Thinking in Graphs with CoMAP: A Shared Visual Workspace for Designing Project-Based Learning

ANONYMOUS AUTHOR(S)

(a) Restricted Thinking: Fragmented ideas in a linear flow

(b) Empowered Thinking: Structured connections in a flexible space

Fig. 1. This figure conceptually contrasts two distinct paradigms for human-AI co-creation.(a) Represents the traditional paradigm, where the design process relies on linear tools and isolated conversational AI. This approach fails to capture the non-linear nature of creative design. (b) Shows the CoMAP paradigm, which utilizes a shared visual workspace as a persistent cognitive artifact. This allows for the externalization of non-linear thoughts, providing users with a distinct alternative for co-creation that addresses the limitations of the traditional paradigm.

Designing project-based learning (PBL) demands managing highly interdependent components, a task that both traditional linear tools and purely conversational AI struggle with. Traditional tools fail to capture the non-linear nature of creative design, while conversational systems lack the persistent, shared context necessary for reflective co-creation. Grounded in theories of distributed cognition, we introduce CoMAP, a system that embodies a graph-based co-design paradigm. By providing a shared visual workspace with dual-modality AI support, CoMAP transforms the human-AI relationship from a prompt-and-response loop into a transparent and equitable partnership. Our study with 30 educators shows CoMAP significantly improves teachers' design expression, divergent thinking, and iterative practice compared to a dialogue-only baseline. These findings demonstrate how a non-linear, artifact-centric approach can foster trust, reduce cognitive load, and empower educators to take control of their creative process. Our contributions are available at: https://comap2025.github.io/ .

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; *Laboratory experiments.*

Additional Key Words and Phrases: Human-AI co-creation, Instructional design,Large language models, Visual authoring, Educational technology

Manuscript submitted to ACM

ACM Reference Format:
Anonymous Author(s). 2026. Thinking in Graphs with CoMAP: A Shared Visual Workspace for Designing Project-Based Learning. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 30 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Project-based learning (PBL) is an instructional approach where students engage in extended, authentic projects that demand managing highly interdependent components like objectives, activities, and assessments. While PBL is highly effective [44, 68], its design process is complex. Most current instructional design tools still rely on linear text formats, which fail to represent the dynamic and multi-dimensional relationships inherent in PBL projects. As philosopher Marshall McLuhan noted, "the medium is the message" [49]—the linear medium itself imposes a cognitive burden on teachers, forcing them to mentally juggle fragmented information and hindering the iterative refinement of project plans [18, 45]. This creates a significant gap between how teachers naturally think and how current tools allow them to work.

To bridge this gap, we draw on theories from distributed cognition [35] and schema theory [38], which suggest that human reasoning and knowledge representation are fundamentally non-linear, structured around interconnected concepts [46, 58]. This aligns with how teachers think about PBL, naturally clustering components like roles, resources, and assessments. Following this insight, we propose a graph-based instructional design paradigm. In this paradigm, a graph serves as a shared visual workspace where nodes represent project elements and edges represent their relationships. This approach enables teachers to fluidly externalize their non-linear thoughts, explore alternatives, and manage complexity, thereby reducing the mental burden of managing multi-dimensional plans.

While recent advancements in AI offer promising avenues [77], existing tools still present significant challenges from a human-computer interaction (HCI) perspective. Learning design platforms model rigid workflows, and recent AI-driven generators often function as a black box, producing monolithic outputs with limited user control [10, 47]. This paradigm of AI as an opaque "task-executor" or "answer-engine" is ill-suited for the dynamic, collaborative nature of PBL design, where teachers require continuous adaptation and fine-grained control [1, 62]. To foster effective human-AI co-creation, tools must be designed to balance efficiency with user agency, transparency, and a persistent, shared context.

To move beyond the limitations of linear tools and opaque AI, we present **CoMAP**, a system that operationalizes a graph-based paradigm for human-AI co-creation. CoMAP is built around a single, central design principle: providing a **shared, visual workspace** that serves as a cognitive artifact for both the human user and the AI. This non-linear graph canvas not only externalizes the teacher's thoughts—allowing for incremental construction and intuitive reorganization—but also serves as the grounding representation for a multi-agent AI. By integrating a fine-grained GUI for direct manipulation with a high-level CUI for broad ideation, CoMAP enables **negotiable control** and facilitates a fundamental shift from AI as a black-box answer engine to a transparent, collaborative partner in the creative process.

This work makes three main contributions:

(1) We propose and operationalize a **graph-based paradigm for human-AI co-design in education**, where the graph serves as a shared visual workspace that externalizes teachers' non-linear cognitive processes and grounds the interactions of a multi-agent AI system. We argue this paradigm fundamentally shifts the human-AI relationship from opaque execution to transparent collaboration.

(2) We implement the **CoMAP system**, which embodies this paradigm through a novel integration of graph-based visualization, modular decomposition, and dual-modality AI support across GUI and CUI modes. The system's design is tailored to empower teachers, enabling negotiable control and supporting a fundamental shift from linear dictation to non-linear construction.
(3) We evaluate CoMAP in a mixed-methods user study (n=30), providing **empirical evidence** for a new co-creation model. Our findings demonstrate how a non-linear, artifact-centric approach influences design practices, significantly enhancing teachers' design expression, reducing their cognitive load, and strengthening their sense of trust and control within the human-AI co-creation process.

2 Related Work

This work is grounded in schema psychology, distributed cognition, and instructional design models, providing a theoretical framework for analyzing complex instructional processes. Visual graph structures are employed as a shared design space for human–AI collaboration, supporting teachers in clearly representing and organizing instructional ideas, thereby fostering design insights and ideation. Accordingly, this section reviews related work in two areas: (1) theoretical foundations in schema psychology, distributed cognition, and instructional design models; and (2) visual representations in instructional design and HCI.

2.1 Theoretical Foundations

2.1.1 Distributed Cognition. The theory of distributed cognition [35] offers a crucial perspective for understanding the complex interaction between human thought and the external world. A core tenet is that cognitive activity is not a centralized process confined to the individual's brain [37]. Instead, it is dynamically **distributed across multiple agents**, including people, artifacts, and the environment. Cognitive processes and their elements—such as information, memory, and and decision-making—flow and transform between these agents. Researchers emphasize that cognition depends on the continuous transformation of information between **internal and external representations** [36]. This interaction enables the efficient flow and conversion of information among the system's elements.

Within this framework, the importance of **artifacts** is highlighted [69]. They are not just passive tools. They are **active participants** that carry and shape cognitive processes. An effective artifact, like a visual chart or design tool, can externalize complex internal thoughts. This significantly reduces an individual's cognitive load [67]. Moreover, well-designed artifacts promote a **cognitive residue** effect [59], allowing individuals to maintain a clear line of thought even without direct interaction.

However, we must also critically examine certain forms of interaction. For example, many interaction models based on conversational AI and flowing documents are inherently centralized and linear [27]. In these models, the core of cognition may shift from the human to the AI system. This leads to what is known as **cognitive offloading** [57]. Users may become dependent on the AI's reasoning, causing their own thinking to stagnate [22]. Although these tools generate a large amount of text, if they don't organize it into an actionable external representation, humans may be left with words but no clear idea of **how to act** [63]. Therefore, a better auxiliary tool should move beyond this centralized model. It should provide clearer and more powerful cognitive support for deeper, more creative human thought.

2.1.2 Schema Psychology. To achieve these goals, we must return to the fundamental principles of human cognition. The concept of **schema** plays a central role in cognitive science [3, 73], with its theoretical roots tracing back to the philosopher Immanuel Kant [51, 58]. It describes how humans organize and interpret information using cognitive frameworks, emphasizing that knowledge is not stored in a fragmented manner but exists as an organized, interconnected framework [26]. Psychologists such as Frederic Bartlett further developed this theory, pointing out that schemas are representations of specific content as **knowledge units** [71] that function in the processing of new information. Therefore, it can be inferred that when knowledge is represented in a **modular** way, it can better match the brain's intrinsic cognitive structure, thereby helping to activate an individual's prior knowledge and supporting more efficient knowledge comprehension and application.

In describing the organization of knowledge, researchers have proposed various models based on schema theory, among which **Semantic Memory Models** [56] are the most influential. A representative model is the Semantic Network Model [6], proposed by Allan Collins and Ross Quillian [16, 17]. This model posits that concepts in semantic memory are organized as a **networked graph**. In this network, each concept is a node, and the relationships between concepts are the **edges** connecting them. This structure allows for the inheritance and transfer of information and enables the quick retrieval and association of information through the mutual **activation** of related nodes [66]. This model has profoundly influenced fields such as knowledge graphs [15] and information structure design, and provides a concrete theoretical blueprint for representing instructional content in a **graphical** format to better reflect and leverage the internal relationships among knowledge points.

These theories collectively reveal the essence of how the human cognitive system processes knowledge [42]: it is stored and used in a structured, networked manner. This cognitive mechanism provides a crucial theoretical foundation for the design of instructional design tools for teachers. Schema theory emphasizes that a teacher's instructional design process is essentially the externalization and concretization of their professional cognitive schema [38]. This research aims to support this process. By using a **structured design approach**, it is expected to better activate a teacher's prior knowledge and, through a **graphical** visualization, facilitate **divergent thinking**, thereby enhancing the quality and efficiency of instructional design.

2.1.3 Instructional Design Models. Schema psychology explains the fundamental principles of how the human brain organizes information, namely through modular cognitive frameworks. Building upon this, instructional design models provide concrete frameworks for how to organize and represent instructional content in a modular form. These models are systematic blueprints for creating effective learning experiences. They vary in granularity and focus, but all provide a clear structure for breaking down the complex task of teaching. For example, the classic ADDIE model (Analysis, Design, Development, Implementation, Evaluation) provides a high-level, generic sequence of steps [7]. The Dick and Carey model offers a more detailed, component-based approach, breaking the process into specific phases such as instructional analysis, learner analysis, and formative evaluation [12, 23]. Backward Design emphasizes first determining desired learning outcomes and assessments before planning instructional activities [74].

Among these, the ASSURE model stands out for its clear, systematic, and modular definition of design elements [64]. ASSURE is an instructional systems design model tailored for teachers to plan and deliver lessons effectively with technology and media. Its acronym represents six key steps: Analyze Learners, State Objectives, Select Methods, Utilize Media and Materials, Require Learner Participation, and Evaluate and Revise. The model's structured, sequential nature is highly suitable as a modular framework for our tool's design representation.

2.2 Visual Representations in Instructional Design and HCI

A long line of HCI and visualization research shows that external visual representations can reduce search, support perceptual inference, and offload memory in complex problem solving—benefits particularly relevant for representing interdependent learning objectives, activities, and assessments. Classic cognitive accounts explain why diagrams outperform text for coordination and reasoning by spatially grouping related information and enabling perceptual inferences that are hard to derive from linear prose [46, 70]. In collaborative settings, making intermediate reasoning translucent improves common ground and team performance, indicating that shared visual artifacts act as "common information spaces" and boundary objects that stabilize coordination [28, 65]. Within learning design specifically, systems such as LAMS [21], CompendiumLD [8], MOT+ [52], and COLLAGE/CLFP [34] have demonstrated that structured, shareable visual languages help instructors externalize pedagogical intent and reuse designs; yet their canvases are largely static, with limited support for dynamic iteration and provenance of design decisions.

Contemporary HCI work closes this gap by turning visualization from a representational surface into a co-creation mechanism: toolkits like IdeationWeb [61], PromptChainer [75] and ChainForge [2] let users compose and scrutinize multi-step AI workflows in node–link spaces, compare alternatives, and debug at multiple granularities—features that increase controllability and transparency during ideation. Complementary strands in sensemaking and explainable AI further argue for exposing model lineage and intermediate states to sustain trust and reduce cognitive burden during iterative design [33].

Taken together, these findings motivate a graph-based shared workspace for instructional design, where relationships among pedagogical elements are first-class, and where the workspace itself becomes a persistent, inspectable locus for human–AI co-design rather than a passive canvas.

3 Formative Study

To inform the design of CoMAP and ensure its alignment with teachers' real-world practices, we conducted a formative study with seven PBL practitioners from diverse disciplines and career stages. The study aimed to identify key **challenges** teachers face when PBL design, as well as to derive corresponding **design goals** to guide system development.

3.1 Participants and Procedure

Seven participants (6 female, 1 male) took part in the study. All had prior experience in designing PBL curricula and in using AI tools for teaching support. To ensure diversity, we recruited participants with different disciplinary and professional backgrounds: two had a background in educational technology, two were pre-service teachers, one was a novice teacher with one year of experience, one was a curriculum researcher, and one was a researcher in science education. Their subject areas included information technology, language, mathematics, and science. Participants ranged in age from 19 to 45 years (M = 26.4, SD = 10.6).

The study had two stages as shown in Figure 2. In the first stage, we conducted a semi-structured interview focusing on (1) participants' typical workflow when designing PBL curricula, (2) how they currently integrated AI tools into this process, and (3) the challenges and limitations they encountered when using existing AI support. Participants were also encouraged to share examples of AI-generated outputs they had used in practice. These insights were expected to surface key challenges in current practice that would inform the subsequent system design.

In the second stage, participants engaged in a think-aloud session using a low-fidelity static prototype as a design probe [72]. The prototype presented a preliminary graph-based representation of instructional design. Participants were invited to reflect from two complementary perspectives: as users, they provided feedback on clarity, potential usage scenarios, and points of confusion; as co-designers, they speculated on missing functionalities and future applications they envisioned for such a system. This step allowed us to elicit prospective design goals, translating user expectations into concrete directions for the development of CoMAP.

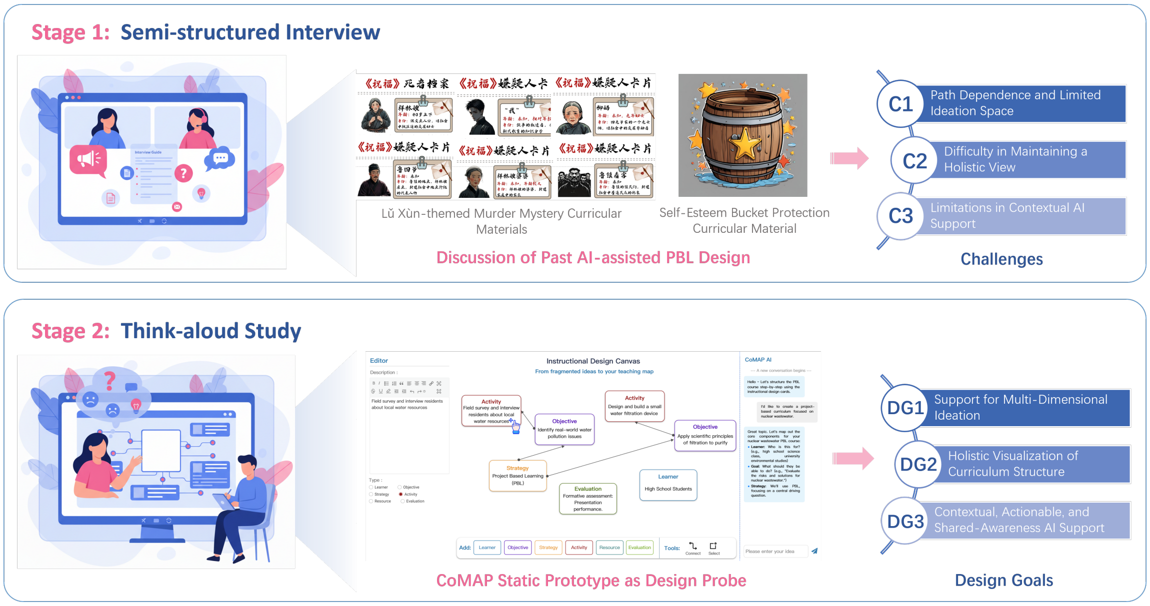


Fig. 2. The two-stage formative study design. In Stage 1, we conducted semi-structured interviews centered on participants' past experiences with AI-assisted PBL design to identify their core challenges. In Stage 2, a static CoMAP prototype was used as a design probe to elicit specific design goals that address these challenges.

Each session lasted approximately 45–75 minutes and was conducted one-on-one via online meetings. All interviews were audio-recorded and transcribed. Two researchers independently coded the transcripts using thematic analysis, and synthesized findings into both challenges for current PBL design and corresponding design goals that guided the design of the CoMAP system.

3.2 Challenges

Based on the formative study, we identified three major challenges that teachers face when designing PBL curricula, whether independently or with conversational AI support.

3.2.1 C1: Path Dependence and Limited Ideation Space. Teachers often fall into familiar patterns when designing PBL projects, or tend to follow AI-generated suggestions, which may unintentionally constrain their own creative thinking. One participant noted, *"When I use the AI, the activities it suggests feel so similar to old courses, there's not much novelty."* Another teacher mentioned, *"Even when I try to come up with new ideas, it's easy to get stuck in habitual structures. I just keep circling back to what I know."* A third participant highlighted the core frustration: *"I feel like I'm in a loop. I want to try something fresh, but I just can't seem to break out of the same old lesson plans."* These insights highlight a common challenge: without external prompts or diverse perspectives, ideation becomes narrow and constrained.

3.2.2 C2: Difficulty in Maintaining a Holistic View. Teachers frequently struggle to perceive the full structure of a PBL project, especially the complex interdependencies among learning objectives, instructional strategies, activities, and assessments. One participant explained, *"Objectives, strategies, and activities are scattered across the document. I have to constantly scroll back and forth and track all these links—it's exhausting."* Another teacher described the cognitive burden: *"I feel so overwhelmed. Without a clear global picture, it's hard to ensure everything is coherent. Sometimes I have all these great ideas, but they're just sections in a document without a clear, cohesive structure."* This challenge increases cognitive load and can hinder reflective planning.

3.2.3 C3: Limitations in Contextual AI Support. Teachers often face difficulties interacting with conversational AI to modify specific curriculum elements. One participant said, *"I have to describe in detail exactly what I want to change, and mistakes or repetitive copy-paste are common. It's a pain."* Another noted, *"Even after gathering AI suggestions, integrating them with my plan is cumbersome. The AI's suggestions don't feel like they truly understand my specific project."* A third teacher mentioned, *"I want the AI to understand that this specific activity connects to two different learning goals, not just one. But the bot just doesn't seem to get it. It's like we're not on the same page."* Teachers expressed the need for AI support that is flexible, context-aware, and able to operate at different granularities, while preserving teacher control.

3.3 Design Goals

Building on the challenges identified above, we derived three design goals to guide the development of CoMAP.

3.3.1 DG1: Support for Multi-Dimensional Ideation. To address the challenge of path dependence and limited ideation space (C1), CoMAP aims to enable teachers to explore diverse ideas across multiple curriculum dimensions, including learner analysis, learning objectives, instructional strategies, activities, resources, and assessment. The system should provide structured inspiration that is suggestive rather than prescriptive, allowing teachers to generate creative solutions while maintaining pedagogical autonomy. This approach is particularly beneficial during early-stage brainstorming, supporting broader exploration and reducing fixation on habitual design patterns.

3.3.2 DG2: Holistic Visualization of Curriculum Structure. To overcome the difficulties in maintaining a global perspective (C2), CoMAP must present all curriculum elements and their interrelationships in an integrated view. The system should allow teachers to seamlessly navigate between a high-level overview and detailed local views, reducing cognitive load and supporting reflective planning. This visualization should externalize complex many-to-many relationships among objectives, strategies, activities, and assessments, enabling teachers to detect gaps or conflicts, compare alternatives, and iteratively refine their plans. This integrated perspective facilitates cross-disciplinary collaboration and helps teachers maintain coherence throughout the design process.

3.3.3 DG3: Contextual, Actionable, and Shared-Awareness AI Support. To address the limitations in current AI interactions (C3), CoMAP should provide AI support that is contextual, flexible, and actionable. The system needs to support a shared awareness of the current design state, so that any modifications made by teachers on the canvas are visible to every AI agent. AI suggestions should be structured and actionable, allowing teachers to adopt recommendations that are immediately applied to the relevant elements on the canvas. The system should support both local agents for fine-grained, component-level edits and a global agent for high-level guidance, ensuring context-aware assistance that preserves teacher control and enables efficient, interactive co-creation throughout the design process.

4 The CoMAP System

Building upon the challenges and design goals derived from our formative study, we designed and implemented CoMAP to address the inherent limitations of both traditional linear tools and purely conversational AI. We propose a distributed cognitive framework that leverages a shared visual representation to resolve the tension between unstructured creative ideation and the need for a structured pedagogical representation. This section first articulates the system's core design

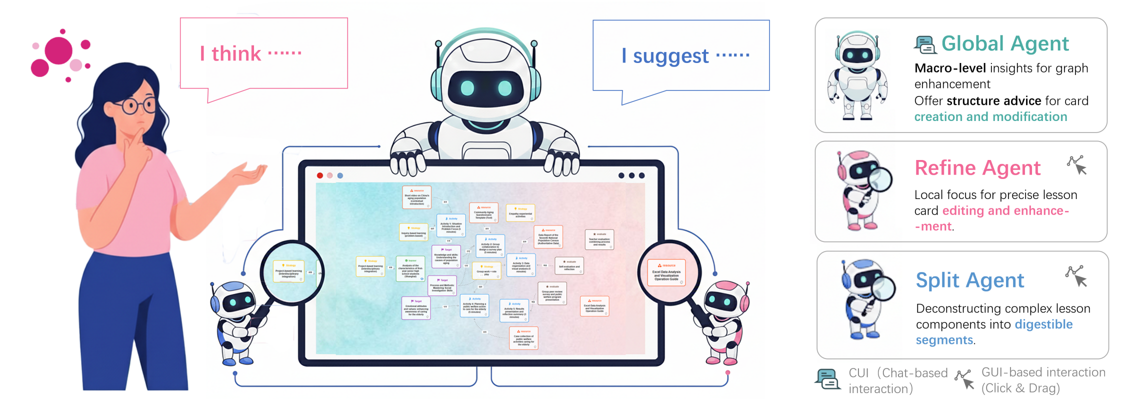


Fig. 3. Overview of the CoMAP system's interactive components and the flow of distributed cognition. This architecture embodies a distributed cognitive framework, where the cognitive load is shared among three key components: the human designer, the Structured Graph Canvas serving as an external cognitive workspace, and a team of specialized AI agents. The human interacts with the canvas to externalize their ideas, while the agents operate on this shared workspace, providing targeted assistance that augments the human's creative process.

philosophy—a distributed cognitive framework. We then detail the key features of the interface and the dual-modality AI support, culminating in a walkthrough scenario that illustrates how CoMAP operationalizes our proposed paradigm.

4.1 The CoMAP Paradigm: A Distributed Cognitive Framework

The design of CoMAP is founded on the principle of empowering educators by augmenting, not automating, their creative process. This is realized through a distributed cognitive framework, where cognition is not confined to a single individual but is instead shared and distributed among the human designer, a **shared graph canvas**, and a team of **collaborative AI agents**. This approach addresses the inherent limitations of both traditional linear tools and purely conversational AI, allowing each component to contribute to the collective intelligence of the system.

The **Structured Graph Canvas** serves as the central hub of this framework, directly addressing the limitations of internal mental models by providing a persistent external representation of the curriculum design. As a cognitive artifact [35], this visual language is grounded in pedagogical semantics, allowing both the human designer and the AI agents to offload complex information from their limited working memory onto a shared workspace. This cognitive offloading directly reduces the mental burden of tracking interdependencies (C2), thereby freeing up cognitive resources for higher-level creative tasks. In essence, the canvas acts as a shared visual language that facilitates the formation of a shared mental model [24] between the human and AI, fulfilling our design goals of enabling a fluid transition from unstructured ideation to a structured, coherent representation (DG1) and supporting transparent, context-aware collaboration (DG2).

The **Collaborative Agents** complement the canvas by tailoring AI support to different levels of the design process. First, a global conversational agent facilitates divergent ideation, offering high-level, contextual suggestions that help educators explore alternative structures and overcome habitual patterns (C1). Second, local GUI-integrated agents support convergent refinement, providing targeted operations such as content enrichment or decomposition that can be directly applied to individual nodes (C3). Together, these agents ensure that AI support is both broad and fine-grained, enabling a workflow that combines creativity with precision and preserving transparency and user control in line with DG1 and DG3. The overall architecture illustrating these components and their interactions is depicted in Figure 3.

4.2 The Structured Canvas: Externalizing Pedagogical Structures

Traditional curriculum design often relies on an internal mental model, which can be limited in its capacity to hold and manipulate the rich interdependencies of complex projects. This reliance on a transient internal representation often leads to cognitive overload and a breakdown in design coherence. To address this core challenge, we engineered the CoMAP canvas as a cognitive tool designed to offload the mental burden of managing complex designs. It achieves this by providing a persistent external representation that is both pedagogically structured and fluidly interactive.

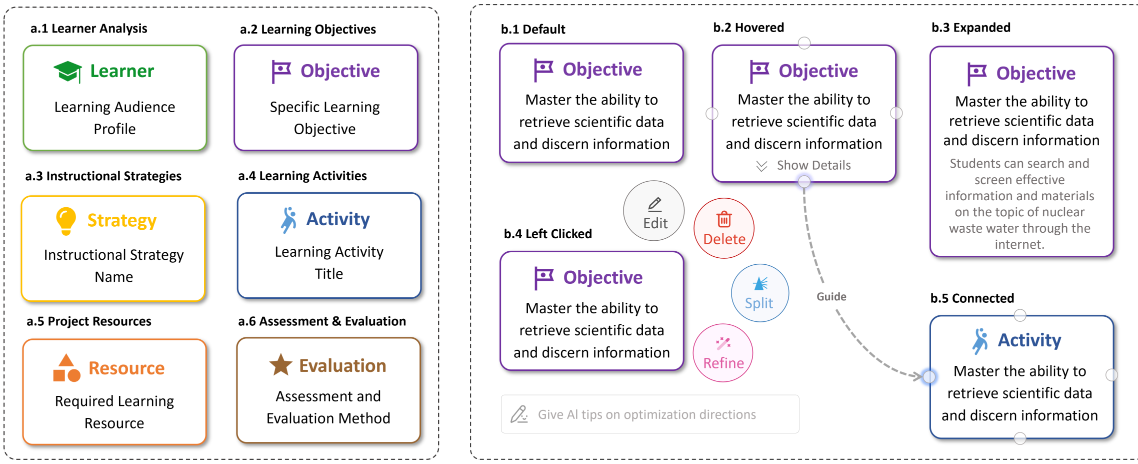
4.2.1 Pedagogical Semantics: A Visual Grammar for Design. To provide a shared, structured language for design, the canvas elements are imbued with pedagogical meaning. Inspired by the ASSURE model, we decomposed the PBL process into six core components.

Nodes (Building Blocks of Cognition): Each node represents a fundamental design unit. As shown in Figure 4 (a), each node type is visually distinguished by a unique icon and color to ensure at-a-glance readability and reduce cognitive load. The six node types are:

- **Learner Analysis (a.1) :** This node anchors the entire design process in the context of the students, prompting educators to consider their prior knowledge, developmental stages, and motivations to ensure the project is both accessible and engaging.
- **Learning Objectives (a.2) :** This node defines the specific, measurable skills, knowledge, or attitudes students are expected to acquire. These objectives serve as the instructional targets, guiding the selection of all subsequent strategies and activities.
- **Instructional Strategies (a.3) :** This node outlines the high-level pedagogical approaches (e.g., inquiry-based learning) that orchestrate the learning experience, representing the macro-level 'how' of the project.
- **Learning Activities (a.4) :** This node details the concrete, hands-on tasks and interactions where knowledge is actively constructed. They are the tangible embodiment of the instructional strategies.
- **Project Resources (a.5) :** This node identifies all necessary materials, media, and tools—from physical equipment to digital applications—required to support the learning activities.
- **Assessment & Evaluation (a.6) :** This node specifies methods for both formative (ongoing) and summative (final) evaluation, which is critical for measuring student progress against the objectives and for revising the design itself.

To further support the creative process, this visual grammar is complemented by a "Design Hints" panel. This panel offers a rich library of curated examples and templates as shown in Table 1, acting as catalysts for creativity and helping to mitigate cognitive fixation (C1) by supporting serendipitous discovery. This dual-support structure—combining curated presets with generative AI—is intentional: the Design Hints offer reliable, pedagogically-vetted starting points that lower the barrier to entry, while the AI agents provide unbounded, context-specific creativity.

Edges (Externalizing Complex Relationships): Beyond individual components, the true complexity of instructional design lies in the rich interdependencies between them. To externalize these relationships, the system supports the creation of directed, labeled edges. This is critical, as it directly offloads the cognitive work of tracking interdependencies from the designer's working memory onto the visual interface [46]. To provide semantic scaffolding, CoMAP suggests conventional relationship labels based on instructional design literature (e.g., a **Resource** "supports" an **Activity**; an



(a) Basic Node Type **(b) Different Display Mode**

Fig. 4. The visual grammar and interactive states of CoMAP nodes. (a) illustrates the six basic node types: **Learner Analysis** (a.1), **Learning Objectives** (a.2), **Instructional Strategies** (a.3), **Learning Activities** (a.4), **Project Resources** (a.5), and **Assessment & Evaluation** (a.6). Each type is distinguished by a unique icon and color to facilitate at-a-glance comprehension of the design structure. (b) demonstrates the different display modes and interactive states of an Objective node: **Default** (b.1), **Hovered** (b.2) to reveal options, **Expanded** (b.3) for detailed content editing, **Left Clicked** (b.4) exposing AI-powered refinement and splitting tools, and **Connected** (b.5) to another node (e.g., Activity) to illustrate pedagogical relationships. These states support fluid, non-linear design exploration.

Objective "guides" an **Activity**; an **Evaluation** "measures" an **Activity**). For any connections without a predefined type, a general "relates to" label is used. Crucially, to preserve designer autonomy, all edge labels are fully editable, allowing educators to articulate the unique logic of their designs.

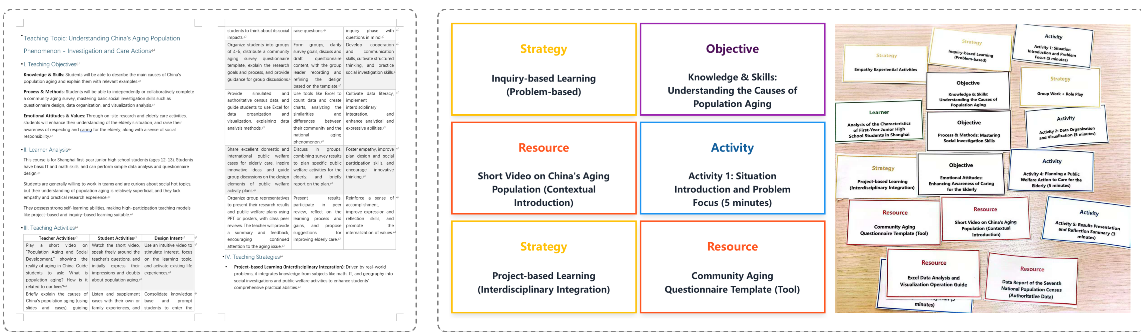
4.2.2 Fluid Interaction: Supporting Non-linear Exploration. CoMAP's interactions are designed to support a non-linear, exploratory workflow. Nodes can be directly manipulated, and each supports a dual-view mode (**summary** and **detail**) for seamless transitions between macro- and micro-level perspectives. As shown in Figure 4b, these interactive states include a default view (b.1), a hovered state (b.2) revealing options, an expanded view (b.3) for detailed content editing, a left-clicked state (b.4) exposing AI-powered refinement and splitting tools, and a connected state (b.5) illustrating pedagogical relationships between nodes. This supports a form of "semantic zooming," allowing teachers to fluidly shift between a holistic overview for structural reasoning and a focused view for detailed work. The canvas itself offers infinite panning and zooming, supported by a **minimap** for constant global context. Crucially, the design can be exported into two formats: printable cards for physical collaboration or a linear lesson plan document, bridging the gap between the digital design space and traditional administrative or collaborative workflows. As shown in Figure 5, these exports provide flexibility, with (a) an editable lesson plan for detailed documentation and (b) printed cards that facilitate hands-on brainstorming and in-person group work.

4.3 Collaborative Agents: A Dual-Modality Approach to AI Co-Creation

Instructional design is not a single, monolithic task. Instead, it involves a dynamic process that shifts between divergent ideation in the early, ambiguous stages and convergent refinement as the plan takes shape. This shift presents educators

Table 1. A selection of Preset Design Hints provided in CoMAP to scaffold the design process across different node types and categories.

Node Type	Category / Level	Example Title & Description
Learner	Primary School	**Primary School Learners**: High curiosity, short attention span, learn through play and concrete operations...
	Junior High	**Junior High Learners**: Developing abstract thinking, sensitive to peer relationships and self-identity...
	High School	**High School Learners**: Advanced abstract thinking, strong logical analysis and independent thought...
	Adult	**Adult Learners**: Goal-oriented, rich prior experience, value practicality and relevance...
Strategy	Constructivism	**Project-Based Learning (PBL)**: Centered on an authentic problem, guiding students in long-term inquiry...
	Cognitivism	**Scaffolding**: Providing structured support to help students complete tasks beyond their current ability...
	Behaviorism	**Direct Instruction**: Teacher-led, highly structured model for teaching foundational knowledge and skills...
	Collaborative	**Collaborative Learning**: Students work in small groups to achieve shared learning goals, emphasizing interdependence...
	Differentiated	**Differentiated Instruction**: Adjusting content, process, and evaluation to meet individual student needs...
Activity	Interactive	**Brainstorming**: Students freely generate ideas on a topic without judgment, which are then categorized and evaluated...
	Collaborative	**Jigsaw**: A large topic is decomposed; each student becomes an "expert" on one part to teach their peers...
	Inquiry-based	**Case Study Analysis**: Students analyze a real-world case to apply knowledge and develop problem-solving skills...
	Practice-based	**Concept Mapping**: Students visually represent their understanding of a topic by connecting concepts and ideas...
	Reflective	**Reflective Journaling**: Students regularly write about what they've learned, challenges faced, and areas for improvement...
Objective	Template	**Knowledge & Skill Goal**: [Actor] Student [Condition] After [a learning activity] [Verb] is able to [perform an action]...
	Template	**Problem-Solving Goal**: [Actor] Student [Condition] In a [collaborative context] [Verb] is able to [design/analyze]...
	Template	**Value & Affective Goal**: [Actor] Student [Condition] After [a reflective experience] [Verb] is able to [express/critique]...
Evaluation	Formative	**One-Minute Paper**: At a session's end, students write down one key learning and one remaining question...
	Formative	**Peer Assessment**: Students provide feedback on each other's work based on established criteria...
	Summative	**Project Report/Presentation**: Evaluates the comprehensive application of knowledge on a final product...
	Summative	**Portfolio Assessment**: A collection of student work over time to evaluate progress and achievement...
Resource	Interactive Tool	**PhET Interactive Simulations**: Free, research-backed science and math simulations for active learning...
	Video Platform	**Khan Academy**: A non-profit educational organization providing free video tutorials and interactive exercises...
	Document Platform	**Smart Education Platform (CN)**: Official platform with high-quality, free lesson plans and courseware...
	Course Platform	**Coursera / edX**: Platforms offering university-level courses for teacher development or student extension...



(a) Editable Lesson Plan **(b) Using Printed Card to Brainstorm**

Fig. 5. Two export modes supported by CoMAP: (a) an editable linear lesson plan document for administrative or documentation purposes, and (b) printable cards for physical collaboration and hands-on brainstorming. These export options bridge the gap between digital design and traditional teaching workflows.

with two distinct challenges: the need for high-level, creative support to overcome a blank canvas (C1) and the demand for precise, contextual assistance to meticulously refine the details (C3). To address these challenges, we designed a dual-modality AI system comprising a global agent and a suite of local agents, each tailored to a specific phase of the design workflow.

4.3.1 The Global Agent: A Conversational Partner for Divergent Ideation. To support the early, divergent stages of brainstorming, the global agent provides high-level support through a conversational user interface (CUI). The agent

(a) observes the entire canvas and acts as an "Ideation Partner," translating the teacher's natural language queries into structured recommendations. As shown in Figure 6, these recommendations can either be for new creations, such as generating cards to fill a gap in the curriculum (b), or for modifications to an existing structure (c). Upon acceptance, the suggestion is instantly rendered onto the canvas as fully formed nodes and edges.

This interaction fundamentally addresses the "gulf of execution" [50] by bridging the gap between a high-level goal and the low-level actions required to represent it structurally. We term this a "Cursor for instructional design," as it allows teachers to use high-level natural language to direct the precise creation and arrangement of structured content, much like a physical cursor translates movement into a specific location on a screen.

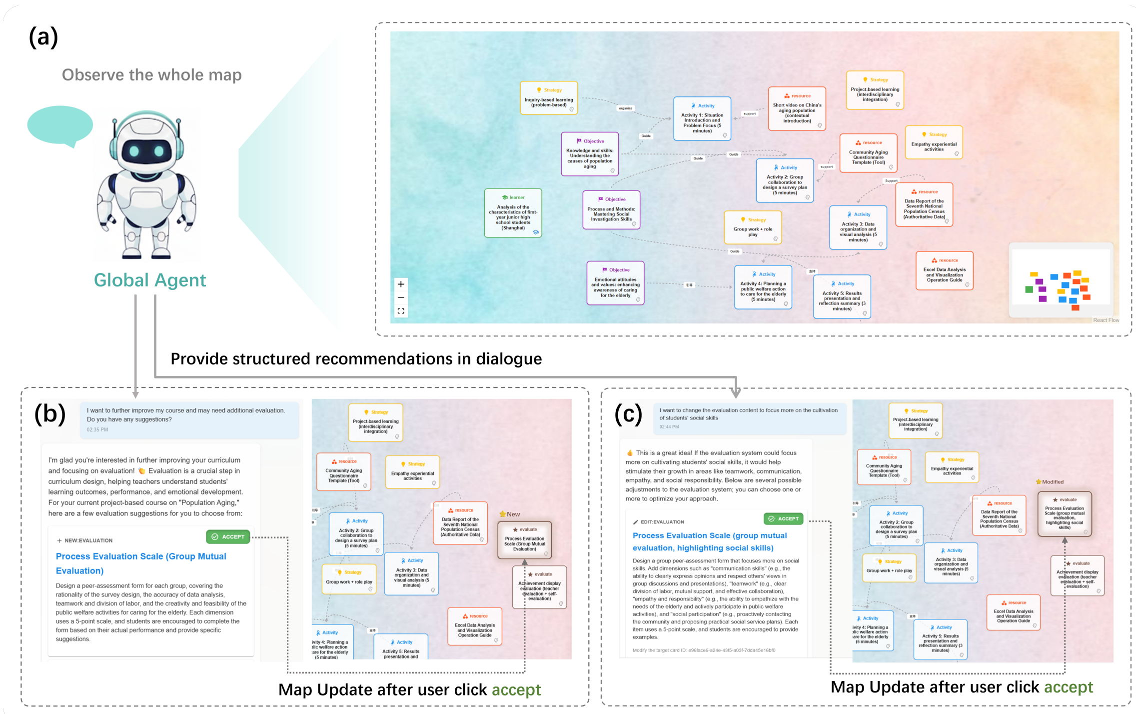


Fig. 6. The CoMAP Global Agent's conversational workflow. **(a)** The Global Agent observes the entire design canvas and provides support through dialogue. The agent's recommendations can be for either creation or modification. **(b)** In response to a teacher's natural language query, the agent generates a new, structured card recommendation. Upon acceptance, this card is added to the canvas. **(c)** Alternatively, the agent provides a structured recommendation to modify an existing card. Upon acceptance, the targeted card on the canvas is updated to reflect the new information. This demonstrates the direct translation from a high-level dialogue intent to both generative and adaptive structured canvas updates.

4.3.2 Local Agents: Contextual Assistants for Convergent Refinement. Once a basic structure is in place, the design work shifts to a phase of convergent refinement. To meet the need for fine-grained, contextual assistance (C3), we provide local agents embedded directly within the canvas's graphical user interface (GUI). These agents are invoked via a context menu on any node and include a Refine Agent for improving content quality (e.g., enriching descriptions, clarifying objectives) and a Splitting Agent for decomposing complex ideas into smaller, more manageable sub-nodes.

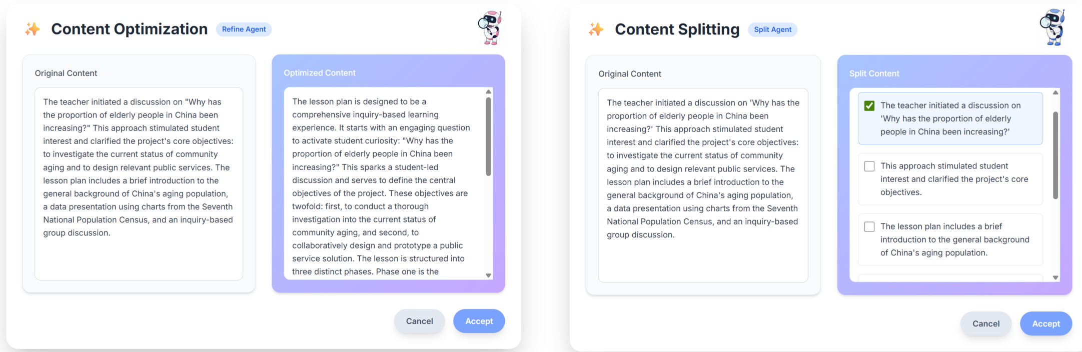


Fig. 7. The "before-and-after" comparison UI for CoMAP's Local Agents. **(Left)** The Refine Agent for 'Content Optimization' transforms an 'Original Content' into an 'Optimized Content' version, enhancing its quality and detail. **(Right)** The Splitting Agent for 'Content Splitting' decomposes a complex 'Original Content' into multiple, selectable 'Split Content' components. This design explicitly preserves user agency by making AI suggestions transparent and providing teachers with fine-grained control to accept or reject changes.

A crucial HCI principle for co-creation is preserving user agency. To ensure the user remains in control and the AI's suggestions are transparent, the local agents' output is presented in a dedicated "before-and-after" comparison UI, as illustrated in Figure 7. This differential view makes the AI's reasoning transparent and gives the teacher explicit control to accept or reject changes. For example, the Refine Agent transforms an 'Original Content' into a more 'Optimized Content' version, while the Splitting Agent breaks down a single 'Original Content' into multiple, selectable 'Split Content' components. By supporting efficient dismissal of unwanted suggestions, this design ensures the AI serves as a co-pilot, not an autopilot, which is a direct implementation of DG3.

4.3.3 Synergistic Workflow: From Divergence to Convergence. Crucially, these two agent modalities are not siloed; they are designed to support a seamless workflow from divergent ideation to convergent refinement. A teacher can use the Global Agent to generate a high-level project scaffold and then immediately select a resulting node to invoke a Local Agent for fine-grained improvements. This interaction is underpinned by a principle of shared awareness: all changes on the canvas, whether made by the user directly or through a Local Agent, update the global state of the design. This ensures that when the teacher re-engages the Global Agent, its subsequent suggestions are informed by the latest, most detailed version of the plan, creating a continuous and contextually-aware feedback loop between the different forms of AI support.

4.4 Walkthrough: Designing a PBL Unit with CoMAP

To illustrate the system in practice, we present a scenario featuring Sarah, a middle school science teacher grappling with the creation of a new PBL unit.

4.4.1 Divergent Ideation: From a Vague Idea to Structured Options. Sarah begins with a common challenge: a vague idea for a PBL unit on "Local Water Pollution," but no clear starting point. Faced with the daunting blank canvas, she turns to the Global Agent. She types a simple query, "Help me brainstorm a project about water pollution for 8th graders." The agent responds not with a single plan, but with several distinct project angles, each represented by a different set of initial nodes. This directly addresses her cognitive fixation (C1) by presenting diverse possibilities. Sarah reviews the options, intrigued by one focused on "community investigation," and accepts it. The AI instantly populates her canvas with an initial, structured map of core nodes, transforming her abstract idea into a tangible and editable starting point.

4.4.2 Convergent Selection and Personalization. With this initial scaffold, Sarah's own creativity is activated. She enters a phase of direct manipulation, reorganizing the nodes and editing content to reflect her teaching style. She adds a new Resource node for a water testing kit she knows is available from the school district, and rephrases an AI-generated Learning Objective to align with specific state curriculum standards. The process becomes a fluid conversation between her domain expertise and the system's structural suggestions, a core aspect of our human-AI co-creation paradigm.

4.4.3 Fine-Grained Refinement and Optimization. Sarah then focuses on a specific Activity node labeled "Field Trip," realizing the concept is too broad for a single class period. To add detail and precision, she right-clicks and invokes the Split Agent. The agent's suggestion UI appears, breaking the activity into three distinct, concrete sub-nodes: "Prepare for Trip," "Collect Samples," and "Analyze Data." Sarah selects all three, and the canvas instantly updates. Next, she turns her attention to the new "Analyze Data" node. Feeling the description could be more robust, she uses the Refine Agent. A before-and-after comparison UI appears, showing a significantly more detailed and pedagogically sound suggestion, which she accepts with a single click.

4.4.4 Holistic Review and Global Alignment. With the details finalized, Sarah zooms out, using the minimap to view the entire project. The holistic visualization allows her to quickly spot a subtle misalignment between her final Assessment and a key Learning Objective on the other side of the canvas. She quickly makes a connection and adjusts the assessment, ensuring the entire project is coherent and aligned. In a short time, Sarah has moved from a nebulous concept to a coherent, detailed, and personalized project plan, feeling a sense of ownership and confidence throughout the process. This scenario demonstrates how CoMAP's fluid interplay between its components empowers teachers to navigate the complex, non-linear process of PBL design.

4.5 Implementation Details

CoMAP is implemented as a full-stack web application. The frontend is built with React, and the interactive canvas is powered by the React Flow library. The backend server is developed using Python with the Flask framework, and user interactions are logged in a SQLite database to support a detailed analysis of the design process. The AI agent functionalities are driven by the OpenAI GPT-4.1 API, integrated through a carefully designed interaction architecture:

- **Agent-driven Canvas Manipulation:** To enable the Global Agent to directly and efficiently modify the canvas, we implemented a function-based interaction model. The agent is provided with a set of predefined functions in its system prompt, such as add_node(type, title, description), modify_node(id, new_title, new_description), and create_edge(source_id, target_id, label). In response to a user's prompt, the agent generates a sequence of these function calls with the appropriate parameters. This sequence, formatted as a JSON array of operations, is then executed by our backend to render the changes on the canvas. This approach supports responsive interaction by breaking down complex requests into a series of discrete, lightweight operations.
- **Contextual Grounding:** To generate relevant function calls, the agent relies on rich contextual information. For the **Global Agent**, the prompt includes the conversation history and a serialized representation of the entire graph, enabling it to make holistically-informed decisions. For **Local Agents**, which perform more targeted tasks like

refinement, the context is focused on the selected node and its immediate neighbors to ensure highly relevant suggestions.

5 User Study

To evaluate CoMAP, we conducted a mixed-methods user study comparing our graph-based, multi-agent system against a conventional dialogue-based AI assistant for PBL design. The study was designed to answer the following research questions:

- **RQ1:** How does the representational format of the design tool (graph vs. text) influence designers' perceived ability to understand and express complex instructional plans?
- **RQ2:** How does CoMAP's graph-based co-design paradigm impact the human-AI interaction experience (e.g., controllability, transparency, cognitive load, collaboration, trust) compared to a standard dialogue-based AI?

5.1 Participants

We recruited 30 participants (23 female, 7 male; age M=24.2, SD=3.63) from 21 different cities across the country via professional educator forums. Participants were recruited individually and were unacquainted with their randomly assigned partners prior to the study. The sample comprised a mix of educational professionals, including in-service teachers, pre-service teachers, and curriculum designers, covering a range of subjects. All participants reported having prior experience in PBL design and the use of AI tools.

5.2 Study Design

We employed a within-subjects crossover design to evaluate the impact of the system interface on user experience. This design allowed each participant to serve as their own control, experiencing both the CoMAP (our experimental condition) and a Baseline condition. The CoMAP condition utilized our designed system, which integrates a graphical canvas for organizing ideas with a collaborative AI. The Baseline condition served as a control, where participants used a standard, dialogue-based AI and were provided with a separate document to serve as a persistent record for their design process. To mitigate potential learning or order effects, participants were counterbalanced by being assigned to one of two groups: Group A completed the Baseline condition first, followed by CoMAP, while Group B completed the conditions in the reverse order. The study procedure was as follows:

(1) **Onboarding and Background Survey:** Participants were briefed on the study's purpose, signed a consent form, and completed a background survey, which included the Intelligent TPACK scale to assess their baseline proficiency.
(2) **Platform Training:** A facilitator provided a brief tutorial and a live demonstration of both the Baseline and CoMAP systems to ensure all participants were familiar with their functionalities.
(3) **Experimental Block 1:**
    - **Task 1:** Participants individually completed the first design task, using the tool assigned to their group for this block. We provided two structured Project-Based Learning (PBL) prompts, chosen for their interdisciplinary nature and universal relatability. The two distinct tasks were counterbalanced with the conditions to control for task-specific effects.
    - **Sharing:** Each participant presented their design idea to their partner in a brief "mock-teaching" format.

- **Interview and Questionnaires 1:** Participants engaged in a semi-structured interview about their experience and then individually completed the post-task questionnaires for the tool they had just used.

(4) **Experimental Block 2:**
- **Task 2:** Participants switched tools, using the second, distinct design task for this block. The group that used the Baseline tool in Block 1 now used CoMAP, and vice versa. This ensured that each participant used both tools.
- **Sharing:** Participants again shared their new designs with each other.
- **Comparative Interview and Questionnaires 2:** Participants completed a final interview, which included comparative questions about the two tools, followed by a final set of post-task questionnaires for the second tool.

To ensure a comparable basis for evaluation across all participants and conditions, we provided two structured Project-Based Learning (PBL) design prompts. The tasks were carefully chosen to be universally relatable and inherently interdisciplinary, accessible to participants regardless of their specific subject expertise. The detailed requirements for each task are as follows:

- **Task A: History of Foreign Food Crops:** Participants were required to design a PBL unit about the history and cultural significance of food crops introduced to their country. The task involved interdisciplinary elements from history, geography, and biology.
- **Task B: Population Aging and Community Care:** Participants were tasked with designing a PBL unit focused on investigating local population aging. This task required integrating concepts from mathematics, politics, and geography, and included planning a community survey and a service event.

5.3 Measurements

We collected quantitative data on the following measures:

- **Background Knowledge:** Prior to the main tasks, we administered an Intelligent TPACK (Technological Pedagogical Content Knowledge) scale as a background survey [13]. The internal consistency for this scale was strong (Cronbach's alpha = 0.87). This allowed us to understand our participants' proficiency in integrating technology, pedagogy, and content knowledge.
- **Human-AI Interaction Experience:** After each task, this was assessed using a quantitative scale covering five core constructs: Controllability, Transparency, Cognitive Load, Collaboration, and Trust. The items were adapted from established scales in human-computer interaction and human-AI teaming literature, specifically a combination of the Trust in Automation Scale [43] and the NASA-TLX scale for perceived mental effort [30] to measure cognitive load. We confirmed the internal consistency of these constructs using Cronbach's alpha coefficients (CoMAP: Controllability = 0.72, Transparency = 0.68, Cognitive Load = 0.78, Collaboration = 0.63, Trust = 0.80; Baseline: Controllability = 0.87, Transparency = 0.84, Cognitive Load = 0.86, Collaboration = 0.91, Trust = 0.78).
- **Perceived Understanding of Design:** Participants' self-reported ability to understand their partner's design during the sharing session. This was measured with custom-designed items that align with the theoretical frameworks of external cognition and distributed cognition, which examine how external representations and artifacts facilitate problem-solving and communication [76]. The Cronbach's alpha coefficients were 0.83 for CoMAP and 0.91 for Baseline.

- **Perceived Expression of Design:** Participants' self-reported ability to clearly present their own design during the sharing session. This was also measured with custom-designed items grounded in the principles of shared understanding in collaborative design, which posits that a shared visual representation is crucial for effective team communication and idea articulation [40]. The Cronbach's alpha coefficients were 0.86 for Graph and 0.87 for Baseline.

5.4 Data Collection and Analysis

The data collection phase yielded a comprehensive, multi-modal dataset to form a robust understanding of the user experience. The collected data included high-fidelity **interaction logs** from platforms (capturing user actions and AI dialogue history), **questionnaire data** from our dependent variable assessments, **audio recordings** of the peer-sharing sessions, and **interview transcripts** from the semi-structured interviews.

5.4.1 Quantitative Analysis. For our primary quantitative analysis, all questionnaire data and metrics derived from the interaction logs were subjected to a normality test (Shapiro-Wilk) to inform the choice of statistical procedures. Paired-samples t-tests were applied to data that were normally distributed, while Wilcoxon signed-rank tests were used for non-parametric data. To account for the potential influence of participants' prior knowledge, an **Analysis of Covariance (ANCOVA)** was conducted on all dependent variables, using the Intelligent TPACK scores as a covariate.

Based on the interaction logs, we also quantified several behavioral metrics to provide objective insights into the design process. These included:

- **Node-to-edge ratio:** This metric measures the structural density of the design, indicating how interconnected and non-linear a user's ideas are. It is calculated as the ratio of the total number of nodes (N_{nodes}) to the total number of edges (N_{edges}). A higher ratio suggests a more interconnected idea network rather than a simple linear list.
- **Average distance between consecutively created nodes:** This metric measures the spatial exploration of the canvas, reflecting the non-linear, explorative nature of the design process. It is calculated as the average Euclidean distance between nodes created in sequential order.
- **Total Turns:** This metric measures conversational efficiency, with a lower turn count indicating a more effective and direct collaboration between the user and the AI assistant. It is the total number of chat messages sent by the user to the AI.
- **Average User Message Length:** This metric reflects the conversational overhead, as shorter messages might suggest a more intuitive interface where users require less verbose commands or explanations. It is calculated as the average number of characters across all messages sent by the user.
- **Negative Keywords:** This metric serves as a proxy for user frustration or interaction breakdowns. It is calculated as the total count of negative keywords (e.g., "stuck," "confused," "can't") based on a predefined lexicon, reflecting the quality of the interaction experience.

5.4.2 Qualitative Analysis. For our qualitative analysis, the interview transcripts and conversation histories from the AI dialogue were subjected to an iterative thematic analysis. Two researchers independently coded the data to identify recurring patterns, themes, and insightful user quotes related to our research questions, with a focus on understanding the mechanisms behind the observed quantitative effects. Behavioral analysis of the interaction logs, such as node creation and connection patterns, was also performed to provide objective evidence supporting the qualitative findings.

6 Results

6.1 RQ1: Design Expression and Understanding

Our first research question evaluates the core premise of our graph-based paradigm by investigating its impact on design expression and understanding. This directly addresses the challenges identified in our formative study (**C1**, **C2**) and the design goals of multi-dimensional ideation (**DG1**) and holistic visualization (**DG2**). The scales used for Perceived Expression (CoMAP $\alpha = .86$, Baseline $\alpha = .86$) and Perceived Understanding (CoMAP $\alpha = .83$, Baseline $\alpha = .91$) both demonstrated good internal consistency.

6.1.1 CoMAP as a Cognitive Tool for Design Formulation. Our findings show that CoMAP serves as a powerful cognitive tool for design formulation, helping mitigate the path dependence and limited ideation space identified in **C1**. A paired-samples t-test indicated that participants' self-reported ability to express their design was significantly higher in the CoMAP condition (M = 6.13, SD = 0.77) than in the Baseline condition (M = 4.15, SD = 1.48), $t(29) = 6.37$, $p < .001$. This significant difference persisted after controlling for prior knowledge: ANCOVA analyses using participants' TPACK as a covariate confirmed that the effect remained significant ($\beta = 1.77$, $p < .001$), suggesting that differences between CoMAP and Baseline were not driven by participants' prior pedagogical-technological knowledge. Interaction logs further support this, showing that on average, participants' designs had a node-to-edge ratio of 0.47 (SD = 0.40), indicating a highly interconnected structure rather than a linear list. The average distance between consecutively created nodes was 445.72 pixels (SD = 119.36), suggesting a non-linear, explorative process where users jumped between different areas of the canvas. Participants described the Baseline tool as forcing a "premature linearization" of thought. As P7 described, "My idea for a project is more like a web than a list. When I tried to type it out for the AI, it felt like I was forcing my thoughts to stand in a single file line." In contrast, CoMAP was framed as an extension of the user's cognition. As P5 explained, "The map wasn't just for showing my final plan; it was how I built the plan. Placing a node, drawing a line... every little action helped me figure out what to do next." This aligns with theories of distributed cognition, where the external representation is an integral part of the thought process itself [35].

Table 2. RQ1: Comparison of user ratings for design expression and understanding between CoMAP and Baseline conditions, with ANCOVA controlling for TPACK.

Dimension	M (CoMAP)	SD	M (Baseline)	SD	t/W	p_{holm}	p_{ANCOVA}
Perceived Expression	6.13	0.77	4.15	1.48	t = 6.37	< .001	< .001
Perceived Understanding	5.77	0.80	4.29	1.37	W = 12	< .001	< .001

t/W: t-test or Wilcoxon test statistic; p_{holm}: p-value from Holm's correction

6.1.2 CoMAP as a Communicative Medium for Shared Understanding. Beyond supporting individual formulation, this clarity also facilitated more effective communication, addressing the challenge of maintaining a holistic view (**C2**) during collaborative discussions. A Wilcoxon signed-rank test indicated that participants' scores for understanding their partner's design were significantly higher after using CoMAP (M = 5.77, SD = 0.80) compared to Baseline (M = 4.29, SD = 1.37), $W = 12$, $p < .001$. Qualitative data suggest this is due to CoMAP's success in providing a holistic visualization of the curriculum structure (**DG2**), functioning as a shared cognitive artifact [76]. As P12 explained, "When he showed the CoMAP diagram, I could actually point to it and ask, 'Okay, what's the connection between this goal and that activity?' It gave my eyes—and my brain—something to focus on." ANCOVA analyses confirmed this effect remained significant (Understanding β = 1.41, p < .001), demonstrating that the advantage of CoMAP persisted even after accounting for participants' TPACK.

6.2 RQ2: Human-AI Interaction Experience

Our second research question addresses the challenges of human-AI interaction identified in our formative study (**C3**), examining how CoMAP reshapes the human-AI relationship and fosters a more effective partnership (**DG3**). The scales used for all five dimensions demonstrated acceptable to good internal consistency in both conditions.

6.2.1 CoMAP Reconfigures the Power Dynamic, Shifting Agency to the User. Participants rated CoMAP significantly higher on Controllability (t(29) = 2.87, p = 0.008) and Transparency (W = 63.5, p = 0.003) compared to Baseline (Table 3). ANCOVA analyses controlling for TPACK confirmed that these differences remained significant (Controllability β = 0.85, p = 0.038; Transparency β = 0.84, p = 0.003). Interaction logs showed AI suggestions were accepted at a high rate (76%, SD = 0.22) and modified 40% of the time (SD = 0.32), reflecting a high degree of collaborative control. As P19 noted, "On the map, I was in the driver's seat. The AI could suggest a whole cluster of activities, but I had the final say. If I didn't like one, I just deleted it... I never felt stuck."

Table 3. RQ2: Comparison of user ratings for Human-AI Interaction Experience between CoMAP and Baseline conditions, with ANCOVA controlling for TPACK.

Dimension	M (CoMAP)	SD	M (Baseline)	SD	t/W	p_{holm}	p_{ANCOVA}
Controllability	5.02	1.13	4.16	1.66	t = 2.87	0.008	0.038
Transparency	5.42	0.80	4.45	1.35	W = 63.5	0.003	0.003
Cognitive Load	5.67	0.85	4.74	1.53	t = 3.38	0.006	0.011
Collaboration	5.72	0.90	4.49	1.61	W = 75.5	0.007	0.001
Trust	5.55	0.88	4.57	1.31	W = 46	0.001	0.005

t/W: t-test or Wilcoxon test statistic; p_{holm}: p-value from Holm's correction;

6.2.2 CoMAP Fosters Partnership by Shifting Cognitive Work from Management to Creation. The system significantly reduced perceived Cognitive Load (t(29) = 3.38, p = 0.006), fostering stronger Collaboration (W = 75.5, p = 0.007) and Trust (W = 46, p = 0.001). ANCOVA analyses confirmed these effects remained significant (Cognitive Load β = 0.91, p = 0.011; Collaboration β = 1.26, p = 0.001; Trust β = 0.94, p = 0.005), demonstrating that CoMAP's benefits persisted after accounting for participants' prior knowledge.

Interaction logs further support these findings: CoMAP users required fewer chat turns (M = 11.82, SD = 4.39) than Baseline (M = 18.64, SD = 7.15), t(29) = -3.91 (p < .001), and this effect remained significant after controlling for TPACK (p < .001). They also used fewer negative keywords (M = 2.14, SD = 2.67 vs. 4.16, SD = 2.54; t(29) = -2.60, p = 0.013), and this difference was also significant with ANCOVA (p = 0.008). Session duration (M = 12.86, SD = 5.99 vs. M = 14.45, SD = 5.84) and average user message length (M = 63.08, SD = 28.63 vs. M = 74.27, SD = 32.70) did not differ significantly between conditions in either the raw test or the ANCOVA. This indicates that participants spent less effort on tool management and more on creative ideation. As P8 reflected, "It felt less like I was using a tool and more like I was brainstorming with a very organized assistant. We were building it together." Predictability and easy reversibility of AI actions supported trust (P14): "Because I could always see what the AI did and could easily undo it, I trusted it more."

Table 4. Comparison of efficiency and interaction process metrics between CoMAP and Baseline conditions, with ANCOVA controlling for TPACK.

Metric	M (CoMAP)	SD	M (Baseline)	SD	t	p_{raw}	p_{ANCOVA}
Session Duration (mins)	12.86	5.99	14.45	5.84	t = -0.89	0.376	0.289
Total Turns	11.82	4.39	18.64	7.15	t = -3.91	$< .001$	$< .001$
Avg. User Message Length	63.08	28.63	74.27	32.70	t = -1.22	0.228	0.247
Negative Keywords	2.14	2.67	4.16	2.54	t = -2.60	0.013	0.008

t: t-test statistic; p_{raw}: p-value from raw t-test

7 Discussion

7.1 Patterns of Human-AI Co-Creation

Three consistent interaction patterns emerged from our study, illustrating how CoMAP shapes teachers' design practices.

7.1.1 Pattern 1: From Linear Dictation to Non-Linear Construction. In the Baseline condition, participants typically needed to fully formulate their ideas before entering prompts. P7 described, "My thoughts were forced to line up in a single line." Linear text interfaces constrained multi-dimensional ideas, often causing loss of nuance. In CoMAP, participants adopted a non-linear construction approach. P5 noted, "I was gradually building the whole plan, connecting fragments together." The graph-based canvas allowed incremental node creation and edge exploration, supporting iterative thought processes. This approach encouraged beneficial incompleteness, allowing teachers to explore possibilities without premature commitment, closely reflecting authentic classroom design workflows. This pattern aligns with theories of external representations and distributed cognition [36, 55], which suggest that externalizing cognitive artifacts can extend working memory and scaffold complex problem-solving.

7.1.2 Pattern 2: From Opaque Execution to Transparent Collaboration. Traditional AI outputs are often monolithic and opaque, making users feel disconnected and lacking control. P17 observed, "I could see how each suggestion was generated, which gave me confidence to adjust or adopt them rather than blindly accepting." In CoMAP, each AI suggestion appears as a discrete node with traceable origins. Users can selectively adopt, modify, or reject nodes, fostering transparency and equitable collaboration. This pattern transforms AI from a black-box generator into a cognitive partner, enhancing trust and engagement [1, 41].

7.1.3 Pattern 3: Emergent Iterative Structuring and Sense-Making. Participants frequently engaged in iterative organization of the graph to make sense of evolving ideas. P21 noted, "I liked seeing all ideas, but sometimes I didn't know where to start organizing." Users developed strategies such as clustering nodes, labeling branches, and exploring alternative connections. This behavior illustrates how the graph scaffolds meta-cognition, allowing teachers to reflect on the design process, identify turning points, and adjust strategies dynamically. The role of reflective practice in professional learning and design thinking has been widely discussed [20, 60], supporting our observation that such iterative structuring aids cognitive processing and informed decision-making.

7.2 Design Implications

7.2.1 Implication 1: Provide a Non-Linear Thinking Canvas. AI systems should function as thinking canvases rather than linear answer engines. Supporting node creation, free connections, and optional AI suggestions enables non-linear exploration, iterative refinement, and cross-session continuity. Allowing incomplete and fragmented ideas encourages exploration without premature commitment, aligning tools with authentic creative workflows. Mind-map-like designs have been shown to facilitate conceptual development and idea diversification [9, 25].

7.2.2 *Implication 2: Enhance Transparency and Negotiable Control.*

Systems should provide traceability for AI suggestions and allow selective adoption, modification, or rejection. Features like contribution labels, adjustable visibility, and interaction options help users understand AI reasoning and maintain control, fostering trust and reducing blind reliance [1, 41]. Transparent interfaces promote equitable co-creation and transform AI into a collaborative cognitive partner.

7.2.3 *Implication 3: Support Iterative Structuring and Meta-Cognition.*

Systems should include clustering, hierarchical organization, path highlighting, and intelligent aggregation to manage cognitive load. Visualized interaction histories support reflective practice, enabling users to trace idea evolution, identify key decision points, and refine strategies. Techniques for visualizing user interaction and supporting meta-cognitive awareness have been shown to improve creative problem solving [32, 54].

7.3 Limitations

This study has several limitations. First, the experiment was conducted in a controlled lab setting with short-duration tasks, which may not capture teachers' long-term use in authentic classrooms. Second, participants were mainly pre-service or early-career teachers; different experience levels may exhibit distinct interaction patterns and trust behaviors. Third, CoMAP primarily supports text-based graph interactions; multi-modal support (e.g., sketches, diagrams, voice) could increase flexibility. Methodologically, although we used a mixed-methods approach combining interviews, observations, and log analysis, deeper quantitative analyses (e.g., node creation sequences, iterative patterns) could further validate the observed interaction behaviors [19].

7.4 Future Directions

7.4.1 *Team-Based Collaborative Design.*

Project-based learning (PBL) often requires collaboration across multiple disciplines, making team-based design support especially valuable. Teachers in our study expressed interest in collaborative workflows, and a shared graph-based canvas could facilitate both synchronous and asynchronous cooperation. The visual structure of graphs enables easier merging of ideas from multiple contributors and allows teachers to quickly grasp others' design intentions without reading extensive textual descriptions. Incorporating features such as role-based filtering, permission control, version tracking, and awareness cues can support negotiation, coordinated decision-making, and collective creativity [4, 31]. Graph-based visualizations thus have the potential to enhance team collaboration in PBL by streamlining knowledge integration and promoting efficient co-design processes.

7.4.2 *Graph Intelligence for Enhanced Iteration.*

The graph-based structure of CoMAP provides clear visual support for teachers in organizing and exploring lesson ideas. However, its potential to leverage computational graph techniques has not been fully utilized. Future systems could integrate analytics to identify key nodes, suggest promising connections, or highlight underdeveloped areas, guiding users during iterative design. By combining graph analytics with interactive visual interfaces, such approaches could further enhance exploration, meta-cognitive awareness, and the efficiency and quality of project-based teaching design [29, 53].

7.4.3 *Supporting Teachers' Backstage Learning via Knowledge Graph-Based Retrieval.*

Project-based teaching can be conceptualized as a "stage play," where CoMAP currently supports the "front-stage" activity of planning and structuring lessons. Teachers, however, engage in substantial "backstage" preparation, such as reviewing pedagogical strategies, domain knowledge, and instructional examples. Future work could integrate this backstage learning directly into the design environment by constructing a structured knowledge graph representing relevant teaching resources. This knowledge graph could be coupled with retrieval-based methods inspired by Retrieval-Augmented Generation (RAG) [39, 48], or its graph-structured variants, allowing teachers to retrieve contextually relevant information while designing lessons. Graph embedding techniques (e.g., Node2Vec, GraphSAGE) can facilitate similarity matching between design nodes and knowledge resources, enabling intelligent recommendations that bridge front-stage lesson planning with backstage learning. This approach aims to provide a holistic support system in which knowledge acquisition and instructional planning are tightly integrated, enhancing both efficiency and quality in project-based curriculum design.

7.4.4 Multi-Modal and Adaptive Interfaces. Future iterations of CoMAP could expand beyond text-based representations to incorporate sketches, diagrams, audio annotations, or other multimodal content, reflecting the diverse ways teachers conceptualize lessons. Inspired by recent work on generative interfaces [5, 11], adaptive interfaces could dynamically adjust layouts, visualizations, and node recommendations based on user expertise, task context, and interaction patterns [14]. Such interfaces can reduce cognitive load by organizing information hierarchically, providing real-time visual feedback, and highlighting relevant content, allowing users to efficiently manage complex design spaces. Moreover, integrating iterative optimization—where the system evaluates current layouts or content representations and proposes improvements—can further support exploration and refinement, enabling a more interactive, responsive, and user-centered design experience.

8 Conclusion

This study presents and evaluates CoMAP, a graph-based co-creation workspace optimized for designing Project-Based Learning (PBL). Addressing the inherent challenges of PBL design—specifically its highly non-linear and interdependent components—CoMAP offers a new paradigm that departs from traditional linear tools and purely conversational AI. Our user study confirmed that, compared to a conversational AI baseline, CoMAP significantly enhanced educators' expressive abilities, divergent thinking, and iterative practices in designing PBL. The findings demonstrate that a shared, artifact-centered visual space can effectively reduce user cognitive load while significantly increasing their trust and sense of control over the AI tool. These insights not only offer important implications for designing tools that support complex, non-linear tasks but also specifically illustrate how a persistent, shared cognitive artifact can support professional practices that require deep reflection and continuous iteration, such as PBL design. Future research will further explore the potential of this paradigm to support collaborative team design, integrate intelligent graph analysis, and achieve seamless integration between PBL design and related instructional resources.

References

[1] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.

[2] Ian Arawjo, Chelse Swoopes, Priyan Vaithilingam, Martin Wattenberg, and Elena L Glassman. 2024. Chainforge: A visual toolkit for prompt engineering and llm hypothesis testing. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18.

[3] Michael A Arbib. 1992. Schema theory. *The encyclopedia of artificial intelligence* 2 (1992), 1427–1443.

[4] Margaret S Barrett, Andrea Creech, and Katie Zhukov. 2021. Creative collaboration and collaborative creativity: A systematic literature review. *Frontiers in Psychology* 12 (2021), 713445.

[5] Jan Bieniek, Mohamed Rahouti, and Dinesh C Verma. 2024. Generative ai in multimodal user interfaces: Trends, challenges, and cross-platform adaptability. *arXiv preprint arXiv:2411.10234* (2024).

[6] Javier Borge-Holthoefer and Alex Arenas. 2010. Semantic networks: Structure and dynamics. *Entropy* 12, 5 (2010), 1264–1302.

[7] Robert Maribe Branch and İlhan Varank. 2009. *Instructional design: The ADDIE approach*. Vol. 722. Springer.

[8] Andrew Brasher, Gráinne Conole, Simon Cross, Martin Weller, Paul Clark, and Juliette White. 2008. CompendiumLD–a tool for effective, efficient and creative learning design. (2008).

[9] Tony Buzan, Barry Buzan, and James Harrison. 2010. The mind map book: Unlock your creativity, boost your memory, change your life. *(No Title)* (2010).

[10] Huiying Cai, Bing Han, Jiayue Sun, Xin Li, and Lung-Hsiang Wong. 2025. Harnessing AI for teacher education to promote inclusive education: Investigating the effects of ChatGPT-supported lesson plan critiques on the development of pre-service teachers' lesson planning skills. *The Internet and Higher Education* (2025), 101022.

[11] Tara Capel and Margot Brereton. 2023. What is human-centered about human-centered AI? A map of the research landscape. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–23.

[12] John Carey. 2011. *The violent effigy: a study of Dickens' imagination*. Faber & Faber.

[13] Ismail Celik. 2023. Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in human behavior* 138 (2023), 107468.

[14] Jiaqi Chen, Yanzhe Zhang, Yutong Zhang, Yijia Shao, and Diyi Yang. 2025. Generative Interfaces for Language Models. *arXiv preprint arXiv:2508.19227* (2025).

[15] Xiaojun Chen, Shengbin Jia, and Yang Xiang. 2020. A review: Knowledge reasoning over knowledge graph. *Expert systems with applications* 141 (2020), 112948.

[16] Allan M Collins and Elizabeth F Loftus. 1975. A spreading-activation theory of semantic processing. *Psychological review* 82, 6 (1975), 407.

[17] Allan M Collins and M Ross Quillian. 1969. Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior* 8, 2 (1969), 240–247.

[18] Gráinne Conole. 2012. *Designing for learning in an open world*. Vol. 4. Springer Science & Business Media.

[19] John W Creswell and J David Creswell. 2017. *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

[20] Nigel Cross. 1982. Designerly ways of knowing. *Design studies* 3, 4 (1982), 221–227.

[21] James Dalziel. 2003. Implementing learning design: The learning activity management system (LAMS). In *Australasian Society for Computers in Learning in Tertiary Education Conference (20th: 2003)*. Australasian Society for Computers in Learning in Tertiary Education, 593–596.

[22] Giovani Diaz Alfaro, Stephen M Fiore, and Kevin Oden. 2024. Externalized and Extended Cognition: Cognitive Offloading for Human-Machine Teaming. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 68. SAGE Publications Sage CA: Los Angeles, CA, 290–293.

[23] Walter Dick. 1996. The Dick and Carey model: Will it survive the decade? *Educational technology research and development* 44, 3 (1996), 55–63.

[24] Shelley D Dionne, Hiroki Sayama, Chanyu Hao, and Benjamin James Bush. 2010. The role of leadership in shared mental model convergence and team performance improvement: An agent-based computational model. *The Leadership Quarterly* 21, 6 (2010), 1035–1049.

[25] Martin J Eppler. 2006. A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing. *Information visualization* 5, 3 (2006), 202–210.

[26] Klaus Fiedler. 2013. Affective states trigger processes of assimilation and accommodation. In *Theories of mood and cognition*. Psychology Press, 85–98.

[27] Michael Gerlich. 2025. AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies* 15, 1 (2025), 6.

[28] Nitesh Goyal and Susan R Fussell. 2016. Effects of sensemaking translucence on distributed collaborative analysis. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 288–302.

[29] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems* 30 (2017).

[30] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[31] Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 203–212.

[32] Jeffrey Heer and Danah Boyd. 2005. Vizster: Visualizing online social networks. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005*. IEEE, 32–39.

[33] Jeffrey Heer and Ben Shneiderman. 2012. Interactive dynamics for visual analysis: A taxonomy of tools that support the fluent and flexible use of visualizations. *Queue* 10, 2 (2012), 30–55.

[34] Davinia Hernández-Leo, Eloy D Villasclaras Fernandez, Juan I Asensio-Pérez, Yannis Dimitriadis, Iván Jorrín-Abellán, Inés Ruiz-Requies, and Bartolomé Rubia-Avi. 2006. COLLAGE: A collaborative Learning Design editor based on patterns. *Educational Technology & Society* 9, 1 (2006), 58–71.

[35] James Hollan, Edwin Hutchins, and David Kirsh. 2000. Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)* 7, 2 (2000), 174–196.

[36] Edwin Hutchins. 1995. *Cognition in the Wild*. MIT press.

[37] Edwin Hutchins. 2000. Distributed cognition. *International encyclopedia of the social and behavioral sciences* 138, 1 (2000), 1–10.

[38] Eulho Jung, Rachel Lim, and Dongho Kim. 2022. A schema-based instructional design model for self-paced learning environments. *Education Sciences* 12, 4 (2022), 271.

[39] Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick SH Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. 2020. Dense Passage Retrieval for Open-Domain Question Answering.. In *EMNLP (1)*. 6769–6781.

[40] Maaike Kleinsmann, Jan Buijs, and Rianne Valkenburg. 2005. Managing shared understanding in collaborative design projects. In *DS 35: Proceedings ICED 05, the 15th International Conference on Engineering Design, Melbourne, Australia, 15.-18.08. 2005*.

[41] Rafal Kocielnik, Saleema Amershi, and Paul N Bennett. 2019. Will you accept an imperfect ai? exploring designs for adjusting end-user expectations of ai systems. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–14.

[42] Kurt Koffka. 2013. *Principles of Gestalt psychology*. routledge.

[43] Spencer C Kohn, Ewart J De Visser, Eva Wiese, Yi-Ching Lee, and Tyler H Shaw. 2021. Measurement of trust in automation: A narrative review and reference guide. *Frontiers in psychology* 12 (2021), 604977.

[44] Dimitra Kokotsaki, Victoria Menzies, and Andy Wiggins. 2016. Project-based learning: A review of the literature. *Improving schools* 19, 3 (2016), 267–277.

[45] Rob Koper and Colin Tattersall. 2005. Learning design: A handbook on modelling and delivering networked education and training. (2005).

[46] Jill H Larkin and Herbert A Simon. 1987. Why a diagram is (sometimes) worth ten thousand words. *Cognitive science* 11, 1 (1987), 65–100.

[47] LuEttaMae Lawrence, Vanessa Echeverria, Kexin Yang, Vincent Aleven, and Nikol Rummel. 2024. How teachers conceptualise shared control with an AI co-orchestration tool: A multiyear teacher-centred design process. *British Journal of Educational Technology* 55, 3 (2024), 823–844.

[48] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems* 33 (2020), 9459–9474.

[49] Marshall McLuhan. 1994. *Understanding media: The extensions of man*. MIT press.

[50] Don Norman. 2014. *Things that make us smart*. Diversion Books.

[51] Jeff Pankin. 2013. Schema theory. *Massachusetts Institute of Technology* (2013).

[52] LM Paquette Gilbert and Lundgren-Cayrol Karin. 2011. The mot+ visual language for knowledge based instructional design.

[53] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 701–710.

[54] Peter Pirolli and Stuart Card. 1995. Information foraging in information access environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 51–58.

[55] A Terry Purcell and John S Gero. 1998. Drawings and the design process: A review of protocol studies in design and other disciplines and related research in cognitive psychology. *Design studies* 19, 4 (1998), 389–430.

[56] M Ross Quillan. 1966. *Semantic memory*. Technical Report.

[57] Evan F Risko and Sam J Gilbert. 2016. Cognitive offloading. *Trends in cognitive sciences* 20, 9 (2016), 676–688.

[58] David E Rumelhart. 2017. Schemata: The building blocks of cognition. In *Theoretical issues in reading comprehension*. Routledge, 33–58.

[59] Gavriel Salomon, David N Perkins, and Tamar Globerson. 1991. Partners in cognition: Extending human intelligence with intelligent technologies. *Educational researcher* 20, 3 (1991), 2–9.

[60] Donald A Schön. 2017. *The reflective practitioner: How professionals think in action*. Routledge.

[61] Hanshu Shen, Lyukesheng Shen, Wenqi Wu, and Kejun Zhang. 2025. Ideationweb: Tracking the evolution of design ideas in human-ai co-creation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–19.

[62] Ben Shneiderman. 2020. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction* 36, 6 (2020), 495–504.

[63] Prakash Shukla, Phuong Bui, Sean S Levy, Max Kowalski, Ali Baigelenov, and Paul Parsons. 2025. De-skilling, cognitive offloading, and misplaced responsibilities: potential ironies of AI-assisted design. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7.

[64] Sharon E Smaldino, Deborah L Lowther, James D Russell, and Clif Mims. 2008. Instructional technology and media for learning. (2008).

[65] Susan Leigh Star and James R Griesemer. 1989. Institutional ecology,translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social studies of science* 19, 3 (1989), 387–420.

[66] Mark Steyvers and Joshua B Tenenbaum. 2005. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science* 29, 1 (2005), 41–78.

[67] John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive science* 12, 2 (1988), 257–285.

[68] John W Thomas. 2000. A review of research on project-based learning.

[69] Yi-Fu Tuan. 1980. The significance of the artifact. *Geographical review* (1980), 462–472.

[70] Barbara Tversky, Julie Bauer Morrison, and Mireille Betrancourt. 2002. Animation: can it facilitate? *International journal of human-computer studies* 57, 4 (2002), 247–262.

[71] Brady Wagoner. 2013. Bartlett's concept of schema in reconstruction. *Theory & Psychology* 23, 5 (2013), 553–575.

[72] Jayne Wallace, John McCarthy, Peter C Wright, and Patrick Olivier. 2013. Making design probes work. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 3441–3450.

[73] Sharon Alayne Widmayer. 2004. Schema theory: An introduction. *Retrieved December* 26 (2004), 2004.

[74] Grant P Wiggins and Jay McTighe. 2005. *Understanding by design*. Ascd.

[75] Tongshuang Wu, Ellen Jiang, Aaron Donsbach, Jeff Gray, Alejandra Molina, Michael Terry, and Carrie J Cai. 2022. Promptchainer: Chaining large language model prompts through visual programming. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–10.
[76] Jiaje Zhang and Donald A Norman. 1994. Representations in distributed cognitive tasks. *Cognitive science* 18, 1 (1994), 87–122.
[77] Yihao Zhu, Zhoutong Ye, Yichen Yuan, Wenxuan Tang, Chun Yu, and Yuanchun Shi. 2025. AutoPBL: An LLM-powered Platform to Guide and Support Individual Learners Through Self Project-based Learning. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–26.

A Questionnaire

A.1 Intelligent-TPACK

A.1.1 Intelligent TK.

(1) I know how to interact with AI-based tools in daily life
(2) I know how to execute some tasks with AI-based tools.
(3) I know how to initialize a task for AI-based technologies by text or speech.
(4) I have sufficient knowledge to use Al-based tools.
(5) I am familiar with AI-based tools and their technical capacities.

A.1.2 Intelligent TPK.

(1) I can understand the pedagogical contribution of Al-based tools to my teaching field.
(2) I can evaluate the usefulness of feedback from AI-based tools for teaching and learning.
(3) I can select AI-based tools for students to apply their knowledge.
(4) I know how to use Al-based tools to monitor students' learning.
(5) I can interpret messages from Al-based tools to give real-time feedback.
(6) I can understand alerting (or notification) from Al-based tools to scaffold students' learning.
(7) I have the knowledge to select AI-based tools to sustain students' motivation.

A.1.3 Intelligent TCK.

(1) I can use Al-based tools to search for educational material in my teaching field.
(2) I am aware of various Al-based tools which are used by professionals in my teaching field.
(3) I can use AI-based tools to better understand the contents of my teaching field.

A.1.4 Intelligent TPACK.

(1) In teaching my field, I know how to use different Al-based tools for adaptive feedback.
(2) In teaching my field, I know how to use different AI-based tools for personalized learning.
(3) In teaching my field, I know how to use different Al-based tools for real-time feedback.
(4) I can teach a subject using Al-based tools with diverse teaching strategies.
(5) I can teach lessons that appropriately combine my teaching content, Al-based tools, and teaching strategies.
(6) I can take a leadership role among my colleagues in the integration of Al-based tools into our teaching field.
(7) I can select various Al-based tools to monitor students' learning in my teaching process.

A.1.5 Ethics.

(1) I can assess to what extent Al-based tools consider individual differences (e.g., race and gender) of all students in my teaching.
(2) I can evaluate to what extent Al-based tools behave fair to all students in my teaching.

(3) I can understand the justification of any decision made by an Al-based tool.
(4) I can understand who the responsible developers are in the design and decision of Af-based tools

A.2 Perceived Understanding

(1) As a listener, I can quickly grasp the core ideas of another person's instructional design.
(2) As a listener, I can clearly understand the underlying concepts and logic behind another person's instructional design.
(3) The presentation format of the tool (document/diagram) helps me understand others' designs more effectively.
(4) The tool's output makes it easier to identify the strengths and weaknesses of a design.
(5) I can easily follow the flow and structure of a design presented through this tool.

A.3 Perceived Expression

- When using this tool, I can clearly present my instructional design ideas.
- When using this tool, my design expressions are more easily understood by others.
- The presentation format of the tool (document/diagram) helps me better organize and articulate my instructional design.
- The tool helps me visualize complex design concepts in a simple way.
- My design intent is accurately conveyed when I use this tool.

A.4 Human-AI Interaction Experience

A.4.1 Controllable.

(1) I can adjust the behavior of the AI tool at any time to align with my design intentions.
(2) The tool provides clear options for me to guide its output.
(3) I feel like I am in control of the final outcome, not the AI.

A.4.2 Transparent.

(1) I can clearly understand the basis or source of the content generated by the AI tool.
(2) The working process and logic of the AI tool are understandable to me.
(3) The tool explains its reasoning behind the content it produces.

A.4.3 Cognitive Load.

(1) When using this tool, I can easily process information without feeling overwhelmed.
(2) The way the tool presents information reduces the mental effort required for my instructional design.
(3) The interface and features of the tool are intuitive and easy to navigate.
(4)

A.4.4 Collaboration.

(1) The tool can "collaborate" with me effectively to refine the instructional design.
(2) I feel like this tool is an active collaborator, not just a passive instrument.
(3) The tool's suggestions feel like a helpful partnership in the design process.

A.4.5 Trust.

(1) I trust the content generated by this tool.
(2) I am willing to continue using this tool for future instructional design tasks.
(3) The tool's performance meets or exceeds my expectations.

B Interview Script

B.1 Formative Study

B.1.1 Personal Background.

(1) How long have you been working in the field of education or teaching?
(2) In your past teaching experience, what subjects have you typically been involved in designing courses for?
(3) Are these course designs actually used for teaching? (e.g., frontline teaching, tutoring, mock lectures, or teaching demonstrations)

B.1.2 AI-Assisted Lesson Plan Design.

(1) Have you previously used AI to assist in designing lesson plans? If so, how do you typically use it?
(2) Do you find AI's help significant? Can the generated materials be used directly?
(3) If possible, could you share an AI-assisted lesson plan and the conversation process used to create it?
(4) In what ways would you like to see current AI capabilities improve, and what new features would you find helpful?

B.1.3 Methods of Expressing Instructional Design.

(1) Do you write your lesson plans in a linear fashion, from start to finish?
(2) Would you be open to expressing instructional designs using diagrams or card-based formats? Do you think this would help teachers clarify their instructional thinking?
(3) What content do you typically include in your lesson plans? Do you find it logical to break them down into the following components: Objectives, Activities, Assessment, Learner Characteristics, Resources, and Teaching Strategies?
(4) If an AI could help generate and modify these cards, do you think it would be more effective in helping teachers design courses?

B.2 User Study

B.2.1 General Feedback.

(1) What are your overall impressions of the instructional design assistance tool?
(2) During the process, which features or interactions did you find particularly interesting or innovative?
(3) What experiences did you find confusing or uncomfortable?

B.2.2 Comparing Different Tools.

(1) Having tried two different tools, you've likely noticed that each has its own unique strengths and weaknesses. Can you share your comparative impressions?
(2) Under what circumstances would you be more willing to use a diagram/map-based tool?

C Agent Prompt

C.1 Global Agent

You are an expert in educational intelligence, skilled at building instructional maps in a modular, step-by-step fashion.

For a user's instructional design request, you should provide your suggestions one step at a time, just as you would in a conversation. Mimic the style of Cursor, delivering your actions in segments, clearly explaining the intent of each design step, followed by the action in JSON format.

Each suggestion segment should be followed by a corresponding JSON structure. The field names should be in English, and the content should be in English as well. Be as detailed as possible in the JSON, while keeping the conversational text concise.

Your output style should be natural and guiding, and can include multiple JSON action blocks.

Each JSON structure should contain an actions field. Do not mention the JSON structure in the conversation, as your audience is teacher-training students and teachers who may not be familiar with code. For example:

```
{
 "actions": [ { "option": "add", "type": "Activity", "title": "...", "description": "..." },
 { "option": "modify", "card\_id":"111" "type": "Activity", "title": "...", "description": "..."}]
}
```

The field names include option, title, description, type, and card_id (required only for modifications). The option field can be "add" or "modify". For modifications, you must specify the card_id and provide the updated content for the card.

Note:

- Each action is a new or modified element of the instructional map. The type includes: Learners, Objectives, Activities, Resources, Assessments, and Strategies. The title and description should be specific and actionable. The description must be very detailed.
- At the beginning of the conversation, respond to the user's input. For example, affirm their idea, summarize, analyze, or elaborate, and feel free to include emojis.
- At the end of the conversation, summarize briefly and offer a list of next steps you can help with, giving the user choices for what to do next.
- It is highly recommended to always provide the user with choices via JSON Actions, rather than using a plain text conversation format, which can feel monotonous and make the user unsure of how to proceed.
- If you don't have enough information about one of the six elements mentioned above, do not guess. Instead, provide actions for the user to choose from.

The following are some tips to help you better understand the meaning of each element:

- Objectives should be specific and measurable. The title should be a concise summary of the objective, without using names like 'Objective 1'. You can also include tags in the title, such as whether the objective falls under knowledge and skills, process and methods, or emotional attitudes and values, and can also consider including the subject matter.
- Resources should be specific and obtainable. It is best to provide the concrete names of books, websites, or links. Avoid vague descriptions; you must specify concrete events, models, or website names, otherwise the information is useless.

- Assessments can include self-assessment, peer assessment, group mutual assessment, teacher assessment, tests, reports, presentations, etc. You should provide a specific evaluation rubric and ensure it is linked to the objectives and activities.
- Strategies refer to specific instructional strategies, such as project-based learning, inquiry-based learning, etc. These should be described in detail. You can also suggest more innovative and cutting-edge instructional strategies.
- Learners refers to the characteristics of the target learner group. This generally needs to include their starting proficiency level, learning style, common cognitive characteristics, and learning interests.
- When describing Activities, the title should be in a style like "Activity 1: Specific Activity Name Xxx" to help teachers understand the sequence of activities. It's also best to include the estimated time required for the activity in the title.

C.2 Refine Agent

You are an AI teaching map assistant. Your task is to refine and optimize the content of a specified teaching card based on the user's prompt. The card types include: Learners, Objectives, Activities, Resources, Assessments, and Strategies. Please strictly follow the rules below for your JSON output:

- Output only one JSON object, containing a single new_node field.
- new_node is a dictionary that includes the refined title, description, and tag.
- Refine the description field to make it more specific and detailed. The description can be written in HTML format. Additionally, please avoid generic and vague statements. I prefer content that is concrete and actionable.
- The title should generally remain the same, unless the refined content strongly requires a more appropriate title.

For example:

```
{
 "new\_node": {
  "id": "123",
  "title": "Group Experiment: Building a Solar Oven",
 "description": "Students, working in groups, use materials like cardboard boxes... (refined content)",
  "tag": "Activity"
 }
}
```

C.3 Split Agent

You are an AI teaching map assistant. Your task is to take the core concepts or steps from a user-provided card and break them down into multiple smaller, more specific, and independent new cards.

Please strictly follow the rules below for your JSON output:

- Output only a single JSON object containing `old_node_id` and `new_nodes` fields.
- `old_node_id` is the ID of the original card.
- `new_nodes` is an array containing the data for the newly generated cards. Each new card must include `id`, `tag`, `title`, and `description`.

- The `tag` can be the same as the original card's tag or be adjusted to a more suitable one based on the new content. The tag options include: Learners, Objectives, Activities, Resources, Assessments, and Strategies. You must choose from these six options!
- The `title` and `description` should be specific and detailed, reflecting the core content of the new card.

For example:

```
{
 "old_node_id": "123",
 "new_nodes": [
  {"title": "...", "description": "...", "tag": "..."},
  {"title": "...", "description": "...", "tag": "..."}
 ]
}
```