Google's Pedagogical Framework for Generative AI to Support Human Learning

Julia Wilkowski[1]
[1]Google

While contemporary generative artificial intelligence (genAI) tools are powerful information synthesizers, their default tendency toward immediate task completion can bypass critical elements, like cognitive friction, essential for effective knowledge construction. To bridge the gap between AI product development and research-based learning science, Google established a pedagogical framework designed to build genAI environments for deeper human learning. Grounded in constructivist and cognitive learning theories, this paper presents five foundational principles: encouraging active learning, managing cognitive load, stimulating curiosity and motivation, adapting to the learner, and deepening metacognition. We examine the theoretical underpinnings and practical considerations of each principle, contrast their execution across various learning technologies, and outline Google's evaluation methodologies for benchmarking pedagogical effectiveness. This framework provides a systematic, assessable roadmap to ensure AI tools scaffold deep understanding, foster learner agency, and cultivate durable skills for the future.

Keywords: education, evaluation, Gemini, LearnLM

1. Introduction

Generative artificial intelligence (genAI) tools are highly capable information synthesizers, but their default optimization for immediate user satisfaction often directly contradicts established pedagogical best practices. Because these models are largely trained to act as helpful assistants, completing tasks and providing direct answers, they bypass steps necessary for knowledge construction. These tools often behave as assistants, trying to complete actions for the user, instead of applying more pedagogical behaviors that encourage the learner to do the work. Assistant behavior is usually not aligned with what we know fosters great learning experiences. At the same time, the vast diversity of learners and contexts, spanning ages, subjects, culture, language, and goals, makes it impossible to create a single definition of "good pedagogy" [1, 2]. We set out to establish a set of foundational, research-based design principles to enable product teams and developers to create a shared language and vision for effective learning environments. These principles guide the development of AI models and tools as well as inform robust evaluation frameworks to measure whether products are meeting these pedagogical goals. These pre-launch evaluations serve as a design validation assessment: do the product's content and interactive behavior align with the learning sciences rationale that informed its design, indicating readiness for human pilot testing? In other words, do we believe the product is likely to lead to high-quality conditions for learning and positively affect learning outcomes? We apply these principles to tools, queries, or conversations that have an academic intent, or that are built for educators.

When genAI chatbots for the general public emerged in 2022, students immediately adopted these tools. While many students use them to further their own learning, some seek to take shortcuts of their education processes. The rapid proliferation of genAI in education contexts has exposed a fundamental tension between productivity and cognitive development. In their recent report,

Corresponding author(s): wilkowski@google.com

the Brookings Institution analyzed over 400 studies across 50 countries to conclude that unguided, general-purpose AI risks leading to "AI-diminished learning". Cognitive offloading and automated answer tools can undermine critical thinking, knowledge retention, and learner agency [3]. In this landscape, educators remain divided about student genAI tool use. The Rand Corporation surveyed a nationally representative sample of students in the United States and found that 62% of middle school, high school, and college students reported using genAI tools for homework at the end of 2025 [4]. While more than half of teachers surveyed report productivity gains in lesson planning and administrative workflows, over 55% express serious concerns about cognitive offloading, diminished critical thinking, and the erosion of authentic assessment [5]. Empirical research strongly reinforces this concern: controlled studies at Wharton and Harvard revealed that while students using unguided LLMs performed 48% better during AI-assisted tasks, their scores dropped by 17% on in-class exams [6, 7]. Conversely, when AI tools are engineered with deliberate "pedagogical friction" (refusing direct answers in favor of dialogue-based hints and step-by-step scaffolding), student performance increased by 127% with sustained retention [8].

Frameworks developed by governments and non-governmental organizations largely offer high-level ethical guardrails and policy taxonomies rather than operational, assessable pedagogical standards [9, 10, 11, 12]. This creates a critical design and evaluation gap: product teams lack a systematic, empirically validated framework to ensure AI learning technologies scaffold deeper understanding rather than superficial answer retrieval. Google's pedagogical framework addresses this challenge by translating learning science principles into concrete heuristics for designing, benchmarking, and evaluating student-centered AI tools at scale.

The Learning Science team at Google focuses on two key areas: first, closing the gap between academic research and product development by guiding product teams to create learning environments informed by evidence; and second, evaluating the effectiveness of our products in fostering learning impact. This document concentrates on the former, laying the groundwork for our approach to the latter. Before we could assess the impact of Google's launched products on teaching and learning, we needed a design readiness metric to inform product development. The evaluation rubrics derived from these principles served as this quantifiable metric. We standardized on a set of principles that we could use for developing and evaluating whether it was likely for our products, models, and features to create high-quality conditions in which learning was likely to occur. Although we have integrated these principles into features designed for both learners and educators, we believe that the most substantial learning outcomes will occur when educators guide students in using genAI tools.

In this paper we explain the research background that informed the development of Google's learning science principles, how we believe those principles create sufficient conditions for learning, how we apply those principles in our products, and how we evaluate the quality of our products from a pedagogical perspective.

2. The Five Principles

We attempted to distill a century of learning science research into give core, actionable principles that we could use to build and evaluate product/model behavior. These principles represent research-based design principles that are likely to create opportunities for learning in the vast amount of learning contexts worldwide. These principles now guide product teams in developing learning-focused features, including foundational model capabilities like LearnLM (now infused into Gemini core models). They also serve as the basis for our evaluation framework, helping assess the likelihood of a learner acquiring knowledge when interacting with a given feature. These principles primarily draw upon constructivism (Piaget) and cognitive learning theories, and we aim to incorporate elements of experiential learning (Dewey), social constructivism (Vygotsky, Bandura), and constructionism (Papert) [13]. From Vygotskian social constructivism, we draw the foundational concept of mediated learning: that cognitive development occurs through dialogic, assisted performance, where an instructional partner provides graduated scaffolding and contingent feedback within the learner's Zone of Proximal Development before skills are internalized independently [14]. While these principles derive from empirical classroom research, translating them into genAI environments requires deliberate adaptation, as instructional techniques that are effective in human-led classrooms do not transfer automatically to computer-based learning environments [15].

The principles we present below were selected based on several key criteria.

- They are backed by empirical evidence showing they contribute to high-quality learning experiences.
- They are well-suited for products that reach learners directly or via educators.
- They apply broadly across multiple disciplines and geographies.
- They can be effectively implemented within technology-forward learning environments.
- They describe education tool behavior (not necessarily learner behaviors or outcomes).

The Cambridge Handbook of the Learning Sciences mentions that the literature has converged "on a small number of overarching principles for facilitating learning: the importance of repetition and practice; managing the demands on cognitive and attentional resources; engaging the learner in the active construction of meaning and knowledge; and metacognitive awareness" [13, p. 30]. Google's similar principles for product and model behavior were distilled from a thorough reading of research and focus on what we can deliver through AI-generated learning tools.

- **Encourage active learning:** Helping learners actively construct understanding through pedagogical approaches like retrieval practice, dialogic inquiry, and problem-solving: tasks in which the learner makes connections, explains concepts, or creates artifacts.
- **Manage cognitive load:** Directing learner cognitive efforts on the focal learning goal while managing distractions unrelated to the learning goal; using well-structured segments and purposeful dual coding to support working memory and schema construction.
- **Stimulate curiosity and motivation:** Fostering perceived task value and expectancy of success by anchoring new concepts in prior knowledge, authentic real-world contexts, and learner goals.
- **Adapt to the learner:** Assessing what the learner knows and can do, providing calibrated challenges and fading scaffolds to build proficiency, and changing subsequent interactions to build on what the student knows while addressing areas of growth; helping the learner achieve their goals while offering the autonomy to explore.
- **Deepen metacognition:** Providing explicit opportunities for learners to set goals, plan strategies, monitor comprehension, and evaluate their learning process.

These principles serve as a flexible yet structured foundation for designing and evaluating education technology that augments and supports learning.

3. Encourage active learning

3.1. Definition of the principle

Active learning is an educational approach that requires learners to manipulate and construct information rather than receiving it passively. We use two primary frameworks of learning: the ICAP framework, which outlines four modes of interacting with content (interactive, constructive, active, passive) [16] and Bloom's taxonomy, which outlines six types of cognitive processes (remember, understand, apply, analyze, evaluate, create) [17]. We believe we can provide opportunities for high quality learning experiences where learners actively construct knowledge and apply it at all levels of the taxonomy.

3.2. Summary of relevant literature

Fundamentally, active learning requires a learner-centered approach. It shifts the educational focus from the teacher as a knowledge transmitter to the student as a participant in building their own knowledge and skills. Research in learning science indicates that meaningful, transferable learning requires learners to engage deeply with content. Active learning is largely rooted in constructivist learning theory, which posits that individuals construct their own knowledge by connecting new information to existing understanding and schema [13, 18].

The ICAP framework differentiates modes of engagement (Passive, Active, Constructive, Interactive) based on overt behaviors, predicting that learning increases as engagement moves from passive to interactive [16]. Critics of the ICAP framework note that it is not possible to see learners' cognitive processes in order to know whether or not a student is engaging actively in a learning experience based on overt behaviors; they advocate for formative assessment as the alternative to assess cognitive progress [19]. Either way, technology-mediated learning experiences can foster active engagement across a range of activities, helping learners move beyond passive reading to discussing, summarizing, recalling information, and reflecting on how new knowledge connects to personal experience and prior knowledge. Engaging with new information through various methods like discussion, practice, inquiry, and creation is more likely to lead to transfer and longer-term learning [16]. Ideally these approaches emphasize higher-order thinking through guided inquiry, simulations, and experiments, using retrieval practice not as an isolated endpoint, but as a mechanism to support deeper conceptual construction and application [17]. As cognitive scientist Daniel Willingham notes, "Memory is the residue of thought," [20] emphasizing that durable learning necessitates effort. Techniques to enhance active learning in text-based interactions include engaging in activities with desirable difficulty, including spaced repetition, retrieval practice, dual coding, concrete examples, interleaving, and elaboration [21, 22].

Encountering new information requires students to make connections with prior knowledge. While learners frequently assimilate new information smoothly into existing schema, encountering counter-intuitive concepts can trigger Piagetian disequilibrium (or conceptual conflict). Navigating these moments can require calibrated, productive struggle supported by targeted instructional scaffolding [23, 24, 25, 26, 27, as cited in 28]. Effective learning experiences minimize unhelpful disequilibrium while providing ample opportunities for this productive struggle.

While active learning is a broad instructional design principle, conversational AI tools can operationalize social-constructivist mediation by structuring turn-by-turn assisted performance within the learner's Zone of Proximal Development (ZPD) [14]. Rather than treating "scaffolded inquiry" as an abstract label, system instructions specify concrete, checkable dialogic behaviors including but not limited to withholding direct solutions; decomposing complex, multi-step problems into manageable sub-goals; delivering graduated hints contingent on the learner's specific error; and fading support as the learner demonstrates independent competence [29].

Research has consistently demonstrated that active learning improves student outcomes, leading to enhanced critical thinking skills, increased retention and transfer of new information, higher motivation, improved interpersonal skills, and reduced course failure rates. A comprehensive meta-analysis of 225 studies across undergraduate science, technology, engineering, and mathematics (STEM) courses found that active learning increased student performance on exams and concept inventories by an average of 0.47 standard deviations [30]. In an analysis of a massive open online course (MOOC), researchers found that students who engaged in interactive "learning by doing" activities learned significantly more than those who passively watched video lectures or read text without embedded generative tasks [31]. Recent research demonstrates that graduate students who engaged actively with content in an online course (projects, quizzing, problem-based assignments) demonstrated stronger learning outcomes, critical thinking, and attitudes about science than students who did not engage in these course options [32] [1].

3.3. Considerations for genAI models and learning features

As we build genAI-powered learning tools, we should encourage learners to engage actively with the content in a variety of ways (including activating prior knowledge, encouraging summarizations, fostering debate, evaluating reasoning, and creating new artifacts). We must also contend with the fact that they may or may not be willing to engage, especially in the absence of a human to motivate active behavior. While we believe that educators should be in the lead when students use AI tools, we also encourage learners to engage with content in rigorous, proven ways.

A key challenge is shifting the model's default behavior from "information provider" to "learning facilitator." This is addressed by training models for pedagogical instruction following [34]. By using specific system instructions like "ask probing and guiding questions" or "don't solve problems for your students, but offer scaffolds," developers can guide the model to behave more like a dialog-based learning partner that provides scaffolded inquiry. Significant care needs to be put into identifying when those scaffolds are necessary, when they can start to be removed, and what specific questions and behaviors the tool should try to identify in order to scaffold thoughtfully.

3.4. Assessing active learning

Leading up to launch, we validate the system's design and tool behavior through rigorous automated and human evaluations. After we have gained confidence in tool and model behavior, we move to piloting with learners, followed by testing at scale with more learners and educators.

When evaluating the capabilities of LearnLM, Google's capabilities that power multiple features and products, educators evaluated a variety of scenarios where learners interacted with the chatbot interface. They sought evidence that the model behaved in accordance with active learning best practices.

- **Opportunities for engagement:** The tutor[2] provides opportunities for engagement from the student.
- **Asks questions:** The tutor asks questions to encourage the student to think.

[1] There is often an effect where students initially perceive that they learn less due to the increased cognitive effort required, their actual learning is often greater in active learning environments [33]. This makes it difficult to ask learners to self-assess their learning, as they may not be the best judge. This also presents problems for learners using tools for academic support outside the classroom or when not directed by an educator: they may be unwilling to engage in tasks that are difficult or that are perceived not to lead to learning, even when well-designed.

[2] Note that the use of "tutor" is copied directly from the LearnLM technical reports. We are therefore using this terminology but advocate for this term to be replaced with "AI tool". While our evaluation criteria adopt the benchmark persona of "the tutor" to assess specific dialogic behaviors (e.g., asking probing questions, providing formative feedback), Google's framework centers generative AI as a conversational learning partner designed to augment human instruction, not a substitute for human 1:1 tutoring.

- **Guides to answer:** The tutor does not give away answers too quickly.
- **Active engagement:** The tutor promotes active engagement with the material [35].

The quality of opportunities the genAI tool provides to engage at different levels of Bloom's taxonomy are also assessed:

- **Recall & comprehension:** the AI tool encourages the learner to recall information and demonstrate comprehension.
- **Analysis & application:** the AI tool engages the learner in activities to analyze and apply skills.
- Evaluation **& creation:** the AI tool engages the learner to evaluate or create content.

All of these rubric dimensions work together to assess how well the model or genAI tool attempts to engage the learner in interacting with information, making connections, and therefore constructing knowledge.

3.5. Examples of active learning in Google's products

Many of the examples of active learning launched in Google's products center on the lower levels of Bloom's taxonomy (such as recall-based quizzes).

- **Gemini Guided Learning mode** provides multiple opportunities for learners to actively engage with content, from answering simple questions to completing scenarios relevant to the topic at hand.
- **Gemini Study Notebooks** and **Quizzes in Search AI Mode** enable learners to upload documents, course materials, and notes then take quizzes based on these materials, fostering active recall.
- **Gemini Notebooks (formerly NotebookLM)** foster active learning through studio artifacts like quizzes and flashcards, along with Interactive Learning Overviews.
- **Classroom Choice Boards Generator** enables educators to rapidly create multiple ways for learners to engage with and represent, and express their understanding of concepts.
- **YouTube Ask** enables learners to ask questions about the video they're watching, or to receive quizzes about the video to assess how well they are comprehending video concepts.

4. Manage cognitive load

4.1. Definition of the principle

Two primary factors of technology-facilitated instruction influence learning: how content is presented, and how learners interact with that information. The second is addressed above in the active learning section. This section addresses how we can present instructional material effectively to help students incorporate information into existing schemas so it can be retrieved in novel contexts. "The total information-processing demands imposed by a given task or set of tasks — also known as cognitive load — can easily exceed what people can manage. When people's limit is exceeded, they are left with insufficient attention and other cognitive resources to complete the task effectively" [36].

4.2. Summary of relevant literature

The brain processes information when we pay attention to, and interact with it; these are prerequisites for moving information to longer-term storage [20, 22]. The process of temporarily holding and manipulating information replies on working memory, which has a limited capacity. Cognitive load theory [37] is based on the premise that overloading working memory reduces the effectiveness of teaching and learning.

Richard E. Mayer provides evidence-based principles for designing multimedia learning experiences that manage cognitive load well. This cognitive theory of multimedia learning is based on several assumptions: humans have a limited capacity for processing information (working memory), and we must actively process information [38]. Mayer also builds on the dual-coding theory that we process information via multiple channels including auditory/verbal and visual/pictorial [39].

Sweller explains that multiple simultaneous processing tasks require significant working memory, which leads to intrinsic or extraneous load [40]. Sweller identifies three primary types of cognitive load:

- **Intrinsic load:** the inherent difficulty of the subject, topic, and task. This also depends on the learner's prior knowledge, affect, and aspects of the learning environment. Intrinsic load is determined by element interactivity and the need for simultaneous processing. Intrinsic load cannot be modified without changing the goals or scope of the learning experience [37, 41].
- **Extraneous load:** non-essential load caused by how information is presented (e.g. distracting graphics, confusing layout, irrelevant information).
- **Germane load:** the effort required for deep cognitive processing and constructing mental models (schemas) in long-term memory.

Google's products for learning aim to direct cognitive load demand toward the focal learning goal while minimizing extraneous load. A learner's prior knowledge can determine the intrinsic load of instructional material. Once concepts are consolidated into schemas in long-term memory, working memory can process a complex system as a single unit [37]. Scaffolding is another technique that can relieve extraneous cognitive load and enable the learner to focus on the most important elements, with the goal of fading this support as the learner becomes more proficient [37].

4.3. Considerations for genAI models and learning features

In conversational AI, managing cognitive load requires addressing the medium's primary failure mode: the AI "wall of text." Language models tend to default to dense, multi-paragraph explanations that impose extraneous reading load. Effective cognitive load management in genAI relies on two concrete mechanisms: 1) response chunking and turn-by-turn pacing, and 2) strategic cognitive offloading of routine sub-calculations to ensure learners' cognitive effort stays focused on the focal learning goal. We acknowledge that we have not fully solved this problem yet. To reduce extraneous load, we use many of Richard Mayer's techniques: the coherence principle (reducing extra, unnecessary, or irrelevant materials); the signaling principle (guiding learners' attention to essential material by highlighting important details); the redundancy principle (avoiding redundant information); and spatial and temporal contiguity principles (placing text with images or relevant narration) [38]. We also apply many of these techniques to manage the intrinsic load of instructional materials: segmenting information into smaller chunks; pre-training to introduce concepts before exploring them thoroughly; and the modality principle (presenting words as spoken text, as in videos). We attempt to also calibrate germane load for schema building by personalizing (presenting language in a conversational tone over a formal tone) and information in multimedia (presenting words and pictures together).

One of the ultimate goals of teaching is to help novices develop schemas, or organizing structures in long-term memory. We seek to promote germane load through the techniques mentioned in the active learning section above. Scaffolding is a primary way that novice learners can interact with challenges and content that may be too difficult for them [42]. Additionally, the tool can apply instructional strategies that are well-matched to the learning goal. Similarly, the tool can provide analogies (only helpful when related to learners' prior knowledge). While we recognize that some techniques that do not apply equally well for novices and experts (expertise reversal effect [43]), we build most often for novices, with the theory that experts are more able to adjust responses to meet their level of need.

AI tools should be designed to scaffold information effectively. Some techniques apply broadly across learners, such as structuring text with clear signals (headings, bullet points) and purposeful dual coding. Pairing a concise verbal explanation with a single, aligned visual representation can reduce extraneous load without overwhelming the learner. Other techniques are more subjective and situation-dependent, such as adjusting the amount of scaffolding or the examples used to make a concept more relatable. AI tools can explicitly ask the learner for feedback or interpret verbal cues, such as a learner's request to "go back" or "explain that again," to detect understanding or confusion. By dynamically modulating hints and explanations in response to these signals, the conversational learning partner can try to keep the student operating within their ZPD [14]. The AI tool can provide the temporary scaffolding necessary for the learner to successfully solve problems and construct understanding that they could not yet achieve unassisted. By doing so, the dialog-based learning assistant can adapt accordingly, providing calibrated challenges to ensure the learner operates within a region of proximal learning without inducing cognitive overload [37, 44].

4.4. Assessing cognitive load management

When we evaluated the LearnLM capabilities (that now power multiple features and products), educators evaluated a variety of scenarios where learners interacted with the chatbot interface. They assessed how well the tools applied the design principle of managing cognitive load.

- **Appropriate response length:** The tutor's responses are an appropriate length for the student.
- **Manageable chunks:** The tutor uses bullet points and other formatting to break information down into smaller, manageable chunks.
- **Straightforward response:** The tutor responses are clear and easy to follow.
- **No irrelevant information:** The tutor avoids irrelevant information.
- **Analogies:** The tutor uses narratives, case studies, or analogies[3] to illustrate key concepts.
- **Information presentation:** The tutor presents information in an appropriate style and structure.
- **Information order:** The tutor develops information in a logical order, building on previous concepts.
- **No repetition:** The tutor avoids repeating information unnecessarily.
- **No contradiction:** The tutor avoids contradicting information from earlier parts of the conversation [35].

Depending on the feature, evaluators may use also rubrics that include additional components:

- **Redundancy & repetition:** The AI tool avoids redundancy or repetition.
- **Complementary representations:** The AI tool presents information using purposeful, complementary representations.
- **Instructional strategies:** The AI tool presents information using well-matched instructional strategies.
- **Content presentation:** The AI tool presents content in a chunked, simple, logical way.

[3] Analogies can be used most effectively when they connect with students' prior knowledge.

A combination of all of these dimensions are used to assess how well Google's features for teaching and learning help manage cognitive load.

4.5. Examples of cognitive load management in Google's products

- **Solvers in Search:** Break down complex problems into bite-sized, step-by-step instructions that the user can follow at their own pace.
- **3D models in Search:** Allow for self-paced exploration of complex concepts, where users can click on different parts to get information in manageable chunks.
- **YouTube:** Features like chapters and moments act as signals, helping users focus on the most essential material and navigate complex information.
- **Guided Learning in Gemini and Classroom:** Offers step-by-step support with procedural STEM problems, shorter responses featuring a chunked section of longer topics, and dynamic scaffolding to help users engage with complex tasks and queries.
- **Gemini App:** Education queries often incorporate visuals alongside text, adhering to the multimodality principle, to help explain abstract and complex topics.

5. Stimulate motivation and curiosity

5.1. Definition of the principle

Why do humans of any age do anything? An internal or external force compels us to act. Motivation is an external or internal force that directs and sustains behavior toward a goal. We might be motivated **intrinsically**, by our own curiosity, desire to learn, excitement to explore, or to build a better life. We might be motivated by others, who help us see the importance of what we're learning, or to be part of a community of learners, or to contribute to the local culture or society. We might be motivated extrinsically, by an external factor such as grades, monetary rewards, or other benefits for completing a task, assignment, project, or work. "Motivation affects the amount of time that people are willing to devote to learning. Humans are motivated to develop competence and solve problems" [45].

5.2. Summary of relevant literature

Learning science supports fostering intrinsic motivation, as it can lead to deeper and more sustained engagement. Self-determination theory posits that three characteristics are required for optional functioning, growth, and wellness: autonomy, competence, relatedness [46].

Expectancy-value theory centers on the idea that motivation for a specific activity or goal is determined by two factors: expectancy and value. This means in order to engage with a task, an individual must have a reasonable expectation that they will succeed at the task, and that it is worthwhile for them to complete the task [47, 48].

ARCS theory of motivation is grounded in expectancy-value theory and provides a systematic approach to incorporate motivational strategies into educational activities. It consists of four key components: attention (stimulating interest and curiosity), relevance (to learners' lives and the real world), confidence (learners have a reasonable expectation that they can complete the task), and satisfaction (intrinsic and extrinsic rewards) [49].

We have summarized this research into several key recommendations for product teams, against which we assess how well our features perform. Strategies include connecting new experiences to prior knowledge, linking content to real-world applications, offering calibrated challenges to foster choice and perseverance aligned with learners' region of proximal learning [44].

5.3. Considerations for genAI models and learning features

Generative AI tools can connect content with learners' goals and their intrinsic motivation. It can also try to make connections to lived experiences by offering real-world content (realistic or actual scenarios, case studies, data, examples, images) and opportunities to apply skills that mirror everyday life. It can also provide a variety of assignments based on the same underlying content and learning objective but that are at the right level of difficulty for the learner. We are skeptical that we can turn extrinsically motivated learners into intrinsically motivated learners, but it's possible to try to tap into a learner's existing motivation and connect learning to their goals.

There is a lot of interest in gamification for education, with the idea that if people are motivated to engage in a learning activity when they aren't otherwise motivated, they will engage. While this theory has some merits, we are cautious to not implement gamification elements for the sake of getting people to use a product, or to increase the use of the product with adverse outcomes. We need to ensure that gamified elements are closely connected with learning objectives and not based on superficial rewards, particularly in the classroom [50]. It seems that extrinsic rewards can harm intrinsic motivation, where learners' initial interest in the task is replaced by their desire for the extrinsic reward; external rewards may also undermine learners' perceptions of autonomy and control [45].

5.4. Assessing curiosity and motivation

When evaluating the capabilities of LearnLM, Gemini's capabilities that power multiple features and products, educators evaluated a variety of scenarios where learners interacted with the chatbot interface. We sought evidence that the model and tools behaved in ways that would foster curiosity and motivation.

- **Stimulates interest:** The tutor tries to stimulate the student's interest and curiosity.
- **Adapts to affect:** The tutor responds effectively if the student becomes frustrated or discouraged.
- **Encouraging feedback:** The tutor delivers feedback (whether positive or negative) in an encouraging way [35].

Depending on the feature, evaluators may also use rubrics to assess additional elements of fostering curiosity and motivation.

- **Encouragement:** The AI tool provides goal-directed, task-specific feedback.
- **Manageable challenges:** The AI tool presents challenges that are at the appropriate level of difficulty (not too easy, not too difficult).
- **Relevance:** The AI tool presents real-world application of concepts, explains why it matters, and helps the learner identify the usefulness and relevance of the content to their life, goals and/or interests.
- **Autonomy:** The AI tool enables learners to shape their learning journey (choice of topics, learning activities, adjusting pace and difficulty).

A combination of all of these dimensions are used to assess how well the features are likely to foster motivation and curiosity in our learners.

5.5. Examples of curiosity and motivation in Google's products

- **Learn Your Way** can anchor abstract concepts in authentic, real-world contexts and connect to learners' existing prior knowledge, while offering relatable examples.

- **Learn About (Labs)**[4] provides callouts to "why it [the topic] matters" throughout a concept exploration.
- **Classroom Real World Connections Gem** identifies opportunities for authentic connections in educators' lessons, assignments, and assessments.
- **Gemini Guided Learning** poses questions meant as invitations to further exploration and learning (especially after an information-seeking, "quick fact" kind of query).

6. Adapt to the learner

6.1. Definition of the principle

The central aim of many education technology tools, as well as classroom instruction, is to offer individualized learning experiences to meet learners where they are and help them achieve their goals [51]. Crucially, these adaptive systems must preserve learner agency by giving learners meaningful autonomy to direct their inquiry, select modalities, and co-construct their study plans rather than subjecting them to opaque algorithmic tracking and decisions made entirely for them [52, 53]. Education technology products can help learners chart a journey from where they are today to achieve their task or longer-term goal with various levels of learner input and control. We can adapt foundational classroom strategies (e.g. differentiation, mastery learning or criterion-referenced progression with formative feedback, Universal Design for Learning) to scaffold demonstrated proficiency in digital environments. (Note: While this paper operationally distinguishes within-session "adaptation" from cross-session "personalization" for product design clarity, both rely on learner modeling across what Intelligent Tutoring Systems literature defines as the inner loop (step-level feedback and hints within a task) and the outer loop (selecting the next task or problem across a larger learning experience like a curriculum) [54].

6.2. Summary of relevant literature

Overall, research studies are mixed on which personalization and adaptive technologies lead to the best learning outcomes, with many systems requiring full implementation and having a large discrepancy in different populations. The expectations and reported outcomes of personalization are generally positive regarding learning experiences and comprehension but there are nuanced and varied effects on academic performance. Most research is based on students using these with instructional content outside of their regular classroom instruction. This limits an understanding of longer term implications. Notably, when Hattie [55] examined the effect of individual instruction on learning outcomes, he found a lack of evidence that this strategy is effective in classrooms, with more benefit from the presence of feedback at the right level of instruction, among other factors. Research is needed to fully understand the conditions under which personalization is most effective.

Some research indicates that adaptive AI systems can positively influence student outcomes. A mixed-methods study involving 300 students and 50 educators across primary to higher education found a substantial improvement in student performance, with average post-assessment scores increasing from 68.4 to 82.7 for students using adaptive learning systems that leveraged AI (compared to a 7.8 point improvement for those using traditional methods) [56]. While these studies show promising immediate benefit, the effect on information retention remains to be seen. A meta-analysis compiling 45 independent studies found that AI-based adaptive learning generated a medium-to-large positive effect size ($g = 0.70$) in favor of students' cognitive learning outcomes compared to non-adaptive learning interventions [57]. Another meta-analysis of 36 studies from 31 published

[4] This core functionality may be incorporated into other products and features in the future.

papers indicates that AI-enhanced adaptive learning experiences can have a modest effect on students' cognitive proficiency and emotional development [58]. Another meta-analysis of Intelligent Tutoring Systems found more mixed results, with generally positive results when systems are designed with thoughtful pedagogical practices; but these systems are difficult to generalize [59].

Significant research and theories exist about how to select the next challenge for learners (outer loop). Research indicates desirable challenges are beneficial for learners; encouraging learners to engage in activities that require more thinking or effort (such as practicing recalling information) lead to better outcomes than activities that require lower effort (such as reading, highlighting, or watching a video) [16, 21, 44]. A number of other research studies have demonstrated the value in providing tasks with just the right amount of difficulty, in diverse contexts like infant attention [60] and practicing motor skills [61]. Empirical studies of error-driven learning demonstrate peak learning rates under specific experimental conditions when tasks are calibrated so learners succeed approximately 85% of the time [62]. Systems that think about the inner loop (feedback) and outer loop (problem selection) of discrete and larger problem solving journeys can also help lead to more effective learning outcomes [54, 63].

6.3. Considerations for genAI models and learning features

GenAi tools can foster learning experiences that are tailored to their skills, needs, and goals. These customized environments, which also hold learners to grade-level expectations and standards, will lead to more effective and satisfying experiences than a tool that does not take these factors into account. The best kind of personalization involves having the learner drive their own learning experience, telling the tool what they want, how they want it, and adjusting the response to fit their needs (which requires a significant degree of metacognitive skills). In absence of the learner making all the choices, we can provide a high-quality customized environment based on information that we infer and receive explicitly about the learner. The tool should provide transparency about automated choices that are being made on their behalf. Because there are downsides to over-personalizing learning experiences, we should ensure that we are thoughtfully using data and not providing too many scaffolds for learners.

Effective personalization requires three components [52]: a student model (or learner profile); a domain model; and a pedagogical model. Aspects of the learner profile could include cognitive, motivational, socio-cultural, and affective variables. A domain model consists of knowledge components that could include subjects, units, chapters, concepts, skills, and the edges between them. The pedagogical model consists of ways the AI tool changes to meet the learner's current and desired future state. In a single-session interaction, adaptive interventions include offering targeted hints (inner loop), adjusting task pace, or providing relatable real-world examples (outer loop). Here, learning science provides an essential distinction: while automated problem-leveling algorithms select unassisted tasks within a student's Region of Proximal Learning [44], dialogic AI tools may be able to support the learner directly within their ZPD [14]. By providing dialogic assisted performance, the AI tool offers graduated hint sequences and step-level scaffolding that fades as the learner demonstrates increasing proficiency. However, unlike a human More Knowledgeable Other, current generative AI models and tools cannot autonomously perform metacognitive monitoring of a student's internal mental state. Therefore, effective adaptation requires explicit check-in prompts, learner feedback, and learner co-regulation (which emphasizes the need to build metacognitive skills in learners). In multi-session journeys, saved information might be used to help a learner plan a longer journey, demonstrate progress over time, and space repetition for practicing tasks. Signals about the learner can be collected implicitly and explicitly to gain a picture of the learner in order to adapt aspects of the learning environment such as difficulty, pace, task selection, and motivational strategies. Ideally the tool will acknowledge learners' experiences, cultures, and backgrounds (funds of knowledge) to help learners connect more deeply to the content they are learning [64].

We recognize that there are a number of risks when personalizing learning environments, including technical challenges (e.g. modeling complexity and resource constraints, data quality, systematic biases, aberrant student behavior); pedagogical challenges (narrowing educational focus, reductionism, losing sight of bigger educational objectives, loss of student control, dependency, inhibition of productive struggle, and over-reliance on a narrow set of interests for examples); social and equity risks (exacerbation of inequality due to the global digital divide, overburdening self-regulation skills); and ethical and privacy concerns (e.g. surveillance (or the perception thereof)).

6.4. Assessing adaptation

When evaluating the capabilities of LearnLM, Gemini's capabilities that power multiple features and products, educators evaluated a variety of scenarios where learners interacted with the chatbot interface. We tried to find evidence that the model was behaving in accordance with the need to assess, adapt, and create a short-term and longer-term plan to help learners effectively.

- **Leveling:** The tutor's explanations are appropriate for the level of the student.
- **Unstuck:** The tutor effectively adapts its approach to help the student when they are stuck[5].
- **Adapts to needs:** Overall, the tutor adapts to the student's needs[6].
- **Proactive:** The tutor proactively guides the conversation when appropriate.
- **Guides appropriately:** The tutor does not withhold information unproductively [35].

Depending on the feature, evaluators may use rubrics that also assess

- **Proactivity:** The AI tool proactively helps the learner get from their current state to their goal
- **Responsiveness:** The AI tool responds to a learner's implicit needs
- **Explicit requests:** The AI tool responds to a learner's stated requests (e.g. to modify the approach or learning plan)

A combination of all of these dimensions are used to assess adaptation and personalization in our features.

6.5. Examples of adaptation in Google's products

For most of our development process, we have only been able to adjust during a single interaction, though we are experimenting with what kind of information we can and should save to customize learning environments and experiences. In a single session, the AI tool might

- notice learner difficulties and offer scaffolded support
- adjust next tasks to keep learners in their ZPD
- ask about the student's grade level or prior knowledge to tailor its explanations
- provide a set of test preparation questions aligned with the learner's study guide or uploaded course document

[5] Learners may be stuck for different reasons that require different strategies, such as confusion (responding with a clarifying question or worked step, for example), frustration (responding by recalibrating task difficulty or offering a different entry point), or disengagement (suggesting that the student may be ready to take a break).

[6] Ideally, learner strengths should also be considered.

- provide calibrated practice problems with tiered scaffolding based on user performance, while maintaining grade-level standards
- offer users choices in content format (e.g., "Do you want to watch a video, read an explanation, or try a practice problem?")
- offer a range of content-specific entry points into a topic

Because adaptation and personalization require information about the learner's current understanding, we also support adaptation by making it easier for teachers to access and synthesize information about their learners. To support educators' differentiation initiatives, we provide insights through Google Classroom about students' performance on assignments and assessments.

7. Deepen metacognition

7.1. Definition of the principle

Metacognition is "thinking about thinking". This higher-order skill involves two key dimensions:

1. **Metacognitive knowledge:** Awareness of one's own cognitive processes, task demands, and strategy repertoire.
2. **Metacognitive regulation:** Using this knowledge to plan, monitor, reflect on, evaluate, and adjust one's learning process.

Metacognition is a crucial skill for lifelong learners, not only because it improves academic performance and enables skills transfer from one situation to another but also enables learners to become more self-directed and effective when learning new concepts and skills.

7.2. Summary of relevant literature

Metacognition has been revealed to be an essential component of learning. A review study by Wang, Haertel, and Walberg identified metacognition as one of the areas with the strongest influence on learning [65]. Hattie's Visible Learning project found that metacognitive strategies have an effect size of 0.53 [55, 66]. Metacognition (specifically, the capacity to retrieve and deploy a strategy in a similar but new context) is what enables a student who has been taught a particular strategy in one problem context to utilize it in another. Metacognitive skills can help students transfer what they have learned from one context to the next.

Winne and Azevedo [67] describe how metacognition is applied in different forms of knowledge identified by cognitive scientists, including declarative and procedural. Declarative metacognition includes beliefs about self-efficacy, knowledge, tasks, applicability of strategies and the probability they will be successful; they note, "A core characteristic of metacognitive declarative knowledge is that declarative knowledge can be verbalized" [67, p. 65]. Just having declarative metacognitive knowledge does not guarantee that the learner will apply the most appropriate learning strategies, however, due to a lack of motivation, capability, or other factors. Metacognitive declarative knowledge does not predict learning outcomes absent the learner acting upon this knowledge. Procedural metacognition can be scaffolded for novice problem solvers or people learning procedural skills to explicitly externalize the procedures. For example, students can articulate Polya's steps while solving procedural problems [68].

Flavell mentions four key classes of cognitive phenomena: a) metacognitive knowledge; b) metacognitive experiences; c) goals (or tasks); and d) actions (or strategies) [66]. Winne and

Azevedo also make three distinctions in the form of thinking involved: a) monitoring; b) control; c) self-regulated learning [67]. Zimmerman [69, p. 13-14] found that higher-achieving students applied more self-regulatory practices than lower-achieving students. "Self-regulated learning theories of academic achievement are distinctive from other accounts of learning and instruction by their emphasis (a) on how students select, organize, or create advantageous learning environments for themselves and (b) on how they plan and control the form and account of their own instruction.... These self-regulated students are distinguished by their systematic use of metacognitive, motivational, and behavioral strategies; by their responsiveness to feedback regarding the effectiveness of their learning; and by their self-perceptions of academic accomplishment" [69, p. 13-14]. All of these aspects of self-regulated learning depend on students being open to learning about them and applying them to their cognitive processes.

7.3. Considerations for genAI models and learning features

Metacognitive strategies can be developed by teaching students the value of self-monitoring and structuring interactions through a gradual release of responsibility [70, 71]. This progression operationalizes Vygotskian scaffolding: the AI first models metacognitive planning and error detection ("I do"), moves to collaborative monitoring with guided prompts within the learner's ZPD ("we do"), and ultimately fades prompts as the student independently plans and reflects on their own learning ("you do"). In the planning phase, the AI tool can set clear instructional goals and communicate them to the students; suggest effective study strategies; outline the plan the AI tool will follow; and follow a sequence of logical instructional steps. The AI tool can provide opportunities to practice metacognitive actions through scaffolded structures like thinking routines. The AI tool can check in on progress and help the learner monitor how well it's going. There are a number of reflection questions the tool can ask: "What did you learn today?" "What strategies did you use today?" "Which strategies were most effective?" Although, as mentioned above, it is questionable to what extent learners will choose to engage with these prompts. Even if they are likely to facilitate great outcomes, learners may be reluctant to apply them due to the cognitive effort required.

AI tools can communicate the importance of metacognition and teach some of the strategies; giving basic feedback (correctness) and hints when solving problems. At a deeper level, tools can model by communicating the plan, checking on progress, and asking the learner to reflect. At an even more advanced level, learners should drive the process of building the plan, monitoring their process, and reflecting on their experiences.

Our goal is to move beyond providing feedback on correctness of responses, for example, and instead encourage the learner to reflect on their own process. System instructions can direct the model to identify their own mistakes or reflect on the conversation. This may require a scaffolded approach if the learner is not yet comfortable with these techniques.

7.4. Assessing metacognition

When evaluating the capabilities of LearnLM, Gemini's capabilities that power multiple features and products, educators evaluated a variety of scenarios where learners interacted with the chatbot interface. When evaluating model behaviors, it is essential to distinguish cognitive scaffolding from genuine metacognitive support. In early LearnLM benchmarks, several rubric dimensions assessed foundational cognitive scaffolding that creates preconditions for reflection and self-knowledge. We also believe this is a frontier in learning science that AI tools have made more accessible to study.

- **Guide mistake discovery:** The tutor guides the student to discover their own mistakes.

- **Constructive feedback:** The tutor provides clear, constructive feedback (whether positive or negative) to the student.
- **Acknowledge correctness:** The tutor acknowledges when part or all of the student's response is correct.
- **Communicates plan:** The tutor communicates a clear plan or objective for the conversation [35].

Depending on the feature, rubrics may also assess other elements closer to metacognitive support:

- **Planning & goal setting:** The AI tool models the strategy of planning by communicating the aims and plan for the session
- **Learner reflection:** The AI tool prompts the learner to reflect on their progress, strengths and opportunities for growth
- **Metacognitive instruction:** The AI tool communicates the importance of metacognition and explains some of the strategies

A combination of all of these dimensions are used to assess the capability of the tools to build metacognitive skills in our learners.

7.5. Examples of building metacognition in Google's products

As this is a newer area we are exploring, we don't have many examples of supporting metacognitive knowledge in practice. The product ideas are categorized by levels of sophistication:

- **Level 1 (Basic):** The tool outlines key aspects of the learning process, such as formulating a plan, tracking progress, assessing the overall learning trajectory, and flagging learner errors, offering varying degrees of transparency regarding how it derived these elements.
- **Level 2 (Modeling):** The tool explicitly demonstrates its methodology by detailing its reasoning for the plan, checking in on progress, and pointing out that an error exists while guiding the student to discover it independently; it communicates the strategy for the learning session and continually monitors development.
- **Level 3 (Involving):** The tool directly engages the learner in co-creating their study plan, prompting them to self-monitor comprehension and fostering personal reflection on their learning journey.

8. Prepare learners for the future

While not strictly a "learning science" principle (it can be considered more of an aim or outcome), we believe that it's critical to prepare students for life as effective, caring, kind, compassionate humans who contribute to their society and reach their full potential.

Google aims to help students develop skills that are valuable outside of a purely academic context, often referred to as durable skills or 21st Century Skills. These are not necessarily new skills, but rather skills that have become increasingly vital in a modern, information-based economy. The skills are a combination of knowledge, values, and attitudes. Key skill groups include:

- Higher order thinking skills (critical thinking, analytical thinking, problem solving)
- Communication and collaboration
- Creativity

- Digital and AI literacy

All of the learning science principles above are methods for preparing learners for the future. While not an explicit principle itself, we do not want technology to replace these critical skills for learners.

8.1. Assessing how AI tools prepare learners for the future

While we have not formally added these items to our evaluation methodology, we have identified multiple dimensions at a high level that we might observe in our genAI tools.

- **Higher-order thinking skills:** The tool encourages application of problem solving, analytical and/or critical thinking skills, and provides adequate support for development of those skills.
- **Creativity:** The tool fosters creativity by modeling it, providing opportunities for learners to be creative, and helping learners develop and use their unique creative voice.
- **Communication & collaboration:** The tool provides opportunities for collaboration between learners and teaches them to communicate effectively.
- **Digital literacy:** The tool promotes safe, responsible and effective use of technology.
- **Preparation for the future:** The tool effectively prepares learners for their post-academic journey.

8.2. Examples of Google's products that prepare learners for the future

We can prepare learners for the future by building explicit products focused on durable skills, as we're starting to do with more experimental products like Vantage. In addition, we can start to build these techniques to support the critical user journeys and ways people use our tools. This could look like prompting a learner to evaluate the credibility of different sources before writing an essay, asking a learner to solve a problem that has multiple possible solutions and justify their chosen approach, or guiding a student through a design thinking process to foster creativity.

9. Future work

As we continue on our mission to develop high-quality, research-backed learning tools, we will attempt to address questions like these:

- What tools work for which students, under which conditions, and why?
- What additional principles should we consider when designing, building, and evaluating learning products?
- How might we incorporate domain-specific pedagogical practices into our tools?
- How do these design principles correlate with learning outcomes?
- How can we foster durable skills, including critical human-human collaboration and interpersonal connections?
- How might we move from supporting learners in traditional education systems of today to empower learners and educators to apply new techniques that enhance learning experiences?

10. Conclusion

The paper demonstrates that Google's approach to education feature product development is grounded in established learning science research. The path forward involves a continuous cycle of:

- **Model improvement:** Using strategies like pedagogical instruction following to make models better at explaining concepts and enhancing learning experiences.
- **Product design:** Intentionally designing features that embody these principles.
- **Rigorous evaluation:** Using robust evaluations to assess learner-perceived helpfulness, pedagogical effectiveness throughout the life cycle, including post-launch, to measure progress and identify areas for improvement.

We have moved beyond evaluations that measure the quality of the pedagogical behaviors toward external evaluations of real users interacting with the product that measure direct impact on learning content knowledge and skills; increasing affective factors and motivation; and building academic skills for the future. We are cautiously optimistic that we can help realize the potential of generative AI tools to create more effective, engaging, and equitable learning experiences for everyone.

11. A note on genAI use in this paper

- A combination of genAI tools were used to support the development of ideas and language in this paper (including Gemini Notebook as well as GeminiApp with a custom Learning Science Coach gem that refers to internal documents written by the pedagogy team, position papers, and literature reviews).
- The tools were used in various ways throughout the process; including refining text, identifying supporting research, and drafting / reviewing early versions.
- Additional reviews were conducted by specialized agents trained to review technical papers and trained on the author's own work.
- When text was suggested by AI tools, it was thoroughly reviewed and edited by the authors / contributors.

12. Contributors and acknowledgments

12.1. Core contributors

The following individuals contributed to the work described in this report. These lists are ordered alphabetically and do not reflect ranking of contributions.

The following individuals made core contributions to the Google pedagogy and learning science team, including conducting literature reviews, consulting with product and model-development teams, and determining how we could incorporate promising practices into our work:
Ana Petrec, Andrea Pavković, Antonia Ĉoić Zrinśćak, Antonija Matić, Brian Veprek, Gracen Grundfest, Kelsey Hammond, Laurel Ozersky, and Mia Mesar.

Google employees across organizations also contributed to this work:
Diana Akrong, Dixie Ching, Irina Jurenka, Kevin R. McKee, and Rachel Hashimshoni.

The following contractors also made key contributions to Google's pedagogy and learning science initiatives:
Ari Dolid, Ateka Ali, Lisa Batista, and Natalie Rodriguez.

12.2. Acknowledgments

Thank you to Google leadership on the Learning, Education, Research, and DeepMind teams for embracing learning science and its applications to product development:

Benedict Gomes, Chris Phillips, Irina Jurenka, Jason Mothersill, Jennifer Lee, Maureen Heymans, Shakir Mohamed, and Yossi Matias.

12.3. Reviewer

Thank you to Dr. Jeremy Roschelle, Co-Executive Director, Learning Sciences Research at Digital Promise, for his thoughtful review and comments of this paper.

References

[1] Ji Young Lee, Jone Sagastui, Christoph Mischo, Allyson S Masters, Svenja Peters, Katrin Wolstein, Roberta Michnick Golinkoff, and Kathy Hirsh-Pasek. Understanding cross-cultural differences in pedagogical beliefs: A comparison among south korean, spanish, and german early childhood teachers. *Teach. Teach. Educ.*, 148:104704, 2024.

[2] Ghulam Omar Qargha and Rachel Dyl. Invisible pedagogical mindsets: Developing a contextual understanding of pedagogies. working paper# 187.1. SPARKS working paper 1. *Center for Universal Education at The Brookings Institution*, 2024.

[3] Mary Burns, Rebecca Winthrop, Natasha Luther, Emma Venetis, and Rida Karim. A new direction for students in an AI world: Prosper, prepare, protect. *Center for Universal Education at The Brookings Institution*, 2026.

[4] Heather Schwartz and Melissa Kay Diliberti. Student use of AI for homework rises as concerns grow about critical thinking skills. https://www.rand.org/news/press/2026/03/student-use-of-ai-for-homework-rises-as-concerns-grow.html, March 2026. Accessed: 2026-8-22.

[5] Mallory Newall. NPR education and AI survey. Technical report, Washington, 2026.

[6] Hamsa Bastani, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman. Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26):e2422633122, 2025.

[7] Nataliya Kosmyna, Eugene Hauptmann, Ye Tong Yuan, Jessica Situ, Xian-Hao Liao, Ashly Vivian Beresnitzky, Iris Braunstein, and Pattie Maes. Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv preprint arXiv:2506.08872*, 4:206, 2025.

[8] Gregory Kestin, Kelly Miller, Anna Klales, Timothy Milbourne, and Gregorio Ponti. AI tutoring outperforms active learning. 2024.

[9] Coalition of leading education organizations introduces five quality indicators for edtech and AI products. https://iste.org/news/coalition-of-leading-education-organizations-introduces-five-quality-indicators- June 2024. Accessed: 2026-8-22.

[10] SAFE benchmarks framework. https://www.edsafeai.org/safe. Accessed: 2026-8-22.

[11] Wayne Holmes, Fengchun Miao, and Others. Guidance for generative AI in education and research, 2023.

[12] Kelly Mills, Sierra Noakes, Megan Pattenhouse, Alison Shell, Andrew Vollavanh, et al. Powerful learning with emerging technology. *Digital Promise*, 2025.

[13] Mitchell J Nathan and R Keith Sawyer. Foundations of the learning sciences. In R Keith Sawyer, editor, *The Cambridge handbook of the learning sciences, second edition*, pages 21–43. Cambridge University Press, Cambridge, England, 2014.

[14] Lev Semenovich Vygotsky and Michael Cole. *Mind in society: Development of higher psychological processes*. Harvard university press, 1978.

[15] AI Policy Perspectives. AI tutors should not approximate human tutors. https://www.aipolicyperspectives.com/p/ai-tutors-should-not-approximate, November 2025. Accessed: 2026-8-7.

[16] Michelene T H Chi and Ruth Wylie. The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educ. Psychol.*, 49(4):219–243, October 2014.

[17] David R Krathwohl. A revision of bloom's taxonomy: An overview. *Theory Pract.*, 41(4):212–218, November 2002.

[18] John Dewey. *The Child and the Curriculum*. The University of Chicago Press, 1902.

[19] Christian M Thurn, Peter A Edelsbrunner, Michal Berkowitz, Anne Deiglmayr, and Lennart Schalk. Questioning central assumptions of the icap framework. *npj Science of Learning*, 8(1):49, 2023.

[20] Daniel T Willingham. *Why don't students like school?: A cognitive scientist answers questions about how the mind works and what it means for the classroom*. John Wiley & Sons, New York, NY, 2021.

[21] Elizabeth L Bjork and Robert A Bjork. 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):56–64, 2011.

[22] Yana Weinstein, Megan Sumeracki, and Oliver Caviglioli. *Understanding How We Learn: A Visual Guide*. Routledge Member of the Taylor and Francis Group, New York, NY, 2019.

[23] James Hiebert and Douglas A Grouws. The effects of classroom mathematics teaching. *Second handbook of research on mathematics teaching and learning: A project of the national council of teachers of mathematics*, 1:371, 2007.

[24] John Dewey. *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*. D. C. Heath & Co., Boston, MA, 1933.

[25] Jean Piaget. *The Psychology of Intelligence*. Littlefield, Adams & Co., Paterson, NJ, 1960.

[26] Leon Festinger. *A Theory of Cognitive Dissonance*. Row, Peterson, Evanston, IL, 1957.

[27] W. A. Brownwell and V. M. Sims. The nature of understanding. In N. B. Henry, editor, *Forty-fifth yearbook of the National Society for the Study of Education: Part I. The measurement of understanding*, pages 27–43. University of Chicago Press, 1946.

[28] Hiroko Kawaguchi Warshauer. Productive struggle in middle school mathematics classrooms. *J. Math. Teach. Educ.*, 18(4):375–400, August 2015.

[29] David Wood, Jerome S Bruner, and Gail Ross. The role of tutoring in problem solving. *Journal of child psychology and psychiatry*, 17(2):89–100, 1976.

[30] Scott Freeman, Sarah L Eddy, Miles McDonough, Michelle K Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the national academy of sciences*, 111(23):8410–8415, 2014.

[31] Kenneth R Koedinger, Jihee Kim, Julianna Zhuxin Jia, Elizabeth A McLaughlin, and Norman L Bier. Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. In *Proceedings of the second (2015) ACM conference on learning@ scale*, pages 111–120, 2015.

[32] Izadora Volpato Rossi, Jordana Dinorá de Lima, Bruna Sabatke, Maria Alice Ferreira Nunes, Graciela Evans Ramirez, and Marcel Ivan Ramirez. Active learning tools improve the learning outcomes, scientific attitude, and critical thinking in higher education: Experiences in an online course during the COVID-19 pandemic. *Biochemistry and Molecular Biology Education*, 49(6):888–903, 2021.

[33] Louis Deslauriers, Logan S McCarty, Kelly Miller, Kristina Callaghan, and Greg Kestin. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences*, 116(39):19251–19257, 2019.

[34] LearnLM Team, Abhinit Modi, Aditya Srikanth Veerubhotla, Aliya Rysbek, Andrea Huber, Brett Wiltshire, Brian Veprek, Daniel Gillick, Daniel Kasenberg, Derek Ahmed, and Others. LearnLM: Improving gemini for learning. *arXiv preprint arXiv:2412.16429*, 2024.

[35] LearnLM Team, Abhinit Modi, Aditya Srikanth Veerubhotla, Aliya Rysbek, Andrea Huber, Ankit Anand, Avishkar Bhoopchand, Brett Wiltshire, Daniel Gillick, Daniel Kasenberg, and Others. Evaluating gemini in an arena for learning. *arXiv preprint arXiv:2505.24477*, 2025.

[36] Marsha C Lovett, Michael W Bridges, Michele DiPietro, Susan A Ambrose, and Marie K Norman. *How learning works: Eight research-based principles for smart teaching*. John Wiley & Sons, 2023.

[37] John Sweller. Cognitive load theory. In *Psychology of Learning and Motivation*, volume 55 of *The Psychology of Learning and Motivation*, pages 37–76. Elsevier, 2011.

[38] Richard E Mayer. Applying the science of learning: evidence-based principles for the design of multimedia instruction. *Am. Psychol.*, 63(8):760–769, November 2008.

[39] Allan Paivio. *Mind and its evolution: A dual coding theoretical approach*. Psychology Press, Philadelphia, PA, December 2006.

[40] John Sweller. Cognitive load theory and individual differences. *Learning and Individual Differences*, 110:102423, 2024.

[41] Center for Education Statistics and Evaluation. Cognitive load theory: Research that teachers really need to understand. *Sydney: Centre for Education Statistics and Evaluation*, 2017.

[42] Brian J Reiser and Iris Tabak. Scaffolding. In R Keith Sawyer, editor, *The Cambridge Handbook of the Learning Sciences*, pages 44–62. Cambridge University Press, Cambridge, England, 2014.

[43] Slava Kalyuga, Paul Ayres, Paul Chandler, and John Sweller. The expertise reversal effect. *Educational Psychologist*, 38(1):23–31, 2003.

[44] Janet Metcalfe and Nate Kornell. A region of proximal learning model of study time allocation. *Journal of memory and language*, 52(4):463–477, 2005.

[45] National Academies of Sciences, Engineering, and Medicine. *How people learn II: Learners, contexts, and cultures*. National Academies Press, Washington, D.C., September 2018.

[46] Richard M Ryan and Edward L Deci. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.*, 55(1):68–78, 2000.

[47] Nancy Frey and Douglas Fisher. Motivation requires a meaningful task. *Engl. J.*, 100(1):30–36, September 2010.

[48] A Wigfield and J S Eccles. Expectancy-value theory of achievement motivation. *Contemp. Educ. Psychol.*, 25(1):68–81, January 2000.

[49] John M Keller. Development and use of the ARCS model of instructional design. *J. Instr. Dev.*, 10(3):2–10, September 1987.

[50] Kushal Khemani. From points to pedagogy: A theoretical framework for gamified learning in secondary education. *EdArXiv*, May 2025.

[51] Jan L Plass and Shashank Pawar. Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education*, 52(3):275–300, 2020.

[52] Radek Pelánek. Adaptive learning is hard: challenges, nuances, and trade-offs in modeling. *International Journal of Artificial Intelligence in Education*, 35(1):304–329, 2025.

[53] Candace Walkington and Matthew L Bernacki. Appraising research on personalized learning: Definitions, theoretical alignment, advancements, and future directions, 2020.

[54] Kurt VanLehn. The behavior of tutoring systems. *International journal of artificial intelligence in education*, 16(3):227–265, 2006.

[55] John Hattie. *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge, London, England, 2008.

[56] Herva Emilda Sari, Benelekser Tumanggor, and David Efron. Improving educational outcomes through adaptive learning systems using AI. *International Transactions on Artificial Intelligence (ITALIC)*, 3(1):21–31, November 2024.

[57] Miling Chen. The impact of AI-assisted personalized learning on student academic achievement. *US-China Educ. Rev. A*, 15(06), June 2025.

[58] Sumei Hu. The effect of artificial intelligence-assisted personalized learning on student learning outcomes: A meta-analysis based on 31 empirical research papers. *Sci. Insights Educ. Front.*, 24(1):3873–3894, September 2024.

[59] Angélique Létourneau, Marion Deslandes Martineau, Patrick Charland, John Alexander Karran, Jared Boasen, and Pierre Majorique Léger. A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education. *NPJ Sci. Learn.*, 10(1):29, May 2025.

[60] Celeste Kidd, Steven T Piantadosi, and Richard N Aslin. The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS one*, 7(5):e36399, 2012.

[61] Mark A Guadagnoli and Timothy D Lee. Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of motor behavior*, 36(2):212–224, 2004.

[62] Robert C Wilson, Amitai Shenhav, Mark Straccia, and Jonathan D Cohen. The eighty five percent rule for optimal learning. *Nature communications*, 10(1):4646, 2019.

[63] Vincent Aleven, Elizabeth A McLaughlin, R Amos Glenn, and Kenneth R Koedinger. Instruction based on adaptive learning technologies. *Handbook of research on learning and instruction*, 2:522–560, 2016.

[64] Luis C Moll, Cathy Amanti, Deborah Neff, and Norma Gonzalez. Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory into practice*, 31(2):132–141, 1992.

[65] Margaret C Wang, Geneva D Haertel, and Herbert J Walberg. What influences learning? a content analysis of review literature. *The Journal of Educational Research*, 84(1):30–43, 1990.

[66] John H Flavell. Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *Am. Psychol.*, 34(10):906–911, October 1979.

[67] Philip H Winne and Roger Azevedo. Metacognition. In R Keith Sawyer, editor, *Cambridge Handbook of the Learning Sciences*, pages 63–87. Cambridge University Press, New York, NY, 2 edition, 2014.

[68] George Pólya. *How to Solve It: A New Aspect of Mathematical Method*. Princeton University Press, Princeton, NJ, 2nd edition, 1957.

[69] Barry J Zimmerman. Self-regulated learning and academic achievement: An overview. *Educ. Psychol.*, 25(1):3–17, 1990.

[70] Douglas Fisher and Nancy Frey. *Better learning through structured teaching: A framework for the gradual release of responsibility*. ASCD, 2021.

[71] P David Pearson and Margaret C Gallagher. The instruction of reading comprehension. *Contemporary educational psychology*, 8(3):317–344, 1983.