cai, Lema Liu, Tingchen Fu, Xinting Huang, Enbo Zhao, Yu Zhang, Yulong Chen, et al. Siren’s song in the ai ocean: a survey on hallucination in large language models. *arXiv preprint arXiv:2309.01219*, 2023.
- [81] Andrew J Peterson. Ai and the problem of knowledge collapse. *arXiv preprint arXiv:2404.03502*, 2024.

- [82] Elizabeth L Bjork, Robert A Bjork, et al. Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society*, 2(59-68), 2011.
- [83] Hiroko Kawaguchi Warshauer. Productive struggle in middle school mathematics classrooms. *Journal of Mathematics Teacher Education*, 18:375–400, 2015.
- [84] Office of Educational Technology U. S. Department of Education. Artificial intelligence and future of teaching and learning: Insights and recommendations, 2023.
- [85] Mei Tan and Hari Subramonyam. More than model documentation: uncovering teachers' bespoke information needs for informed classroom integration of chatgpt. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–19, 2024.